STATS

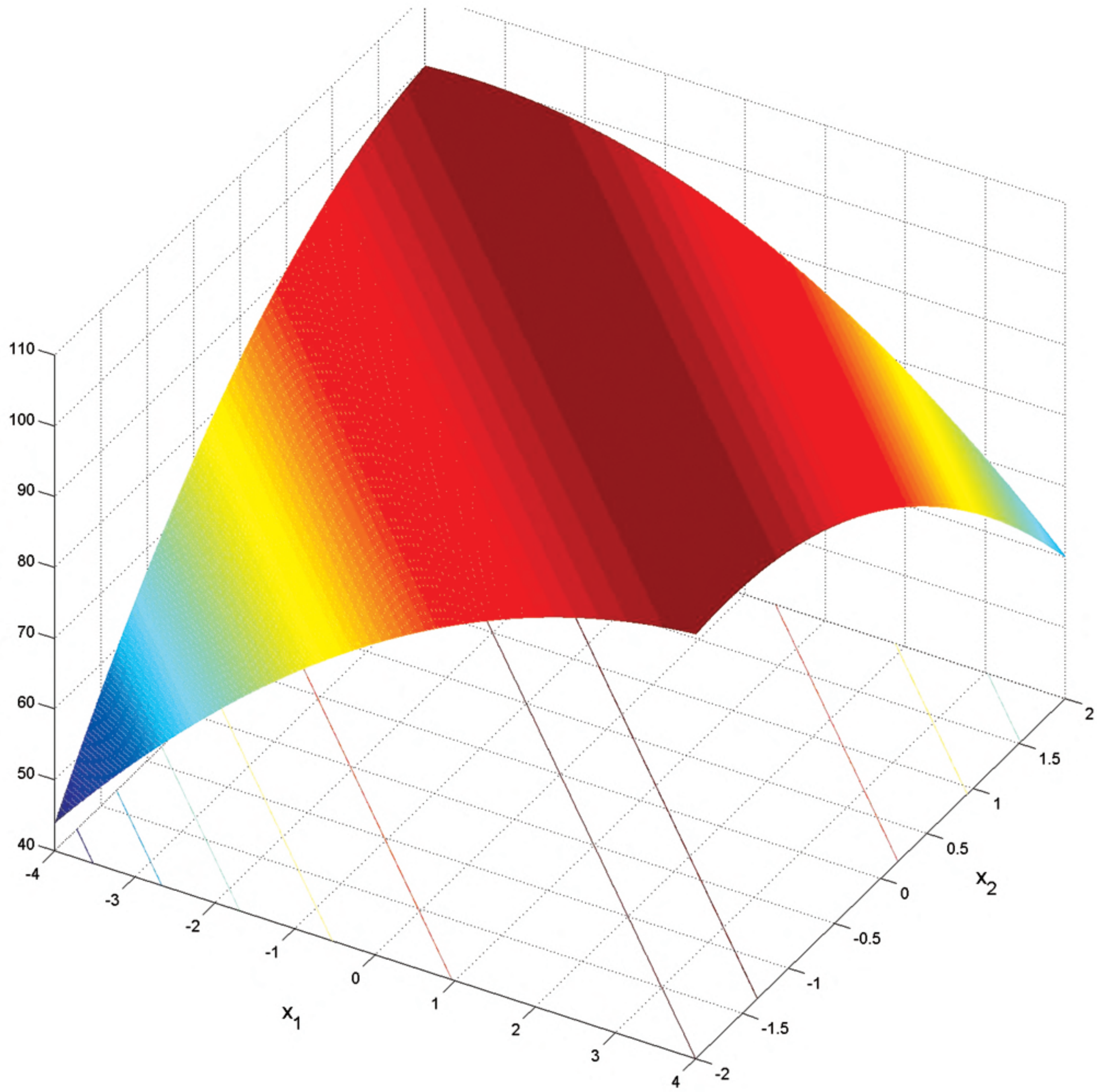
THE MAGAZINE FOR STUDENTS OF STATISTICS : : ISSUE 48

Numb3rs, Sabermetrics, Joe Jackson, and Steroids

Larry Lesser Explains How Learning Stats Is FUN!

Designing a Better Paper Helicopter

Using Response Surface Methodology




STATS

STATS contents

"What nature does blindly, slowly, and ruthlessly, man may do providently, quickly, and kindly. As it lies within his power, so it becomes his duty to work in that direction."

Sir Francis Galton, 1905

features

SPECIAL TOPICS

page 7

columns

COLUMNS

page 22

guest writers

Numb3rs, Sabermetrics, Joe Jackson, and Steroids

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). He is a former editor of *The American Statistician*, enjoys playing tennis, and is an avid baseball fan. His dreams will come true when the Phillies next win the World Series.

Learning Stats Is FUN... with the Right Mode

LAWRENCE (LARRY) LESSER is an associate professor at The University of Texas at El Paso and on the Research Advisory Board of the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE). His web site and contact information are at *www.math.utep.edu/Faculty/lesser/Mathemusician.html.*

Designing a Better Paper Helicopter Using Response Surface Methodology

ERIK BARRY ERHARDT was introduced to response surface methodology while studying for a master's of science degree in statistics at Worcester Polytechnic Institute. Currently, he is a PhD student at the University of New Mexico and a member of the Albuquerque Chapter of the ASA. He enjoys teaching and research, and he likes to continually improve things.

EDITOR

Paul J. Fields
Department of Statistics
Brigham Young University, Provo, UT 84602
pjfields@stat.byu.edu

DIRECTOR OF PROGRAMS

Martha Aliaga
American Statistical Association
732 North Washington Street, Alexandria, VA 22314-1943
martha@amstat.org

EDITORIAL BOARD

Peter Flanagan-Hyde
Mathematics Department
Phoenix Country Day School, Paradise Valley, AZ 85253
pflanaga@pcds.org

Schuyler W. Huck
Department of Educational Psychology and Counseling
University of Tennessee
Knoxville, TN 37996
shuck@utk.educolsen@cr.k12.ia.us

Jackie Miller
Department of Statistics
The Ohio State University, Columbus, OH 43210
jbm@stat.ohio-state.edu

Chris Olsen
Department of Mathematics
George Washington High School, Cedar Rapids, IA 53403
colsen@cr.k12.ia.us

Bruce Trumbo
Department of Statistics
California State University, East Bay, Hayward, CA 94542
bruce.trumbo@csueastbay.edu

PRODUCTION/DESIGN

Megan Murphy
Communications Manager
American Statistical Association

Val Snider
Publications Coordinator
American Statistical Association

Colby Johnson • Lidia Vigyázó
Graphic Designers/Production Coordinators
American Statistical Association

***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, 732 North Washington Street, Alexandria, VA 22314-1943** USA; (703) 684-1221; *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 provided to all student members of the ASA at no additional cost. Subscription rates for others: $15.00 a year for ASA members; $20.00 a year for nonmembers; $25.00 for a Library subscription.

Ideas for feature articles and materials for departments should be sent to Editor Paul J. Fields at the address listed above. Material must be sent as a Microsoft Word document. Accompanying artwork will be accepted in four graphics formats only: EPS, TIFF, PDF, or JPG (minimum 300 dpi). No articles in WordPerfect will be accepted.

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

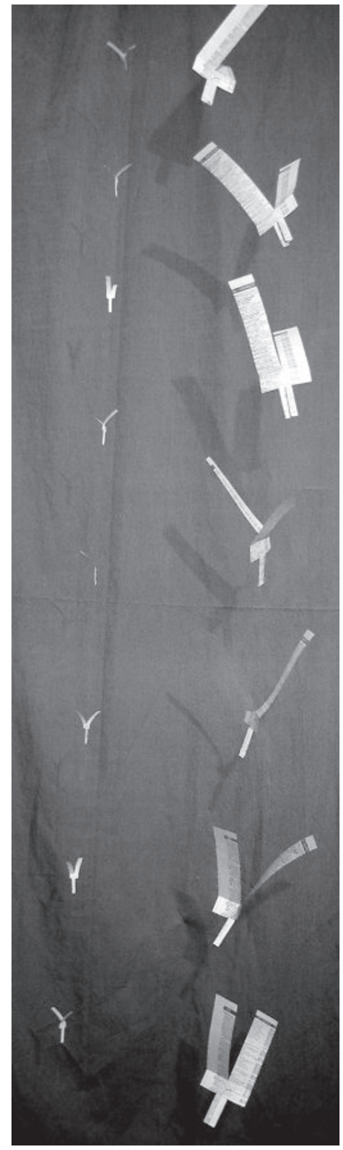
EDITOR'S COLUMN

In this issue, Jim Albert is our "lead-off batter," with an article about statistics on television. Did you see the Numb3rs episode, "Hardball," last November about sabermetrics, the statistical analysis of baseball? Albert is our resident *STATS* sabermetrician. He poses some interesting questions based on that episode. You will be intrigued by his answers. By the way, do you know the origin of the word "sabermetrics"? "Saber" comes from the acronym SABR for the Society for American Baseball Research, and "metrics" refers to measures of performance.

Learning stats can be fun! Just ask Larry Lesser. He has put together a huge collection of jokes, songs, books, games, poems, and more to help us learn statistics and have fun at the same time. There is a long list of references to statistics fun in References and Additional Reading starting on Page 26 and an even longer list on the *STATS* web site, *www.amstat.org/publications/stats*. Think of some more ways to have fun learning stats and send them to us to share with other *STATS* readers.

Jackie Miller, who writes for the column Ask *STATS*, asked Joan Garfield, "What Is Statistics Education?" *STATS* considers Garfield the foremost authority on statistics education in the country—no, make that the world—and she has taken the time to explain to us what statistics education is all about. Read what a true pioneer has to say about this important topic.

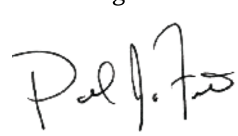
What performance measure would make a paper helicopter "the best"? Maximum hang time, right? Statistics doctoral student Erik Erhardt explains in our cover feature how he and classmate Hantao Mai optimized the design of a paper helicopter to get the best performance. At the right is a strobe photograph of his helicopters in action—a full-size model and a scaled-down "nanocopter." Look at the photo carefully. Which one has the longer hang time?



STROBE PHOTOGRAPH of a full-size model and scaled-down paper helicopters in flight

After you read Erik's article, try the experiment yourself, and then let us know what you discover about paper helicopters.

In order to design good experiments to build a better paper helicopter, develop a new drug, or grow bigger shrimp, we need to speak the language of experimentation. Peter Flanagan-Hyde says vocabulary is more than just knowing the words, though. Read his AP Statistics article and find out what he means by that. He gives us some lessons for learning statistics that will help us "take it to the next level."



Numb3rs, Sabermetrics, Joe Jackson, and Steroids

by Jim Albert

"Numb3rs" is a prime time television show, currently in its fourth season, that follows FBI Special Agent Don Eppes and his brother, Charlie. Together, they fight crime and pursue criminals. There are many shows of this type on television; what makes "Numb3rs" notable is that Charlie is a mathematical genius, and his mathematical insights usually are crucial to solving the crimes. The show illustrates the use of mathematics in different ways and throughout many walks of life. Generally supported by the mathematics community, "Numb3rs" sends the positive message that "math is cool."

The "Numb3rs" episode that aired November 10, 2006, titled "Hardball," focused on the use of steroids in baseball. This is a timely subject, as some professional baseball players recently admitted to using illegal human growth hormones to enhance their performance. In particular, many people have claimed that Barry Bonds, Mark McGwire, and Sammy Sosa—great home-run sluggers over the last 10 years—used steroids during their careers. Some think steroids may have contributed to Bonds' record of hitting 73 home runs during the 2001 baseball season.

Sabermetrics

In the "Hardball" episode, a professional baseball player dies during a practice, and it is found that he received a toxic dose of steroids, which suggests murder. When FBI agents explore the victim's recent email messages, they find an interesting attachment containing a long list of equations. They discover the author of the equations is a young "sabermetrician," an analyst of baseball statistics, who claims to have discovered an equation that uses batting statistics to detect players using steroids. Charlie checks out the equation and discovers it can reliably detect the players who use illegal drugs. When Charlie is surprised at this, “Numb3rs” local physics professor Larry Fleinhardt comments:

> It wouldn’t be the first time math revealed a surprising truth about the sport. In a 1993 article in *The American Statistician,* Jay Bennett, using sabermetrics, analyzed ‘Shoeless’ Joe Jackson’s career. He was able to prove that Jackson played up to his full potential in every one of the 1919 ‘Black Sox’ series. ...Math restored a man’s good name and reputation after 70 years. I find that rather beautiful.

“Hardball” raises some interesting questions:

> *How did Jay Bennett statistically show that Joe Jackson did not intentionally try to lose the 1919 World Series?*
>
> *Is it possible to use statistical methods to detect possible steroid use?*

JOE JACKSON (with shoes) on the right and another legendary baseball great, Ty Cobb, on the left

Joe Jackson

In 1919, the Cincinnati Reds upset the Chicago White Sox in the World Series of baseball. A year later, two key Chicago players confessed to participating in a conspiracy to “throw,” or deliberately lose, the 1919 World Series. Eight Chicago players (labeled the “Black Sox”) were tried and found innocent. However, the commissioner of baseball banned all of the suspected players from Major League Baseball for life.

