NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

American Statistical Association

The Statistics Teacher Network

www.amstat.org/education/stn/index.html

Number 65 ASA/NCTM Joint Committee on the Curriculum in Statistics and Probability **Winter 2004**

Random Babies and Random Bookbags:

Learning Basic Probability Through Simulating the Matching Problem

Allan Rossman and Beth Chance
Cal Poly–San Luis Obispo

One of the classic problems in probability theory is the "matching problem," which is often presented as: A group of men at a party throw their hats into the middle of a room, and at the end of the party each man selects a hat at random before they leave. What is the probability that everyone gets the right hat, or that nobody does, or that at least one person does, and what is the average number of men who get their own hat?

But we have found that our college students do not get excited by the context of men selecting random hats, so we often change the scenario to a more interesting one. One possibility, which is all in jest, and we apologize for the offensiveness if one takes the context seriously, is to ask students to consider a hospital staff that returns newborn babies to their mothers at random (*Workshop Statistics*, Activity 14-1). We use this problem and context to ask students to explore and discover fundamental ideas of probability and randomness. For younger students, the context could easily be changed to something less offensive but still likely to interest them, perhaps a group of students who throw their identical backpacks into a pile before they walk into the cafeteria and then select a backpack at random after lunch.

Randomization

So, how can students as low as middle school grades analyze this problem? We suggest beginning with a small number of students and starting with simulation before proceeding to an analytical solution. Suppose that three students' backpacks are randomly mixed up and that each student then selects a backpack at random. Ask students to simulate this situation with index cards to represent the hypothetical students; let's call them Amy, Barb, and Carol. Give each student three index cards and have them write one hypothetical student's name on each card. Then have them write the same three names on a sheet of paper. The cards represent the backpacks, and the names on the paper represent the hypothetical students. You know where this is leading: students shuffle up the three index cards thoroughly, and then they randomly deal out the three cards to the three names on the sheet of paper. Then they turn over the cards and see how many students got the correct backpack. For example, if the result is that Amy gets Barb's backpack, Barb gets Amy's, and Carol gets her own, then this randomization

Also In This Issue...

produced one match. Ask each student in your class to record the number of correct "matches" for their randomization, and then have each student repeat this process several times. Finally, pool the results for the class. One way to do the pooling is to have each student put tally marks on the board corresponding to his/her results. For example, suppose that a student conducts six randomizations and obtains:

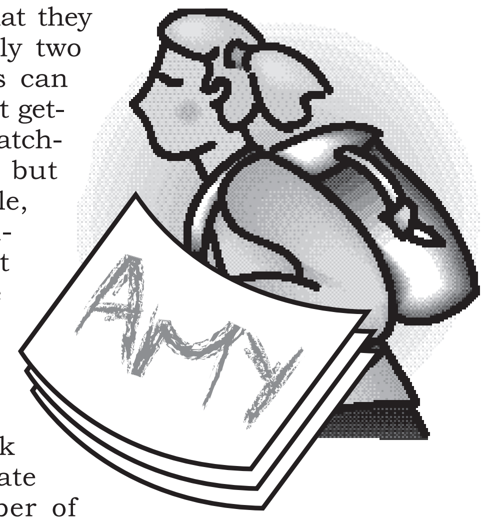
Trial	Name	What backpack	Number of matches
1	Amy Barb Carol	Barb's backpack Amy's backpack Carol's backpack	1
2	Amy Barb Carol	Amy's backpack Carol's backpack Barb's backpack	1
3	Amy Barb Carol	Barb's backpack Carol's backpack Amy's backpack	0
4	Amy Barb Carol	Amy's backpack Barb's backpack Carol's backpack	3
5	Amy Barb Carol	Carol's backpack Barb's backpack Amy's backpack	1
6	Amy Barb Carol	Carol's backpack Amy's backpack Barb's backpack	1

Simulation

Then he/she would put one tally mark under 0, four tally marks under 1, and one tally mark under 3.