The most famous of the banned players was Jackson, who had the third-highest career batting average (.356) in baseball history. Many fans have raised questions about the 1919 World Series. Is there evidence that the Black Sox did not play to their full potential during the series? The data show that Jackson actually performed well during these games, obtaining 12 hits and a .375 batting average. But, even if he played well overall, perhaps he did not perform well during “clutch” situations in the game.

Bennett, in his *The American Statistician* article, described a statistical method for determining the value of Jackson’s play during the 1919 World Series. Here is how it works: At the beginning of the game, the teams have approximately the same ability, so the probability at the start that one team, say Chicago, will win the game is 0.5. But every play in the game will change the game situation, such as the runners on base, number of outs, and runs scored, which will impact the probability that Chicago will win. One can measure a given player’s contribution to the game by looking at every batting and fielding play in which he was involved and the changes in the game-winning probabilities for all those plays. Using this measure of performance, Bennett found that Jackson was the third most valuable player on his team in contributing toward his team winning the games.

Based on this probability of winning metric and using a resampling technique that redistributed Jackson’s at-bat outcomes with the situations in which they occurred, Bennett developed a distribution of possible batting performances for Jackson in the 1919 World Series. If Jackson’s actual performance appeared at the extreme low end of the distribution, it would support Jackson’s detractors in their assertion that Jackson threw the series. However, Jackson’s performance was in the upper half of the distribution, which indicates he actually performed as well as could be expected with respect to the at-bat situations in which he appeared. In hypothesis-testing terms, the evidence does not support rejecting the null

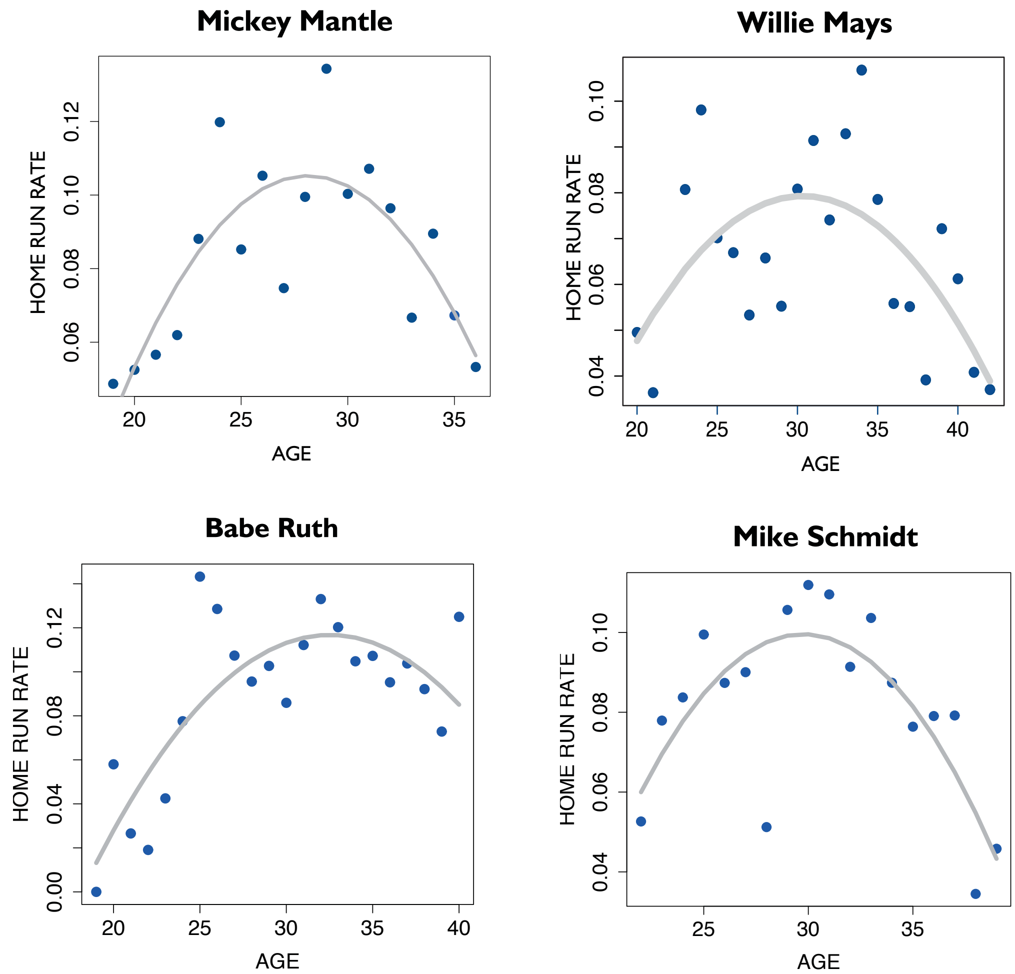


FIGURE 1. Home run rate by age for four all-time great sluggers

hypothesis (Jackson's batting performance showed no adverse clutch effect) in favor of the alternate hypothesis (Jackson did not hit in the clutch).

Bennett has analyzed all the World Series games from 1996 through 2006 using this probability of winning metric. Visit *www.amstat.org/sections/sis* to see the most valuable players (MVPs) in these series according to Bennett's approach. As you might expect, Bennett's objective method of evaluating players' performance to determine the MVP sometimes conflicts with the official MVP chosen by sportswriters.

Steroid Use

Now let's turn to steroid use by professional baseball players. As statistical methodology was useful in assessing Jackson's performance during the 1919 World Series, can it also be used to detect steroid use? For example, suppose a particular player—such as Barry Bonds—exhibits a large increase in hitting home runs for one season. Does this mean Bonds was on steroids? Scott Berry, in a *CHANCE* article titled "A Juiced Analysis," explains that there are many explanations for a jump in home run numbers for a particular player. It could be due to steroids, but it also could be due to the player doing a large amount of training during the off-season or using a new hitting style. It also could just be due to chance variation.

Although it is difficult to say any player is "juiced" based entirely on empirical evidence for a single season, we can look at patterns of hitting over time and for players who exhibit unusual patterns. One useful measure of hitting strength is the home run rate defined as the proportion of home runs hit among all the balls a player hits and puts into play. The home run rate can be written as:

HR RATE = HR / (AB – SO),

where HR is the number of home runs, AB is the number of official at-bats, and SO is the number of strikeouts. If we look at the home run rate for a particular player across all the seasons in his

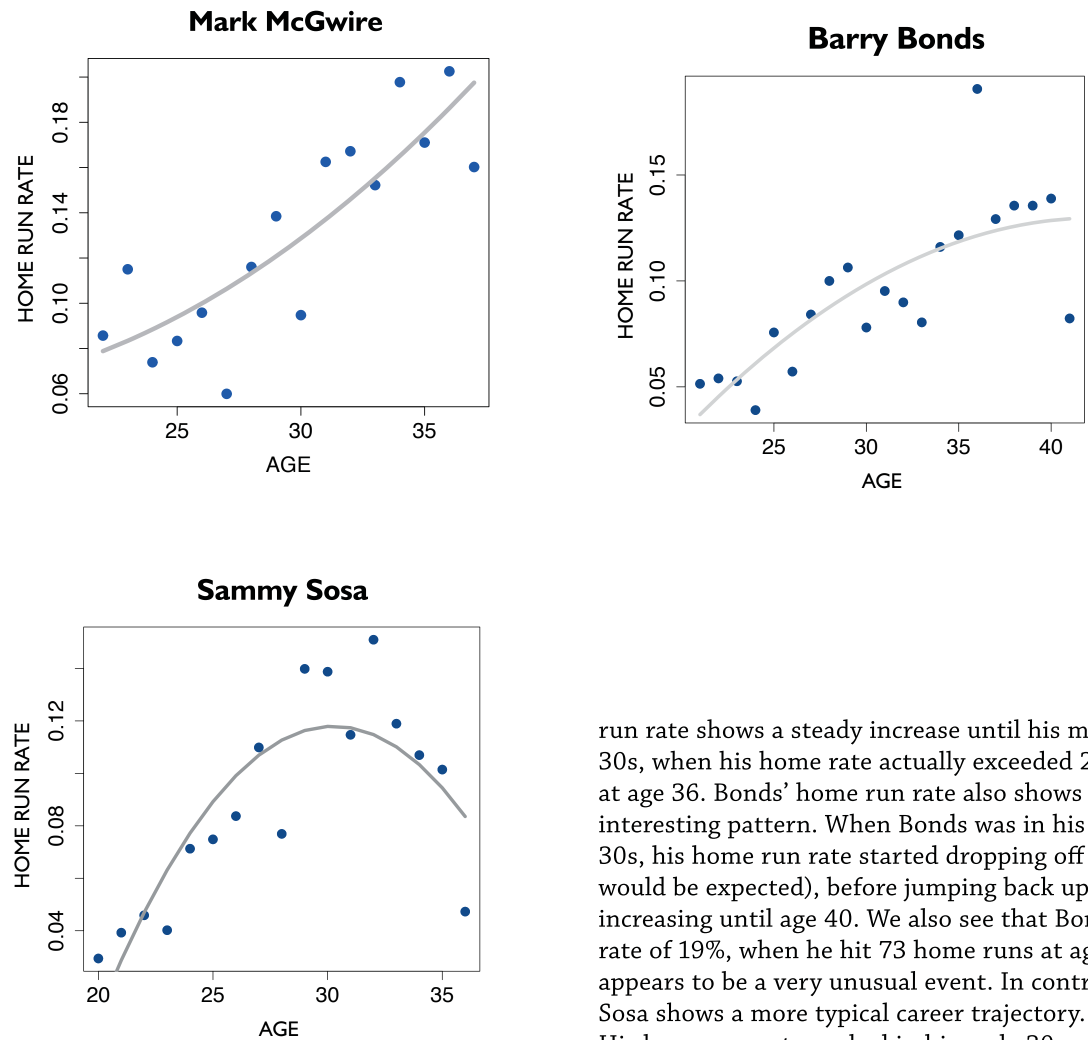


FIGURE 2. Home run rate by age for three recent sluggers

career, a typical pattern emerges. A player's home run rate generally increases until the middle of his career—usually when he is around 30 years old—and then diminishes until retirement. For example, the graphs in Figure 1 plot the home run rates of four historically great sluggers—Babe Ruth, Mickey Mantle, Willie Mays, and Mike Schmidt—with best-fitting quadratic curves placed on top. The career trajectories of most great home run hitters in baseball history show a similar pattern.

Suppose we look at the home run rates in Figure 2 for recent home run leaders McGwire, Bonds, and Sosa, who are suspected of using steroids. The trajectories of McGwire and Bonds show a surprising pattern. McGwire's home run rate shows a steady increase until his mid-30s, when his home rate actually exceeded 20% at age 36. Bonds' home run rate also shows an interesting pattern. When Bonds was in his early 30s, his home run rate started dropping off (as would be expected), before jumping back up and increasing until age 40. We also see that Bonds' rate of 19%, when he hit 73 home runs at age 36, appears to be a very unusual event. In contrast, Sosa shows a more typical career trajectory. His home run rate peaked in his early 30s and dropped off toward the end of his career.

Have we proven that McGwire and Bonds used steroids when they were playing baseball? No. What these graphs show is that McGwire and Bonds have unusual career trajectories, in that they peak in home run hitting at atypically old ages. Few players in baseball history have displayed similar career trajectories. What caused these players to peak at such advanced ages? McGwire's manager, Tony La Russa, responded to the steroid accusation as follows:

> It's fabrication. The product of our good play and strength of our players—Mark was a great example—what we saw was a lot of hard work. And hard work will produce strength gains and size gains.

As many other players also "work hard," statistical analyses such as examining career home run rates may suggest another reason is needed to explain these patterns of behavior. ◕

Learning Stats Is FUN …with the Right Mode

by Lawrence M. Lesser

At one time or another, you may have felt "statistics anxiety" and declared statistics to be part of the "inhumanities," considering its focus on "mean" things. But maybe with the right "mode," we can add some fun to learning statistics in order to make chi-square "chi-cool," turn those factorial signs into exciting exclamation points, and not fall asleep while studying Zs.

Humor and Song

As the word "median" is embedded in the word "comedian," let's start with some statistics jokes you can share with your statistics friends (the punch lines are on Page 21):

❶ How many statisticians does it take to change a light bulb?

❷ Why did the statistician cross the road?

❸ What are the last available graves at a cemetery?

❹ Why did a statistician bring a bomb on the flight? And, what if that jet-setting statistician was thrown in jail?

❺ What is a Western Union message sent by a snake?

❻ Speaking of snakes, what is the most statistical snake?

Another way to make statistics fun is through song. For example, I have written fun songs ranging from a parody about a not-yet-informed lottery player sung to the tune of "The Gambler" to a break-up song I call "Statistician's BLUEs" (remember **b**est **l**inear **u**nbiased **e**stimators), stuffed with two dozen statistics puns. Many people have been moved to write fun songs about statistics, too, and I helped assemble some of their songs into a searchable database on the CAUSE (Consortium for the Advancement of Undergraduate Statistics Education) web site (*www.causeweb.org)* for all to enjoy. You can look up the gambler and BLUEs songs there. The CAUSEweb database also contains loads of cartoons, jokes, and quotations that lend themselves to a lot of fun.

Besides being fun, rhyme and rhythm can be helpful learning aids because they make it easier to recall statistical ideas. Psychology professor Carmen Wilson VanVoorhis researched the effectiveness of using jingles as mnemonic devices. While teaching two sections of students who had grade-point averages that were statistically equivalent, VanVoorhis had three statistics definitions read aloud in one section and statistics jingles sung for the same concepts in the other section. On four short-answer test items, the students in the jingles section did better ($t_{69} = 2.01$, p-value $< .05$) and had a statistically significant correlation ($r_{31} = .37$, p-value $< .05$) between test scores and self-rated familiarity with the jingles.

You might enjoy writing your own jingles for statistical ideas. For example, I wrote "What *P*-Value Means," to be sung to the tune of "Row, Row, Row Your Boat."

It is key to know what *p*-value means.
It's the chance,
With the null,
You obtain

Data that's at least that extreme.

Editor's Note: The References and Additional Reading List starting on Page 26 is an annotated listing of all of the people mentioned in this article and a great resource for ideas about the fun 'mode' for learning statistics.

Books

When you need a break from studying statistics content in "mathematized" theoretical form, you might enjoy reading books that use bits of humor to explain big ideas. Try *Statistics with a Sense of Humor; The Cartoon Guide to Statistics;* or *Winning with Statistics: a Painless First Look at Numbers, Ratios, Percentages, Means, and Inference.*

Or, look at studies with realistic statistics delivered in an absurd or satirical context, such as certain articles in *Annals of Improbable Research* or *The Journal of Irreproducible Results*. This latter journal recently sponsored a "funniest graph" contest. What would your entry have been? Perhaps a pie chart of how you and your friends shared some pie á la 'mode'?

Games

Playing games with probabilistic structure provides a way to have fun while thinking about and discussing statistical ideas. It also can give you an edge over your less-statistical friends. For example, statistics can be used to determine good strategies for playing classic board games, such as Monopoly. Steve Abbott and Matt Richey, mathematics professors, used Markov chain models to determine the spaces on which players most frequently landed, and then calculated expected value per roll and time to break-even by property color groups with hotels. They found, "The orange has far and away the lowest break-even time, owing to its high frequency and its relatively low building costs. Enthusiasm for the green group's high expected return per roll must be tempered by the knowledge that this monopoly may not yield profits for many rolls. [And] it is hardly worth the trouble to purchase the feeble purples (Baltic and Mediterranean)…"

Medical researcher Phil Woodward has shown that Yahtzee has more than 1 trillion outcomes, which is why it is not trivial to find the optimal choices of which dice to re-roll and which scoring group to use. As an example of a nonobvious scoring choice, he asks, "What should the dice combinations 4-4-6-6-6 and 5-5-6-6-6 be scored against if obtained at the end of the first turn?"

At the 2006 Joint Statistical Meetings in Seattle, statistics professor Paul Stephenson described GOLO, a "golf dice" game, with a probability-informed, multiple-rolls strategy. And at the 2005 Joint Statistical Meetings, Mary Richardson, a statistics professor, and David Coffey, a mathematics professor, asked, "What is the probability of pigging out?" They looked at strategies for estimating the outcome probabilities associated with rolling the nonstandard, pig-shaped dice of Pass the Pigs. They also looked at the implications on strategy for ending a turn, considering that one particular roll outcome can wipe out all the points on a player's current turn. A related dice game, HOG, was studied by mathematics professors Larry Feldman and Fred Morgan, who also were looking for the 'best' strategy. So, you can go low, go hog-wild, or just pig out—all the while having statistical fun.

Edward Packel, a mathematics professor who studied backgammon, explains how probability can be used to support the following example of backgammon strategy: "If you must leave a blot [that] you do not want hit, leave it at least seven points from the threatening man, if possible, in which case, the farther, the better (with 11 versus 12 as the lone exception). If you must leave the blot within six points, move it as close as possible to the threat man." This would be difficult to figure out without using statistical analysis.

A multicultural example is the Jewish Hanukkah game of Dreidl, played with a four-sided top. This game can yield increasingly sophisticated analyses, such as finding the expected payoff on the *n*th spin with *N* players or what rule changes would make the game 'fair.' Felicia Trachtenberg—mathematician, statistician, and Dreidl enthusiast—and Robert Feinerman, who is also a mathematics professor, looked at how the game can be fair and unfair.

In card games such as bridge, blackjack, and poker, there are many places for knowledge of probability to inform strategy. There are particularly interesting situations, such as that attributed to World Poker Champion Amarillo Slim, who allows an opponent to choose Hand 1 (2♣, 2♠), Hand 2 (A♠, K♦), or Hand 3 (J♥, 10♥). Woodward explains, "Slim was prepared to let you choose any of these hands. He would then choose a different one. ...He had the odds on his side because Hand 1 beats Hand 2 about 52% of the time, Hand 2 beats Hand 3 with probability around 59%, and Hand 3 will beat Hand 1 on 53% of matchups." He also notes that this phenomenon occurs in other two-player games, such as predicting what coin-tossing sequence of a given length will occur first.

Did you know there actually are magic tricks based on probability? For example, the famous "Kruskal count" card trick is an application of linear algebra and Markov processes. Ivars Peterson, an author and editor, notes, "The underlying mechanics of the Kruskal count also highlights how seemingly unrelated chains of events can lock together in sync after a while—a phenomenon worth watching for in other contexts."

Game Shows

Televised game shows, such as Wheel of Fortune or Who Wants to Be a Millionaire?, can provide another source of statistical fun. In Wheel of Fortune, contestants often must decide whether to risk bankruptcy by spinning for more money when they know the puzzle's solution. Who Wants to Be a Millionaire? yields two types of strategy, explored by education professor Robert Quinn: (1) how much guessing in the preliminary Fastest Finger question may maximize the chance of being the contestant with the correct four-item sequence in the fastest time and (2) the expected value of guessing when you do not know an answer once you are in the hot seat pursuing prize money. Also, during the 2006 Joint Statistical Meetings, mathematics professor Diane Evans explained the strategy for how to maximize the benefit of that "50–50 lifeline" based on the initial assignment of probability estimates to the answer choices and whether the game show uses random or nonrandom elimination of choices.

> "Learning and life are maximal where play and work coincide."
> *L. W. Gibbs*

Of course, the decision of whether to keep one's original choice after one or more options have been eliminated brings to mind the Let's Make a Deal game show's "Monty Hall" problem (a.k.a the Three Doors problem or the Car and Goats problem). It has generated a huge number of articles. Statistics professor Federico O'Reilly's article, "A Look at the Two-Envelope Paradox," published in the Spring 2006 issue of *STATS*, is a good example of another decision scenario involving conditional probability.

The Price is Right game called Plinko involves trying to choose the best place(s) to release a chip at the top of a vertical board filled with rows of pegs so the chip has the greatest chance of finishing the maze in a favorable slot at the bottom of the board. Obviously, this game is similar to Sir Francis Galton's quincunx device, in which beads dropped at the center form a bell-shaped curve. Connections between Plinko and learning statistics have been made by mathematicians Amy Biesterfeld, LaDawn Haws, Susie Lanier, and Sharon Barrs. Biesterfeld has looked at strategy for other The Price is Right games involving probability, such as Range Game or Master Key. Eric Wood, another mathematician, has used probability to decide whether to take a second spin of the giant wheel (containing multiples of five up to 100 in scrambled order) in the multiplayer competition. The goal is to get the highest score without exceeding 100 and earning a spot in the Showcase round.

A current popular game show is Deal or No Deal, in which players try to maximize the deal they get for their unopened briefcase, taking into account its expected value from a set of known monetary amounts ranging from $.01 to $1 million in light of 25 other unopened briefcases.

As one last example, Friend or Foe was analyzed to see whether age, gender, race, and the amount of prize money affect contestants' strategies.

By learning statistics, you will be ready to win big prizes if you are picked for a game show. The possibilities seem endless.

Internet

Speaking of the seemingly endless, try searching the internet. Statistics professors Webster West and Todd Ogden found and catalogued online interactive demonstrations and games for learning statistics. You can find applets to make compositions from "Mozart's Musical Dice Game." A Google search will turn up musical performing groups with names such as "Statistics" and "Random," as well as web sites that randomly generate songs or band names. There are even statistical "rap" videos on the internet. See Page 27 for some sites I found. I am sure you can find many more.

Creative Works

Another way to have fun is to look for examples of statistics in real-world creative works. There are examples of statistics concepts and vocabulary in popular song lyrics, such as in this one from singer-songwriter David Wilcox:

"So you come for a visit, but you lose your nerve out at the edges of the bell-shaped curve."

There are also poems that use statistical images or words. One of these is Wislawa Szymborska's "A Word on Statistics." One of the best-known short stories ever is Shirley Jackson's "The Lottery," which first appeared in 1948 in *The New Yorker* and has since appeared in many compilations. The story's vehicle is an annual village ritual that includes making sampling frames and conducting a sequence of simple random samples.

The classic novel *The Phantom Tollbooth*, by Norton Juster, has some entertaining passages concerning averages in Chapter 16, such as when lead character Milo encounters "one half of a small child who had been divided neatly from top to bottom":

> What is the rest of your family like?" said Milo, this time a bit more sympathetically. "Oh, we're just the average family," he said thoughtfully, "mother, father, and 2.58 children, and, as I explained, I'm the .58.
>
> It must be rather odd being only part of a person," Milo remarked. "Not at all," said the child. "Every average family has 2.58 children, so I always have someone to play with. Besides, each family also has an average of 1.3 automobiles, and since I'm the only one who can drive three tenths of a car, I get to use it all the time.

Look around and I am sure you will find many more examples of statistics in popular songs, stories, and movies.

More Fun

Another diversion is coming up with your own creative activities that can be subjected to hypothesis testing. One of mine is to test the null hypothesis that I cannot juggle three balls, using a value of five seconds for the null mean amount of time before one of the balls hits the ground. Speaking of round things, next time you tear open a bag of m&ms, use a chi-square test to compare the observed color distribution to the official distribution of that particular variety on the company's web site at *http://us.mms.com/us/about/products*.

Another of my favorites is to apply hypothesis testing to everyday events or subjects, such as popular songs. Do hit songs have a mean length of 3.5 minutes? Are they usually cowritten? Is there a correlation between a song's tempo and the length of a song's introduction?

It also can be fun finding and analyzing silly statements about correlation or causality, ranging from birth rates and stork counts to the Super Bowl and Dow Jones. At the 2006 Joint Statistical Meetings, Harry Norton shared the following example:

> DAY ONE: Made loud noise behind frog. Frog jumped 15 feet.
>
> DAY TWO: Immobilized one hind leg of frog, then made same loud noise as on day one. Frog jumped only 3 feet.
>
> DAY THREE: Immobilized both hind legs of frog, then made many loud noises—louder than days one and two. Frog did not jump at all.
>
> CONCLUSION: When both hind legs of a frog are immobilized, it becomes deaf.

The next time you are on a long road trip (perhaps driving to the Joint Statistical Meetings), think of all the opportunities for statistical fun. The "license plate game" can be played with statistical variations (e.g., first person to spot a license plate whose digits yield particular values for the mean, median, mode, or range…). You also might hunt for statistical terminology along the roadside (e.g., Look, a park for 'random variables'—RVs!). Another modified car game is to see how long you can take turns giving the name of a statistical term (or statistician) beginning with the letter of which the previous word ended.

You also can play "free association" word games. My graduate advisor, famous for his empirical Bayes research, often would throw me and my fellow members of S.O.S.—Students of Statistics—a pair of nonstatistical words to see if we could guess which word he was associating with "Bayesian" and which one with "frequentist." We correctly guessed almost every time, to our advisor's surprise.

With statistics, you even can have fun with the clothes you wear (as long as you avoid "lack of fit"). Some of the current ASA T-shirt captions include: "I'm Statistically Significant," "Statisticians Are Spatial," and "Approximately Normal." Get some of your stats friends together and make some of your own T-shirts with stats slogans.

You also can explore paradoxes or statistical situations with striking results. Eric Sowey has shown how surprises are fun and help us learn. Test yourself with the following questions (The answers are on Page 21.):

① If the standard normal distribution is drawn to scale on a sheet of paper so its tails are 1 mm above the axis at z = ± 6, how high must the paper be at the mode to accommodate the drawing?

② How many people must be in the room for there to be at least a 50% chance of at least two people sharing a birthday? This problem was featured in the Spring 2005 issue of *STATS*.

③ There are three cards in a bag. Both sides of one card are green, both sides of one card are blue, and one side is green and one side is blue on the other card. You pull a card out of the bag at random and see that one side is blue. What is the probability that the other side is also blue?

Let's Party

For a couple of decades now, museums and schools throughout the country have celebrated Pi Day every March 14 (3/14). Maybe it's time to create a statistics counterpart to this special day. Perhaps statistics students could celebrate "Happy Median Day" every July 2, enjoying a serving of pie á la 'median'! Or maybe every January 10 and December 22 could be "95% Confidence Days." Suppose we celebrate the birthdays of one or more famous statisticians. Here is a list of 10 to get your started:

JANUARY 13 – GERTRUDE COX
FEBRUARY 16 – SIR FRANCIS GALTON
FEBRUARY 17 – RONALD FISHER
MARCH 27 – KARL PEARSON
MAY 12 – FLORENCE NIGHTINGALE
JUNE 13 – WILLIAM GOSSET
JUNE 16 – JOHN TUKEY
JULY 15 – WILLIAM COCHRAN
OCTOBER 14 – W. EDWARDS DEMING
NOVEMBER 18 – GEORGE GALLUP

Gosset's birthday could be celebrated with a 'students' tea,' making sure the tea is served according to "students' tea distribution." And since no one seems sure what day in 1702 Thomas Bayes was born in London, why not pick a date you find convenient for a "Bayes Day" celebration? Check out the web sites in the references section, where you can look up more birthdays of famous people.

By the way, did you know the second full week of February has been designated Random Acts of Kindness Week? Well, how about picking your own week at random for "Kind Acts of Randomness Week"? Then, get some of your statistics friends together to celebrate by doing a service project for someone. Nothing is more fun than making the world a better place. Have some fun and pass it on.

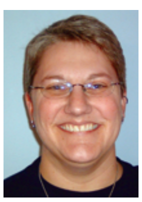
Learning Statistics Can Be Fun

I hope you are now inspired to find your own ways of bringing fun into your statistical journeys, and I hope you have a sense of the benefits and huge variety of options for having fun with statistics—without unduly regressing or deviating! All we need is the right 'mode.' ◕

What Is This Statistics Education STUFF?

JOAN GARFIELD is professor of educational psychology at the University of Minnesota, where she heads a graduate program in statistics education. She serves as associate director of research for the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) and is involved in many national and international statistics education committees. She is coeditor of *The Assessment Challenge in Statistics Education* and *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Currently, she is coauthoring a book that connects research to practical teaching advice. *jbg@umn.edu*

by Jackie Miller



JACKIE MILLER 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. When not at school, Miller enjoys a regular life (despite what her students might think), including keeping up with her many dogs!

As someone with a one-of-a-kind PhD in statistics education from The Ohio State University, I am very interested in statistics education. The statistics education community has been growing and developing over the past two decades. Joan Garfield of the University of Minnesota is a mover and shaker in the statistics education community and someone I consider a mentor. In the hope of getting some of you interested in statistics education, I asked Garfield to answer the following questions.

How is statistics education different from other disciplines, such as mathematics education?

I see most of the focus in mathematics education being [put] on teaching and learning mathematics in kindergarten through high school (K–12) classrooms, or on the preparation and professional development of K–12 mathematics teachers. In contrast, most of the focus in statistics education is on the teaching and learning of statistics in college-level classrooms. While most of the scholarship in mathematics education is by mathematics education professionals who do not actually teach mathematics, but rather teach teachers to teach mathematics, most of the scholarship is by people who teach statistics and are concerned about improving the teaching and learning of statistics in statistics education.

Recently, there has been a growing interest among mathematics educators in statistics education in K–12 settings, and on the preparation of K–12 teachers to teach statistics. However, while there are courses to prepare teachers to teach mathematics in elementary and secondary schools, there is little, if any, formal preparation of teachers to teach statistics at any level.

Is there a different pedagogy that comes with the teaching and learning of statistics? If so, why?

Statistics is a unique discipline, as people such as David Moore and George Cobb have so eloquently written. Statistics is *not* mathematics, although mathematics is used in statistics, just as mathematics is used in other disciplines, such as physics or economics. I believe there are unique methods for teaching statistics, and that it should *not* be taught as a mathematics course. The recommended methods for teaching statistics include:

- Using real data
- Focusing on concepts, rather than computations
- Using statistical software and technological tools
- Involving students in doing statistics, rather than just learning about how to do statistics

One of the pedagogical methods used by many in statistics education is to utilize simulation tools to help students visualize abstract statistical concepts and procedures, such as sampling distributions and the Central Limit Theorem.

Currently, there is a set of guidelines on teaching the introductory statistics course, as well as statistics in K–12 classes, that has been endorsed by the American Statistical Association: the *Guidelines for Assessment and Instruction in Statistics Education (GAISE*). We hope these guidelines will lead to high-quality statistics courses that will help students develop an understanding and appreciation of statistical investigations.

What is the purpose or objective in having a statistics education graduate program?

A graduate program in statistics education, such as the one I direct at the University of Minnesota, is to both prepare excellent teachers of statistics at the postsecondary level and to prepare graduate students to conduct research on the teaching and learning of statistics. To become an excellent teacher of statistics, graduate students need to take at least one course in statistics education. In such a course, they become familiar with the field of statistics education: the current guidelines, important readings, resources, teaching methods, and technological tools. Graduate students also need to learn about good assessment practices and resources.

In addition, while statisticians are trained in quantitative statistical methods, few of these methods apply to conducting research focused on the teaching and learning of statistics in college classrooms. This type of research requires an understanding of current theories of learning so that we can design and conduct studies with a strong theoretical background. Qualitative methods, educational measurement, and a solid background in foundational research in statistics education also are important and part of the coursework in a graduate program in statistics education.

Another purpose of a graduate program is to provide careful and high-quality training and supervision in teaching as graduate students grow and develop into teachers of statistics.

What results have we seen from graduates with degrees in statistics education?

In the past, a graduate student wanting to focus their dissertation research on statistics education had to do this on their own, as part of their course work in existing programs, such as statistics, psychology, or mathematics education. The graduate program in statistics education at the University of Minnesota was launched four years ago, and the first student to complete a doctorate in the program graduated in May 2006. He was hired for a one-year lecturer position at the University of Minnesota and then planned to apply for academic jobs at the college or university level. I expect he will play an active role in the statistics education community and contribute to the research and scholarship in this area. I have two other students in the program, and a new one began in Fall 2006 who I hope will follow a similar path.

> **"Statistics is *not* mathematics."**

How many programs are there, and how many current statistics education PhDs are there from these programs?

As far as I know, my program is the only stand-alone graduate program in statistics education. At Ohio State, there are several students working on doctorates in mathematics education with a focus on teaching statistics at the college level. Also, and new this year, there is a graduate education minor in graduate interdisciplinary specialization in college and university teaching. At the University of Georgia, graduate students in mathematics education can focus on the area of statistics, but I think it is still directed to K–12 settings.

What is the statistics education community like?

I think the statistics education community is a wonderful, vibrant, exciting community. There are many bright, enthusiastic, energetic people who have dedicated their lives to teaching students and to improving students' experiences in statistics classes. There are many wonderful texts, tools, and other resources available today, so there is no excuse for teaching an old-fashioned, traditional course that consists of a teacher lecturing to students and demonstrating how to work out formulas and compute answers while students listen passively and take notes. There is no excuse for teaching a course that does not use real data, real examples, and real statistical tools. These resources are all widely and freely accessible on the internet and elsewhere.

Today, we are fortunate to have *CAUSEweb.org* —the web site for CAUSE—that provides a one-stop site to locate valuable information about current activities, materials, and research.

Is there anything else you would like to tell students of statistics about statistics education?

I encourage all students interested in pursuing a career in statistics education to become familiar with CAUSE and *CAUSEweb.org* and to attend the United States Conference on Teaching Statistics (USCOTS) held at The Ohio State University in May 2007. This conference is held biannually and showcases some of the great thinkers and contributors to the field and offers an opportunity to network with graduate students and statistics teachers at all levels. It is important to become part of the statistics education community and not work in isolation. I welcome any readers with further questions to read about my graduate program at *www.education.umn.edu/EdPsych/QME/stats-intro.html* or contact me at *jbg@umn.edu*. ◕

Designing a Better Paper Helicopter
USING RESPONSE SURFACE METHODOLOGY

by Erik Barry Erhardt

Suppose a group of your friends are having a contest to design a paper helicopter that remains aloft the longest when dropped from a certain height. To be fair, everyone starts with the helicopter pattern given in Figure 1, which is an easy pattern to modify and replicate. Armed with knowledge about response surface methodology and a desire to strive for excellence, you could have an advantage. Let us go through the steps together, and I will show you how my classmate, Hantao Mai, and I designed "a better paper helicopter" to become two-time paper helicopter champions at Worcester Polytechnic Institute. After you see what we did, you can try to do even better.

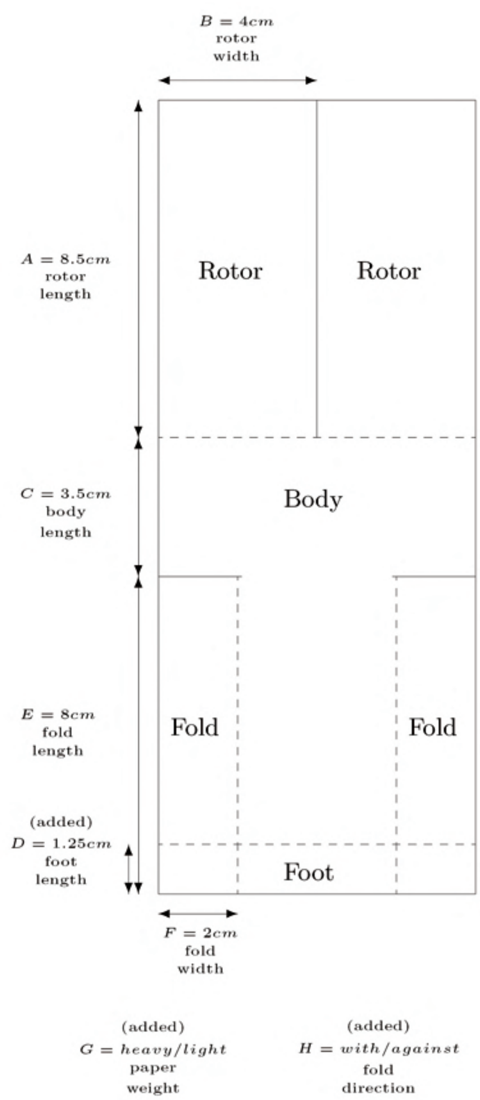


FIGURE 1. Initial helicopter pattern. Cut along the solid lines and fold along the dotted lines. The foot fold, *D*, paper weight, *G*, and fold direction, *H*, were not included as part of the initial pattern, but were added after brainstorming about potential factors that might influence flight time.

Definitions of highlighted words can be found on Page 20

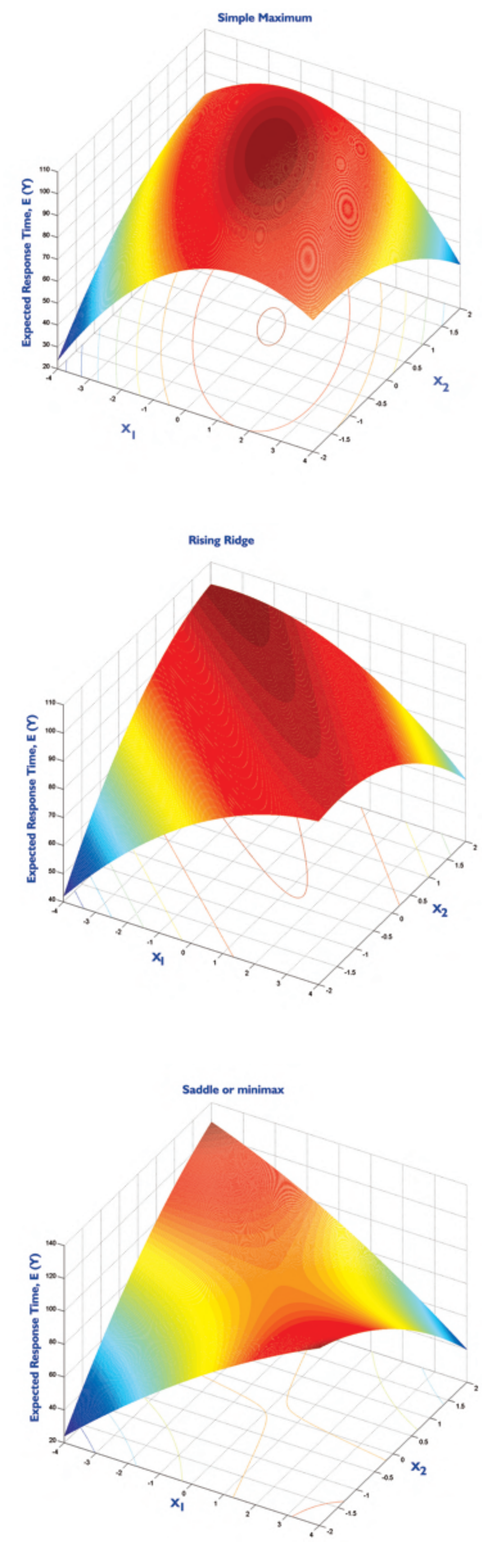


FIGURE 2. Examples of response surfaces in three dimensions: a maximum, a rising ridge, and a "saddle" with a maximum in one direction and a minimum in a transecting direction

Response Surface Methodology

Response surface methodology (RSM) is a collection of statistical and mathematical techniques to explore efficiently the performance of a system to find ways to improve it. A good reference on RSM is *Response Surface Methodology* by Raymond Myers and Douglas Montgomery. A response surface can be envisioned as a curved surface representing how the system's output performance (a dependent variable) is affected by specified input factors (independent variables). Examples of response surfaces in three-dimensional space are shown in Figure 2.

George Box is the statistician credited with first proposing the ideas behind RSM more than 50 years ago. The major tools in RSM are design of experiments, multiple regression, and optimization.

In the paper helicopter project, our goal was to use these tools to find the combination of design factors that would maximize flight time. We designed a set of experiments that allowed us to discover enough about the shape of the response surface to design the winning paper helicopter—twice!

Design of Experiments

Brainstorming

An important question to start with is, "What factors should we test?" A brainstorming session can reveal many factors that might influence the response. We brainstormed about helicopter design using our knowledge of the aerodynamics of flight, and tried to think of everything about the design of a paper helicopter that we could control. Then, we made a few helicopters, dropped them, and made some modifications—a fun way to start a project.

Factorial Designs

In designing experiments, the workhorses of experimental designs are factorial designs. These are efficient designs for assessing the effects of several factors on a response. A 2^k design is a factorial design for k factors, each tested at two levels, coded as +1 for "high" and –1 for "low." The low and high levels are selected to span the range of possible values for each factor. A graphical representation of a 2^4 design is shown in Figure 3. The cube shows the orthogonal relationship of the experimental points.

Center points are experimental runs with all factors at level 0 midway between –1 and +1. Adding center points to the experiment is a logical way of checking curvature and obtaining an estimate of the error variance.

Pareto Principle and Factor Screening

The Pareto Principle says there are a "vital few" important factors and a "trivial many" less important factors.

A screening experiment can distinguish the trivial variables from the vital few important factors influencing the response. The

SPOTLIGHT

The concepts behind optimization through experimentation, now known as response surface methodology, were developed by George E. P. Box (a statistician) and K. B Wilson (a chemist). Box is emeritus professor at the University of Wisconsin-Madison, where he cofounded the Center for Quality and Productivity Improvement. His work has encompassed design of experiments, quality control, time series analysis, and Bayesian inference. It is his name in "Box-Jenkins time series models," "Box-Cox transformation," and "Box-Behnken experimental designs."

The idea for the paper helicopter experiment originated with C. B. "Kip" Rogers at Digital Equipment Corporation. It was popularized by Box as a way to teach experimental design for quality improvement to engineers. Box is coauthor with J. Stuart Hunter and William G. Hunter of *Statistics for Experimenters: Design, Innovation, and Discovery*, which features the paper helicopter experiment in Chapter 12.

SPOTLIGHT

Vilfredo Pareto (1848–1923) was an Italian economist, sociologist, and engineer. In statistics, his name is associated with the Pareto Chart and Pareto Distribution. He is best-known for observing that 20% of the population owned 80% of the property in Italy. Experience has shown that a similar relationship is common in many other situations in the social, biological, and physical sciences. Consequently, Pareto's observation has been generalized as the Pareto Principle. Also know as the "80–20 Rule," it says that "80% of the consequences come from 20% of the causes." The rule has many applications in quality control and design of experiments. For instance, a Pareto Chart is a histogram showing the frequency of occurrence of events with the categories ordered from most frequent to least frequent. When studying the performance of a system, a Pareto Chart can help identify the important factors (the 20% or "vital few") among the less important factors (the 80% or "trivial many"). Based on this same idea, the Pareto Distribution has been used to describe the distribution of incomes, population sizes, particle sizes, word frequencies, customer complaints, and many other empirical phenomena.

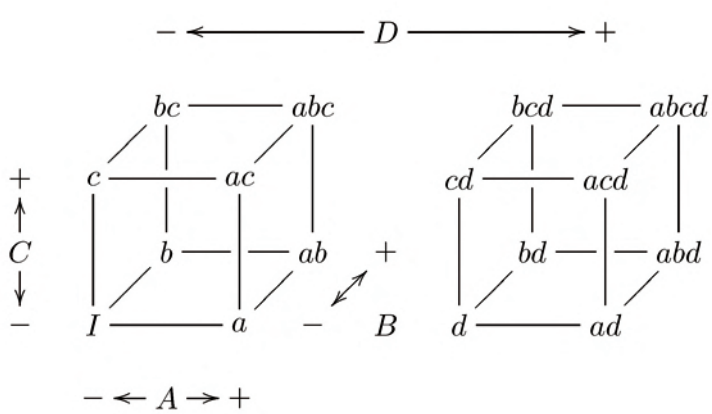


FIGURE 3. A 2^4 factorial design consists of all combinations of four factors, each taking two levels (+1, –1), represented here with a letter for +1 and without a letter for –1. For example, point *abd* in the figure corresponds to factors A, B, and D set to high level +1 and C to low level –1.

dimensionality of the experimental region often is greatly reduced, exponentially decreasing the number of experimental runs required.

Fractional Factorial Designs

A full factorial design contains all possible combinations of the k factors at the tested levels. When the number of factors is large, a 2^k design will have a large number of runs. For example, $k = 8$ requires $2^8 = 256$ runs. Fractional factorial designs can be used for screening to find the factors that contribute most to the response, if a full factorial design requires a prohibitive number of runs. Fractional factorial designs are designated as 2^{k-p}, with $p = 1$ for a design with one-half the full factorial runs, $p = 2$ for one-fourth the full factorial runs, etc.

The price paid for the efficiency of a 2^{k-p} design is that ambiguity is introduced through confounding of the effects. In a useful screening design, the main effects will not be confounded with each other, although there can be confounding with the interaction effects. Because the main effects are not confounded with each other, their relative importance can be distinguished among each other, but not from certain interactions.

Our Screening Experiment

The goal during a screening experiment is to identify those factors with the most influence on the response, and then eliminate the remaining factors from further analysis. To determine the vital few factors that contribute to flight time, we performed a screening experiment using a 2^{8-4} fractional factorial design with the initial helicopter pattern as our center point.

We selected eight factors possibly influencing the flight time of the helicopter: the original five factors plus three that we added (i.e., foot length, paper weight, and fold direction). Paper weight levels were phone book white pages paper light (–1) and standard copy paper heavy (+1). Fold direction levels indicated the fold direction was against (opposite) or with the direction of rotation. Table 1 shows the coded and uncoded values for the eight factors we investigated.

TABLE 1. Factors potentially influencing paper helicopter flight time and the factor levels for the screening experiment in actual units for each coded value. Length and width were measured in centimeters.

	Factors	Coded Values –1	0	+1
A	= Rotor Length	5.5	8.5	11.5
B	= Rotor Width	3.0	4.0	5.0
C	= Body Length	1.5	3.5	5.5
D	= Foot Length	0.0	1.25	2.5
E	= Fold Length	5.0	8.0	11.0
F	= Fold Width	1.5	2.0	2.5
G	= Paper Weight	light	(none)	heavy
H	= Fold Direction	against	(none)	with

We produced 16 helicopters at the design points listed in Table 2. Then, we dropped them in random order from 20 feet and recorded the flight time of each. All flights were conducted indoors in still air.

We used the SAS macro EFFECTS from *Applied Statistics for Engineers and Scientists* by J. D. Petruccelli, B. Nandram, and M. Chen to compute the factor effect estimates, as shown in Table 3.

TABLE 2. Screening experiment with coded factor levels and center points

Factors and Coded Levels

Run	Order	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	Flight Time
1	12	–1	–1	–1	–1	–1	–1	–1	–1	11.80
2	7	–1	–1	–1	1	1	1	1	–1	8.29
3	11	–1	–1	1	–1	1	1	–1	1	9.00
4	15	–1	–1	1	1	–1	–1	1	1	7.21
5	1	–1	1	–1	–1	1	–1	1	1	6.65
6	4	–1	1	–1	1	–1	1	–1	1	10.26
7	16	–1	1	1	–1	–1	1	1	–1	7.98
8	8	–1	1	1	1	1	–1	–1	–1	8.06
9	3	1	–1	–1	–1	–1	1	1	1	9.20
10	10	1	–1	–1	1	1	–1	–1	1	19.35
11	9	1	–1	1	–1	1	–1	1	–1	12.08
12	5	1	–1	1	1	–1	1	–1	–1	20.50
13	6	1	1	–1	–1	1	1	–1	–1	13.58
14	13	1	1	–1	1	–1	–1	1	–1	7.47
15	14	1	1	1	–1	–1	–1	–1	1	9.79
16	2	1	1	1	1	1	1	1	1	9.20

With only a single measurement for each combination of factor levels in the screening design, we could not estimate the error, as all the degrees-of-freedom would be used for estimating the factor effects. A Pareto Chart of the absolute value of the effects (see Figure 4) shows that the main effects *A*, *B*, and *G* should be considered the vital few, as they are more than twice as large as the next largest main effect (i.e., *D*). (The references on Page 28 by Cuthbert Daniel and Russell V. Lenth discuss other ways of analyzing data from unreplicated factorial designs.)

We were left with three main effects, but by inspection, we could see that $G = -1$ (the light paper)

TABLE 3. Factor effects estimated with SAS macro EFFECTS. Factors *A*, *B*, and *G* are indicated as the vital few.

Factor	Estimate	
A	3.9900	*
B	–3.0550	*
C	–0.3475	
D	1.2825	
E	0.2500	
F	0.7000	
G	–4.2825	*
H	–1.1375	
AB	–2.2175	
AC	0.8400	
AD	1.6850	
BC	–0.3850	
BD	–2.0350	
CD	0.2475	
ABCD	1.5625	

always gave a higher response. Also, we fixed *H* = –1 (fold against), as we observed that this setting gave more stable helicopters. Further testing suggested favorable levels of body length *C* = 2, foot length *D* = 2, fold length *E* = 6, and fold width *F* = 2 in their uncoded units of centimeters. So, we had only two factors to consider (*A*, rotor length, and *B*, rotor width) to start our model-building process.

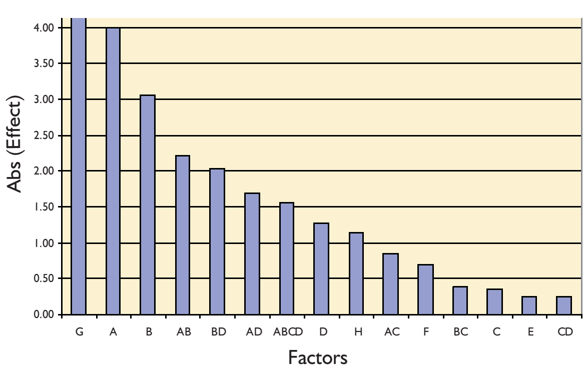


FIGURE 4. Pareto Chart of the absolute value of the effects, in order from largest to smallest. The paper weight, rotor length, and rotor width stand out as the important factors.

Multiple Regression

Regression Modeling

In building a regression model, we assume the response of flight time, *y*, depends upon *k* controllable factors, such as rotor length and width, $\xi_1,\ldots,\xi_k$, in $y = f(\xi_1,\ldots,\xi_k)+\varepsilon$, where *f* is an unknown function and ε is random error with expected value $E(\varepsilon)=0$ and common variance $Var(\varepsilon)=\sigma^2$. The response surface is the expected flight time for varying factor values defined by $E(y)=f(\xi_1,\ldots,\xi_k)$. For convenience, the uncoded variables, $\xi_1,\ldots,\xi_k$, in their original units, are transformed to the coded variables, $x_1,\ldots,x_k$, with zero mean $\mu_i=0$. The response surface is in terms of the coded variables $E(y)=f(x_1,\ldots,x_k)$.

Because we do not know the form of *f*, we need to approximate it. For this purpose, second-order polynomials are useful. A second-order polynomial has the form:

$$E(y)=\beta_0+\sum_{i=1}^{k}\beta_i x_i+\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\beta_{ij}x_i x_j+\sum_{i=1}^{k}\beta_{ii}x_{ii}^2$$

The first-order terms describe a plane, while the cross-product terms (also called interaction terms) and the second-order terms describe curvature.

With only two factors, it is practical to use a 2^2 replicated design. New measurements were taken for all helicopters, as shown in Table 4.

TABLE 4. The results of a 2^2 design with replication to determine the initial estimate of the response surface

Obs	A	B	Time
1	–1	–1	10.24
2	–1	–1	9.11
3	1	–1	16.52
4	1	–1	16.99
5	–1	1	10.20
6	–1	1	9.26
7	1	1	10.02
8	1	1	9.94
9	–1	–1	11.31
10	–1	–1	10.94
11	1	–1	12.58
12	1	–1	13.86
13	–1	1	8.20
14	–1	1	9.92
15	1	1	9.95
16	1	1	9.93

Our Initial Model

We applied multiple regression to the results in Table 4 to determine our initial model to estimate the response surface as:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$
$$= 11.1163 + 1.2881 x_1 - 1.5081 x_2$$

Optimization

Steepest Ascent

What is the shortest path to the top of a hill? Always take the steepest possible step. The goal is to move the process variables, x, to a new region of improved response by following the path of steepest ascent. The procedure is to compute the path of steepest ascent using the first-order model and conduct experiments along this path until a local maximum is reached. The direction and magnitude of steepest ascent is the gradient. For the response surface $E(y) = f(x_1, \ldots, x_k)$, the gradient is

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_k} \right)'$$, a vector of slopes.

In particular, for the first-order model

$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$, the resulting estimated direction of steepest ascent is

$$\left(b_1, b_2, \ldots, b_k \right)'.$$

Therefore, a step change proportional to the regression coefficient for each factor will be the path to follow in subsequent experiments to "climb the hill" to the optimum. For our climb, we decided each step would be 1.0 cm for factor A (rotor length) and correspondingly −0.39 cm for factor B (rotor width). Five steps along the path of steepest ascent, starting at the base, are shown in Table 5. An approximate maximum was found at Step 3.

TABLE 5. Steepest ascent coordinates and results in uncoded units (centimeters). Time measured in seconds. Although the fold width (*F*) was not a factor, it is shown in the table because it had to vary with rotor width (*B*). *F* had to decrease as *B* decreased.

	Factors			
Step	A	B	F	Time
Base	8.50	4.00	2.0	12.99
1	9.50	3.61	2.0	15.22
2	10.50	3.22	2.0	16.34
3	**11.50**	**2.83**	**1.5**	**18.78**
4	12.50	2.44	1.5	17.39
5	13.50	2.05	1.2	7.24

Central Composite Designs

The basic experimental designs to estimate second-order response surface models are central composite designs. A two-dimensional central composite design is shown in Figure 5. These designs consist of a factorial design for estimating first-order and two-factor interactions, $2k$ axial (or "star") points at

$(\pm a, 0, \ldots, 0), (0, \pm a, 0, \ldots, 0), \ldots, (0, \ldots, 0, \pm a)$

for estimating the second-order terms, and replicated center points to estimate the second-order terms and to provide an estimate of error. The value of a is set at $\sqrt{k}$ for rotatability, so the accuracy of prediction with a quadratic equation will not depend on direction.

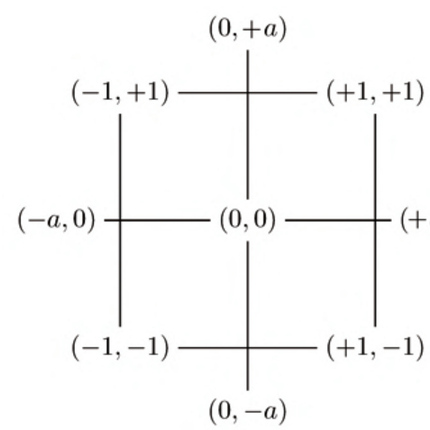


FIGURE 5. A two-dimensional central composite design

For our central composite design, $a = \sqrt{k} = \sqrt{2}$, as shown in Table 6 in coded and uncoded units centered at the maximum steepest ascent point (Step 3, in Table 5). The results are shown in Table 7.

TABLE 6. Central composite design factor levels in coded and uncoded units (centimeters)

Factor	$-\sqrt{2}$	−1	**0**	1	$\sqrt{2}$
A	10.08	10.50	**11.50**	12.50	12.91
B	2.28	2.44	**2.83**	3.22	3.38

TABLE 7. Experimental results from the central composite design. Time in seconds. The order of runs was randomized to minimize the risk of the sequence of experimentation affecting the response.

Run	Order	A	B	Time
1	7	–1	–1	13.65
2	3	1	–1	13.74
3	11	–1	1	15.48
4	5	1	1	13.53
5	9	0	0	17.38
6	2	0	0	16.35
7	1	0	0	16.41
8	10	+1.414	0	12.51
9	4	–1.414	0	15.17
10	6	0	+1.414	14.86
11	8	0	–1.414	11.85

Finding an Optimum

A near-optimum location was found and a second-order model was used to describe the curvature near the optimum to determine optimal conditions.

Near the optimum, the response surface is curved and can be approximated by a second-order model. A maximum or minimum will occur at a stationary point, where the partial derivatives are all 0, $\frac{\partial \hat{y}}{\partial x_i} = 0$, for $i = 1, \ldots, k$. Determining the shape of the surface is easy if k is 1 or 2 by plotting the surface or looking at contour plots. If there are many x variables, then visualizing the surface is difficult, but information about curvature can be understood analytically. Canonical analysis takes a square matrix of the β_{ij} coefficients and calculates eigenvalues $\lambda_1, \ldots, \lambda_k$ and eigenvectors. The eigenvalues tell us the direction and magnitude of curvature, with the sign indicating downward (– = concave) or upward (+ = convex) curvature and magnitude indicating fairly flat (near 0) to steeply peaked (far from 0). Wanting a maximum in our experiment, we hoped all the λs would be negative.

Our Final Model

Using regression analysis, the estimated second-order model is

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 + b_4 x_1^2 + b_5 x_2^2$$

$$= 16.713 - 0.702 x_1 + 0.735 x_2 - 0.510 x_1 x_2 - 1.311 x_1^2 - 1.554 x_2^2.$$

Model Validation

To validate our model, we checked the significance of the model terms, performed a lack-of-fit test, and checked our statistical assumptions.

Using *t*-tests for the regression coefficients, Table 8 indicates that all factors except the interaction term were significant at the 0.05 level. To test lack-of-fit, the pure error in the process can be estimated using the center points. The lack-of-fit test suggests the model adequately describes the data because the *p*-value = 0.3737 in Table 9 is large; we do not reject the null hypothesis (H_0) of fit (i.e., no lack of fit).

TABLE 8. Second-order response surface parameter estimates are all significant, except for the interaction term, indicating the interaction term could be dropped from the model.

Parameter	df	Estimate	Std Err	*T*	*p*-value
Intercept	1	16.713	0.408	40.98	0.0001
$x_1 = a$	1	–0.702	0.250	–2.81	0.0374
$x_2 = b$	1	0.735	0.250	2.94	0.0322
$x_{12} = a \times b$	1	–0.510	0.353	–1.44	**0.2083**
$x_1^2 = a \times a$	1	–1.311	0.297	–4.41	0.0070
$x_2^2 = b \times b$	1	–1.554	0.297	–5.23	0.0034

TABLE 9. Lack-of-fit test indicates the model is adequate (i.e., it does not "lack fit").

Residual	df	SS	MS	*F*	*p*-value
Lack of Fit	3	1.826	0.609	1.82	**0.3737**
Pure Error	2	0.668	0.334		
Total Error	5	2.495	0.499		

We checked that the statistical requirements of independent and identically distributed residuals were satisfied. The residuals plot indicated the residuals were random with constant variance. Additionally, we tested the residuals for normality. The test had a large *p*-value ($p = 0.761$), indicating the normality assumption should not be rejected. Therefore, our second-order model appeared adequate.

We used SAS's RSREG (Response Surface Regression) procedure to predict the stationary point at (a= −0.32, b=0.29) with the predicted response $\hat{y}$=16.9. The eigenvalues $(\lambda_a, \lambda_b) = (-1.15, -1.71)$ indicated a maximum—yes!

Our Optimum Design

The estimated response surface is plotted in Figure 6, with the 90% confidence region for the location of the maximum emphasized at the top. (See the reference on Page 28 by Enrique Del Castillo and Suntara Cahya about computing confidence regions on a response surface stationary point.) For our model, R^2 = .92, which indicates the model should be able to explain about 92% of the variation in the response over the range of factor values we tested. The final optimal helicopter design is shown in Figure 7.

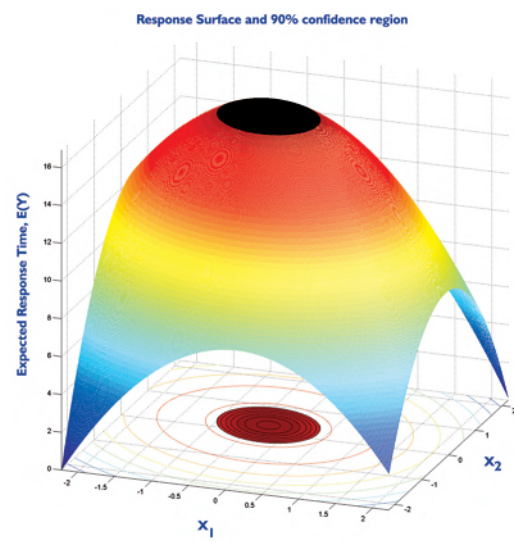


FIGURE 6. Estimated response surface with 90% confidence region for the coordinates of the maximum response

Confirmatory Experiment

Last, we conducted a confirmatory experiment at the optimum location to verify the second-order prediction. The results of our confirmatory experiment consisting of six new helicopters at the predicted optimal setting are shown in Table 10. The mean response is 17.81 seconds with standard deviation 1.67 seconds, with a 95% confidence

TABLE 10. Results from the confirmatory experiment corroborate the model's predicted response.

Flight Time (sec)	15.5	16.4	19.7	19.4	18.6	17.3

Design of Experiments VOCABULARY

Effect: The mean difference in response due to changing a factor's level, such as the mean difference in flight time of paper helicopters when rotor length is changed from short to long.

Experimental Design: A systematic procedure for changing factor levels to measure and compare the responses, such as changing rotor length and rotor width on paper helicopters.

Factor: An independent variable deliberately changed in an experiment to study its effect on the response, such as the length of a paper helicopter's rotors.

Factorial Design: An experiment that measures the responses from all possible combinations of two or more factors at fixed levels, such as the four combinations of two rotor lengths and two rotor widths for paper helicopters.

Interaction: When the response due to one factor is influenced by another factor, such as the influence of rotor width on rotor length for paper helicopters.

Response: The performance of an experimental unit, such as the flight time of a paper helicopter.

Response Surface: A curved surface representing how a system's output performance is affected by specified input factors, such as how paper helicopter flight time is affected by rotor length and rotor width.

interval on the mean response of (16.1, 19.6). Our model's predicted response, $\hat{y}$=16.9 seconds, falls within this interval, thus confirming the usefulness of our response model for the factor levels we tested and under the environmental conditions in our experiments.

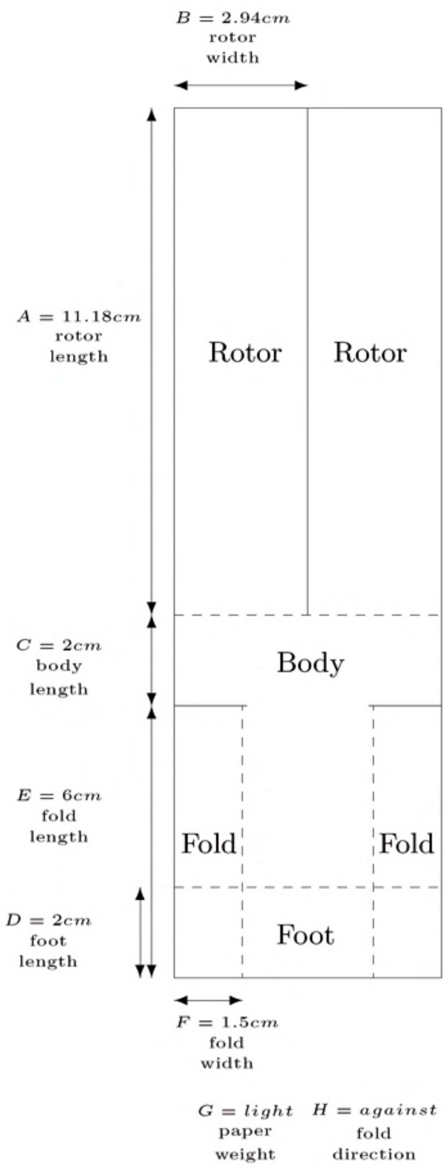


FIGURE 7. The optimal helicopter design is longer and narrower and has a shorter base than the initial helicopter design.

There You Have It

So, there you have it. That is what we did. Now, see if you can do even better. Good luck, and please let us know what you find out about using response surface methodology to design a better paper helicopter. Have fun. ◕

Answers from **Learning Stats Is FUN**

Page 7

❶ One...plus or minus 2.4

❷ To get data from the other side of the median

❸ Residual plots

❹ Just think how unlikely it is for there to be two bombs on board. She'd lose her degrees of freedom.

❺ A hisss-togram

❻ The South-American ANOCOVA

Page 11

① 65.7 km (about 40 miles)

② 23 people

③ ⅔

Vocabulary Is More Than Just **Knowing** the **WORDS**

by Peter Flanagan-Hyde

I have participated in grading the Advanced Placement (AP) statistics examinations for a number of years, and it always amazes me how reading many students' papers and seeing where their misunderstandings cluster help me reach a deeper understanding of some of the most basic ideas of the AP statistics course. The 2006 exam was no exception, as students had consistent difficulty on its Question 5, about the design of an experiment. In particular, students had trouble effectively using the vocabulary of experimental design on this question, especially the meaning of the words factor, treatment, variation, and confounding.

Many of the students who managed to explain the design of the experiment correctly had large gaps in how they expressed these fundamental terms, often confusing and mis-stating their analyses of the design. As I struggled to figure out where they were going off track, it helped me think more deeply about the fundamental ideas of experimental design. It powerfully struck me, while reading this year's exams, that there were many students who could quote the definitions of the terms and quote statements of interpretations without having anything but the most superficial understanding of the concepts involved.

PETER FLANAGAN-HYDE has been a math teacher for 27 years and has taught AP Statistics since its inception in the 1996-1997 school year. 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.

Shrimp Growth Experiment

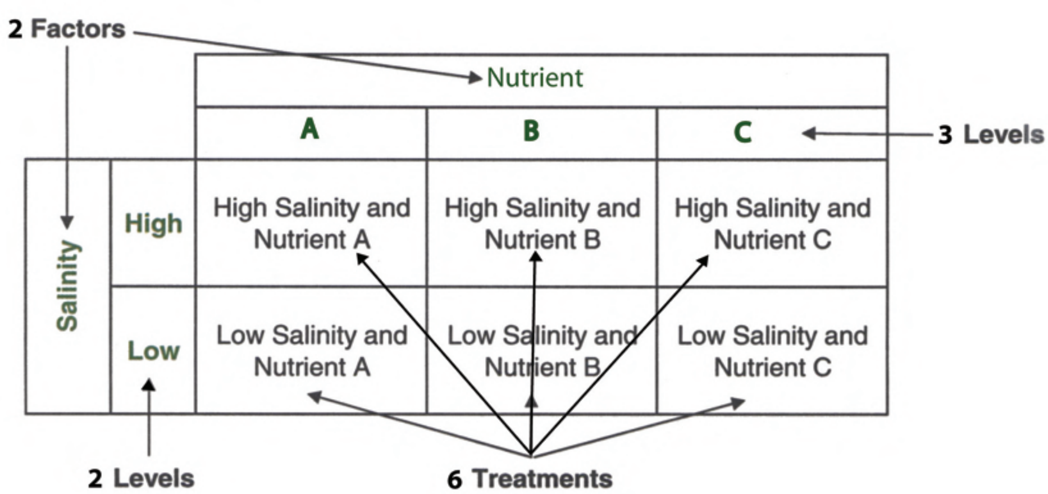


FIGURE 1. Shrimp growth experiment factors, levels, and treatments

The scenario for the question was a two-factor experiment on the growth of shrimp. The experiment was to be conducted on a specific species of shrimp—tiger shrimp—with two factors: salinity and growth-enhancing nutrient. There were two levels of salinity (high and low) and three levels of supplemental nutrients arbitrarily labeled A, B, and C. Students needed to state the treatments for the experiment, design a completely randomized experiment, and comment on both the advantages and disadvantages of restricting the experiment to one species of shrimp.

Factor and Treatment

Many students revealed in their answers that their understanding of the terms factor and treatment did not really distinguish one from the other, as they tended to use the terms interchangeably. In this two-factor design, there were six treatments, one for each of the possible combinations of the factors. Many students, however, listed the factors and their levels when asked about the treatments, just stating high salinity, low salinity, nutrient A, nutrient B, and nutrient C. They completely missed the idea that a treatment consisted of a particular combination of the factors (salinity and nutrient) at particular levels (high salinity with nutrient A, etc.).

However, for many students, the confusion about the vocabulary did not prevent them from correctly designing a completely randomized experiment in which there were, indeed, six treatments. They just did not know exactly what the term "treatment" referred to in the first part of the question. Part of the problem may have been that students were much less familiar with the two-factor setting than with the more typical classroom problem of a single-factor experiment. With only one factor, the distinction between treatment and factor is less clear in students' minds, as they can essentially interchange factor level with treatment.

Variation and Confounding

The second half of Question 5 asked about the statistical advantage and disadvantage of restricting the experiment to one species of shrimp. Students were better at articulating the disadvantage—limited generalizability of the results, because other shrimp species might react differently to the treatments—than they were about the advantage. To correctly answer this part of the question, students needed to have a secure understanding of variation, its sources in an experimental setting, and the strategies used to manage unwanted variation.

> "Confounding is a word for which many students seem to know the definition, but not the meaning of the term."

Factor is an explanatory or predictor variable to be studied in an investigation.

The possible values of a factor are called the levels of that factor

Treatment is a specific combination of the levels of the factors.

Confounding variable is a variable whose effect on the response variable cannot be separated from the effect of the explanatory variable (factor) on the response variable. It is a variable that offers an alternative explanation for the result of the study.

Variation refers to observed changes or inconsistencies in a characteristic, attribute, or action.

Variation due to differences in treatments is the 'good' variation in an experiment. In this experiment, the good variation is differences in weight gain due to differences in salinity levels and added nutrients. We are trying to quantify this variation as specifically as possible, but this is made more difficult by 'bad' variation that arises due to factors other than the treatments.

These other factors are extraneous variables—other characteristics of the conditions and subjects in the experiment that cause differences in the response. In this experiment, there can be differences in the conditions in the tanks in which the experiment takes place, as well as innate differences among the shrimp in the experiment. The design of the experiment should attempt to minimize the effects of these sources of variation.

For the tanks, this means efforts on the part of the experimenters to make the conditions in the tanks as identical as possible. Despite these efforts, there can be no guarantee that there are not differences that remain between the tanks. Randomly assigning the treatments to the tanks is how experimenters can make sure there is no systematic way in which differences in the tanks are associated with particular treatments.

For the shrimp, this is partly addressed by randomly assigning the shrimp to the tanks. Minimizing the differences between shrimp, however, is a second step that helps reduce the variation in weight gain due to difference from shrimp to shrimp. This is the point of limiting the shrimp to one particular species: tiger shrimp. This reduction in variation is what students should be able to identify as the advantage of having only tiger shrimp in the experiment.

However, student responses about the statistical advantage revealed commonly held misconceptions about an important term in experimental design: confounding. Perhaps the most commonly stated incorrect advantage was that having only one type of shrimp "eliminates confounding." Confounding is a word for which many students seem to know the definition, but not the meaning of the term. They are typically able to chant "confounding is when you cannot tell if the response is due to the treatment or another variable," and they know it is 'bad.' They also know it prevents you from easily seeing the true effects of a variable on a response.

Having more than one species of shrimp potentially increases the difficulty of seeing the true effects of salinity and nutrients, but this does not make it confounding. As stated in the AP Scoring Guidelines for this question, "In this completely randomized design, confounding is not possible." This is worded bluntly and strongly, I think, to reinforce the point that confounding is widely misunderstood by students and perhaps their teachers. For confounding to be present, there needs to be a systematic way in which the treatments are tied together or to an extraneous variable. This occurs often in observational studies, but should not in an experiment, when the treatments are randomly assigned and all combinations of factors are tested.

It is possible to imagine situations in which confounding would be present in trying to understand the effects of salinity and nutrients. For example, if the experiment were conducted in such a way that the higher level of salinity was applied inadvertently to only tiger shrimp, while the lower level to only some other type of shrimp, confounding may well be present. However, species could be included in the experiment as a third factor without causing confounding by testing all factor combinations and by randomizing. Then, if a different species of shrimp responds (grows) differently than tiger shrimp in the various treatments, that effect could be determined.

Some students used the term "confounding" in an almost casual way, making statements such as "Having one species of shrimp reduces variability and so eliminates potential confounding variables." This use conveys a sense that the student's concept of a confounding variable is anything that makes the situation murkier in determining an effect. If so, the student has not separated the ideas of variability and confounding in his or her mind, as both can contribute to making it harder to see if an effect is present.

In this problem, with a well-designed, completely randomized experiment, confounding should not be present. Removing the possibility of confounding is really the point of doing an experiment, allowing one to make a conclusion about cause and effect.

Lessons for Learning Statistics

In the students' incorrect responses, there are two lessons for learning statistics. The first is to be exposed to a variety of experimental situations and to make sure some have multiple factors. The second is that it is often difficult to judge from the final product—in this case, a correctly designed experiment—whether all the constituent concepts that support the final product are correctly configured. It takes some probing with careful assessment along the way to make sure that when we have learned to reproduce a particular experimental design, we also are able to apply each of the concepts in somewhat different situations. The depth of our understanding of a particular concept can be measured by the distance we carry it to reach a new situation, and that takes repeated practice.

As teachers and students, we need to find effective ways to learn the vocabulary of statistics that produce something more than recognizing the terms and being able to replay each definition. This is at the heart of teaching and learning for understanding, an idea that has been promoted by Grant Wiggins and Jay McTighe, among others. These authors contend that students "cannot be said to understand their own answer, even though it is correct, if they can only answer a question phrased just so. Furthermore, they will not be able to use what they 'know' on any test or challenge that frames the same question differently."

The idea that we should be teaching and learning for an understanding that is flexible, adaptable, and able to be transferred to new situations is not new. It dates back to at least the early 1950s when Benjamin Bloom developed his taxonomy of educational objectives. His words about how to design useful assessment tasks could have come from the current writers of the AP exam: "Ideally, we are seeking a problem which will test the extent to which an individual has learned to apply the abstractions in a practical way."

The capability to apply concepts in new situations is the appropriate measure of understanding. As we move through the curriculum, there is a natural tendency to reduce the concepts to a phrase or two. At times, we need to rely on memorized formulations to help us reach a beginning understanding of an idea, but we need genuine practice in using the concepts in a variety of situations to develop a deeper understanding.

It is certainly correct to say, "Confounding is when you cannot tell if the response is due to the treatment or another factor," but if this is the extent of our knowledge of the term, we will not have the kind of flexible and adaptable understanding we can transfer to new situations as they arise, either on tests or—much more importantly—when we use the ideas we have learned in new settings and new situations in future coursework or when reading the paper some Saturday morning. ◕

References and Additional Reading List

The references for each article in this issue of *STATS* are included in the listing below, along with suggestions for additional reading on related topics. The page numbers are the numbers in blue.

3 Numb3rs, Sabermetrics, Joe Jackson, and Steroids.

You can see the "Hardball" episode of Numb3rs at *www.cbs.com/primetime/numb3rs*.

Jim Albert's work appeared in *STATS* when he wrote, "Does a Baseball Hitter's Batting Average Measure Ability or Luck?", *STATS*, Issue 44, 2005.

The details of Jay Bennett's methodology appear in "Did Shoeless Joe Jackson Throw the 1919 World Series?", *The American Statistician,* 47(4): 241–250, 1993. It was republished in *Anthology of Statistics in Sports*, edited by J. Albert, J. Bennett, and J. Cochran, American Statistical Association and Society of Industrial and Applied Mathematics, 2005.

Scott Berry's study of steroid use appears in, "A Juiced Analysis," *CHANCE*, 15(4): 2002.

The American Statistical Association's Section on Statistics in Sports web site is at *www.amstat.org/sections/sis*.

Joe Jackson's biography by David L. Fleitz is *Shoeless: The Life and Times of Joe Jackson,* McFarland & Company, Inc., Publishers, 2001.

The web site for the Shoeless Joe Jackson Society is at *www.blackbetsy.com/society.htm*.

J. Albert and J. Bennett's book about baseball statistics is *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*, Copernicus Press, 2003.

W. F. McNeil and P. Palmer's sabermetrics book is *Backstop: A History of the Catcher and Sabermetric Ranking of 50 All-Time Greats,* McFarland & Company, Inc., Publishers, 2005.

The web site for the Society for American Baseball Research (SABR) can be found at *www.sabr.org*.

For the stories behind the stats, see *www.baseballlibrary.com;* baseball stats are available at *www.baseball-reference.com*.

7 Learning Stats Is Fun…with the Right Mode.

Humor and Song

Examples of Larry Lesser's statistical music appear in "Average Love Songs," *Amstat News*, 339(51): 2005; "Stat Song Sing-Along," *STATS*, 33: 16–17, 2002; and "Musical Means: Using Songs in Teaching Statistics" *Teaching Statistics*, 23(3): 81–85, 2001.

Larry Lesser presented "Making Statistics Learning Fun" in a Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) webinar, accessible at *www.causeweb.org/webinar/2006-04*.

For information about CAUSEweb, see *www.causeweb.org* and *www.causeweb.org/resources/fun*.

Carmen Wilson VanVoorhis' research using songs to learn statistics, "Stat Jingles: To Sing or Not To Sing," appears in *Teaching of Psychology*, 29(3): 249–250, 2002.

Randall Garner describes his research about using humor to learn statistics in "Humor in Pedagogy: How Ha-Ha Can Lead to Aha!", *College Teaching*, 54(1): 177–180, 2006.

Several humor collections are in Hershey Friedman, Linda Friedman, and Taiwo Amoo's "Using Humor in the Introductory Statistics Course," *Journal of Statistics Education,* 10(3): 2002, accessible at *www.amstat.org/publications/jse/v10n3/friedman.html*.

General connections between music and statistics are in Jan Beran's *Statistics in Musicology*, Chapman and Hall/CRC, 2004.

Books and Journals

For a statistics book with a humorous style, read Fred Pyrczak's *Statistics with a Sense of Humor,* Pyrczak Publishing, 1998.

Another statistics book with a light touch is Larry Gonick and Woollcott Smith's *The Cartoon Guide to Statistics*, Harper Collins, 1993.

Your library might have a copy of Richard Runyon's *Winning with Statistics: A Painless First Look at Numbers, Ratios, Percentages, Means, and Inference*, Addison-Wesley, 1977.

Sources of irreverent uses of statistics in 'research' include *The Journal of Irreproducible Results* at *www.jir.com* and the *Annals of Improbable Research* at *www.improb.com*.

Games

Steve Abbott and Matt Richey "Take a Walk on the Boardwalk" in *College Mathematics Journal*, 28(3): 162–171, 1997.

Phil Woodward solves the popular dice game Yahtzee in "Yahtzee: the Solution," *CHANCE*, 16(1): 18–22,

2003. He discusses poker in "Call My Bluff," *Significance*, 3(1): 30–32, 2006.

Paul Stephenson, Mary Richardson, and John Gabrosek analyze the dice game GOLO in "How LO Can You GO? Analyzing Probabilities for the Dice-Based Golf Game GOLO," a paper in the *2006 JSM Proceedings* [CD-ROM], 2377–2381. You can find GOLO at *http://igolo.com*.

Mary Richardson and David Coffey analyze the game Pass the Pigs in "What Is the Probability of 'Pigging Out'?" on the *2005 ASA Proceedings* [CD-ROM], 2299–2301.

John Kern provides a Bayesian view of Pass the Pigs in his paper, "Pig Data and Bayesian Inference on Multinomial Probabilities," *Journal of Statistics Education*, 14(3): 2006, accessible at *www.amstat.org/publications/jse/v14n3/datasets.kern.html*.

Larry Feldman and Fred Morgan go 'HOG wild' in "The Pedagogy and Probability of the Dice Game HOG," *Journal of Statistics Education*, 11(2): 2003, accessible at *www.amstat.org/publications/jse/v11n2/feldman.html*.

Edward Packel examines Backgammon and other games in *The Mathematics of Games and Gambling,* Mathematical Association of America, 1981.

Felicia Trachtenberg explains "The Game of Dreidel Made Fair" in *College Mathematics Journal*, 27: 278–281, 1996.

Robert Feinerman gives a statistical 'spin' to "An Ancient Unfair Game" in *American Mathematical Monthly,* 83: 623–625, 1976.

Ivars Peterson explains the probability of "Guessing Cards" on MAA Online, 2001, accessible at *www.maa.org/mathland/mathtrek_12_24_01.html*.

Game Shows

Robert Quinn goes "Exploring the Probabilities of 'Who Wants to Be a Millionaire?'" in *Teaching Statistics,* 25(3): 81–84, 2003.

Diane Evans published "Conditional Probability and the 50:50 Lifeline on 'Who Wants to Be a Millionaire?'" in the *2006 JSM Proceedings* [CD-ROM], 2267–2271.

Ed Barbeau presents the three-door problem in "Fallacies, Flaws, and Flimflam," *College Mathematics Journal,* 24(2): 149–154, 1993.

Matthew Carlton also looks at the three-door scenario in "Pedigrees, Prizes, and Prisoners: The Misuse of Conditional Probability," *Journal of Statistics Education,* 13(2): 2005, accessible at *www.amstat.org/publications/jse/v13n2/carlton.html*.

Federico O'Reilly took "A Look at the Two-Envelope Paradox" in *STATS,* 45: 14–17, Spring 2006.

LaDawn Haws discusses "Plinko, Probability, and Pascal" in *Mathematics Teacher*, 88(4): 282–285, 1995.

Amy Biesterfeld shows that "The Price (or Probability) Is Right" in *Journal of Statistics Education,* 9(3): 2001, accessible at *www.amstat.org/publications/jse/v9n3/biesterfeld.html*.

Susie Lanier and Sharon Barrs say, "Let's Play Plinko: A Lesson in Simulations and Experimental Probabilities" in *Mathematics Teacher,* 96(9): 626–633, 2003, with resources available at *http://mathdemos.gcsu.edu/mathdemos/plinko/index.html*.

Eric Wood explains "Probability, Problem Solving, and 'The Price Is Right'" in *Mathematics Teacher,* 85(2): 103–109, 1992.

You can track your own success in maximizing the offer by playing a simulation of "Deal or No Deal" at *www.nbc.com/Deal_or_No_Deal/game*.

David Kalist looks at "Data from the Television Game Show 'Friend or Foe?'" in *Journal of Statistics Education*, 12(3): 2004, accessible at *www.amstat.org/publications/jse/v12n3/datasets.kalist.html*.

Internet

Webster West and Todd Ogden, "Interactive Demonstrations for Statistics Education on the World Wide Web," *Journal of Statistics Education*, 6(3): 1998, accessible at *www.amstat.org/publications/jse/v6n3/applets/LetsMakeaDeal.html*.

One place to find Mozart's Musical Dice Game is *http://sunsite.univie.ac.at/Mozart/dice*.

A web site that randomly generates songs is *www.amiright.com/cgi-bin/songrandom.cgi*; one that makes band names is *www.bandnamemaker.com*.

Statistics 'rap' lyrics can be found at *www.lawrence.edu/fast/jordanj/rap.html* and statistics 'rap' videos can be found at *http://video.google.com/videoplay?docid=489221653835413043*.

Creative Works

An example of published song lyrics with statistical language is David Wilcox's "Down Here," Underneath, Vanguard 79528–2, 1999.

Wislawa Szymborska's poem, "A Word on Statistics," appears in *The Atlantic Monthly*, 279(5): 68, 1997.

Shirley Jackson's "The Lottery" is in X. J. Kennedy, Dorothy Kennedy, and Jane Aaron's *The Bedford Reader,* 6th ed., Bedford Books, 1997, accessible at *http://fisheaters.com/thelottery.doc*.

Norton Juster's novel, *The Phantom Tollbooth*, published by Random House, Inc., in 1961.

More Fun

You can have fun testing hypotheses about packaged candy, such as m&m's, using resources such as the Mars, Inc., web site at *http://us.mms.com/us/about/products*.

The frog 'causality' example and many other fun ideas appear in Harry Norton's paper, "Keeping an Introductory Statistics Course Interesting: Use of Demonstrations, Examples, Rewards, and a Little Humor," presented at the Joint Statistical Meetings, Seattle, Washington, 2006.

Eric Sowey catalogues examples of intriguing statistical situations in "Striking Demonstrations in Teaching Statistics," *Journal of Statistics Education*, 9(1): 2001, accessible at *www.amstat.org/publications/jse/v9n1/sowey.html*.

Buy humorous T-shirts from the American Statistical Association at *www.amstat.org/asastore*.

Explore the birthday problem with Larry Lesser in "Exploring the Birthday Problem with Spreadsheets," *Mathematics Teacher*, 92(5): 407-411, 1999.

Bruce Trumbo, Eric Suess, and Clayton Schupp use R for "Simulation: Computing the Probabilities of Matching Birthdays," *STATS*, 43: 3–7, 2005.

Let's Party

For "Pi Day" ideas, see "Slices of Pi: Rounding Up Ideas for Celebrating Pi Day," *Texas Mathematics Teacher*, 51(2): 6–11, 2004, or visit *www.math.utep.edu/Faculty/lesser/piday.html*.

Find famous statisticians' birthdays at *www.york.ac.uk/depts/maths/histstat/people* or the math history site, *www-history.mcs.st-andrews.ac.uk/BiogIndex.html*.

12 What Is This Statistics Education Stuff?

Joan Garfield's web site is located at *http://education.umn.edu/EdPsych/faculty/Garfield.html*.

The Guidelines for Assessment and Instruction in Statistics Education (GAISE) can be accessed at *www.amstat.org/Education/gaise/GAISECollege.htm*.

The United States Conference on Teaching Statistics (USCOTS) web site is *www.causeweb.org/uscots*.

For more about statistics education, see D. Ben-Zvi and J. Garfield's (editors) *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, Kluwer Academic Publishers, 2004, and I. Gal and J.B. Garfield's (editors) *The Assessment Challenge in Statistics Education*, IOS Press and the International Statistical Institute, 1997.

For information about the graduate education minor in graduate interdisciplinary specialization in college and university teaching at The Ohio State University, contact Jackie Miller at *jbm@stat.ohio-state.edu*.

14 Designing a Better Paper Helicopter Using Response Surface Methodology.

Raymond H. Myers and Douglas C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley, 1995.

SAS and Minitab macros accompany the text by Joseph Petruccelli, Balgobin Nandram, and Minghui Chen, *Applied Statistics for Engineers and Scientists*, Prentice Hall, 1999. Available at *http://users.wpi.edu/~jdp/downloads/book_macros/main.html*.

Cuthbert Daniel explains the "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments" in *Technometrics*, 1: 311–342, 1959.

Russell V. Lenth gives a "Quick and Easy Analysis of Unreplicated Factorials" in *Technometrics*, 31; 469–473, 1989.

Enrique Del Castillo and Suntara Cahya show "A Tool for Computing Confidence Regions on the Stationary Point of a Response Surface" in *The American Statistician*, 55: 358–365, 2001, available at *http://lib.stat.cmu.edu/TAS/BH*.

To learn about design of experiments and the paper helicopter, see Chapter 12 of *Statistics for Experimenters: Design, Innovation, and Discovery*, by George E. P. Box, J. Stuart Hunter, and William G. Hunter, 2nd ed., John Wiley & Sons, Inc., 2005.

See also the following:

George Box, "Teaching Engineers Experimental Design with a Paper Helicopeter," Report No. 76, December 1991, Center for Quality and Productivity Improvement, University of Wisconsin–Madison.

David Annis, "Rethinking the Paper Helicopter: Combining Statistical and Engineering Knowledge," *The American Statistician*, 59(4): 320–326.

Terry Siorek and Raphael Haftka, "Paper Helicopter—Experimental Optimum Engineering Design Classroom Problem," *American Institute of Aeronautics and Astronautics*, 1998.

J. Adam Johnson, Scott Widener, Howard Gitlow, and Edward Popovich, "A 'Six Sigma' Black Belt Case Study: G. E. P. Box's Paper Helicopter Experiment Part A and Part B," *Quality Engineering*, 18(4): 413–442.

22 Vocabulary Is More Than Just Knowing the Words.

Blooms' Taxonomy is presented by Benjamin Bloom (editor) in *Taxonomy of Educational Objectives: Classification of Educational Goals*, Longman, Green, & Co., 1956.

Six facets of understanding are discussed by Grant Wiggins and Jay McTighe in *Understanding by Design*, Association for Supervision and Curriculum Development, 2005.

2006 AP Statistics Exam questions can be found at *www.collegeboard.com/prod_downloads/student/testing/ap/statistics/ap06_frq_statistics.pdf*.

Scoring guidelines for the AP Statistics exam are available at *http://apcentral.collegeboard.com/apc/public/repository/_ap06_statistics_sg.pdf*. ◕

Subscribe to *STATS* Today!

Receive free online access to *STATS* magazine when you join or subscribe!

☑ YES!

I would like to **join the ASA** & receive
***STATS: The Magazine for Students of Statistics* FREE**

❑ **$50** K–12 Membership
❑ **$10** Student Membership

I would only like to subscribe to
STATS: The Magazine for Students of Statistics

❑ **$20** Individual Subscription Only
❑ **$25** Institutional/School Subscription

www.amstat.org/publications/stats

ASA Member ID# ______ Name ______

Organization ______

Address ______

City ______ State/Province ______ ZIP/Postal Code ______ Country ______

Phone ______ Email ______

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

Credit Card: ❑ VISA ❑ MasterCard ❑ American Express Amount Due $ ______

Card Number ______ Exp. Date ______

Name of Cardholder ______

Authorizing Signature ______

Return form with payment to:

ASA Subscriptions
Department 79081
Baltimore, MD 21279-0081
USA

or fax to: (703) 684-2037

ASA
AMERICAN STATISTICAL ASSOCIATION

(703) 684-1221 • asainfo@amstat.org • *www.amstat.org*