Then looking at the combined results of the entire class, the instructor can lead the students through calculation of the fraction of randomizations that produced zero matches, the fraction that produced one match, and so on. These fractions are empirical probabilities that estimate the theoretical probabilities. In this manner students can begin to develop an understanding that probability refers to the long-run fraction of times that an outcome would happen in the long run if the random process were repeated indefinitely under identical conditions. Students can be led to realize that repeating the random process more times will lead to more precise estimates of the probabilities.

Even in this simple case with three items, some interesting results appear. Students should find that one match is a more common outcome than zero matches, that obtaining three matches is fairly uncommon, and that they never obtain exactly two matches. Students can be led to realize that getting exactly two matches is not unlikely but rather is impossible, because if two students get the correct backpack, then the only remaining choice for the third student is also the correct backpack. You can also ask students to calculate the average number of matches per randomization, and they should find it to be close to 1.



Next, depending on the level of your students, you could lead them to calculate the exact theoretical probabilities for this problem. The key to the exact analysis is to enumerate the sample space (the set of all possible outcomes), and to realize that the "at random" assignment of backpacks to students means that all outcomes in the sample space are equally likely. For this matching problem with 3 students and 3 backpacks, there are 3! = 6 possible outcomes: ABC, ACB, BAC, BCA, CAB, and CBA, where xyz means that Amy gets backpack x, Barb gets backpack y, and Carol gets backpack z. You can ask students to determine the number of correct matches for each outcome; the answers are: ABC (3), ACB (1), BAC (1), BCA (0), CAB (0), and CBA (1). Because each of these outcomes has a 1/6 probability of occurring, the probability of getting zero matches is 2/6, the probability of getting one match is 3/6, and the probability of getting three matches (everyone getting their own backpack) is 1/6. Having done the simulation first, students should realize that this means, for example, that in the long run, 1/2 of of a large number of randomizations will produce exactly one person getting the correct backpack. For more advanced students, you can also introduce the concept of expected value here and calculate the expected number of matches to be 1. Referring to the simulation results can help students to interpret this expected value as the long-run average value that the empirical average from the process converges to.

Another way to extend this analysis is to consider four students and four backpacks. The simulation analysis can proceed as before, using four index cards rather than three. We have also developed a Java applet *(www.rossmanchance.com/applets/randomBabies/Babies.html)* that simulates this situation in a visually appealing manner, in the context of distributing babies to mothers at random. An instructor might demonstrate this applet in class or ask students to use it in a lab or at home. Clicking on the applet's "randomize" button causes a stork to bring four babies to the hospital, and the babies are then distributed to the four homes at random. The colors of the houses and diapers indicate which baby goes with which home, and the number of matches is recorded in a table next to the picture is recorded. This processess can be repeated over and over, and switching off the "Animation" feature can allow for a large number of randomizations to be simulated quickly. The applet generates a histogram of the number of matches.

In addition, clicking on one of the histogram bars reveals a time plot of the fraction of occurrences versus the number of trials and helps students understand the "law of large numbers."

This graph helps students to understand that the fraction fluctuates considerably at first but gradually begins to approach a limiting value. This limiting value is, in fact, the exact theoretical probability, as clicking on the "Theoretical values" button reveals.

The exact analysis for the four-person problem is analogous to the three-person case, but the number of outcomes is now 4, or 24. The exact probabilities for the number of matches turn out to be:

Number of matches	Probability
0	9/24
1	8/24
2	6/24
3	0
4	1/24

The expected value of matches continues to equal 1.

More advanced students can also consider the analysis of the matching problem with more than four items. The exact analysis becomes less feasible rather quickly, but simulation remains an effective means of producing approximate probabilities.

In summary, we have found this "matching problem" to present an engaging context in which students at many levels can explore basic concepts of probability. We strongly recommend starting with hands-on simulations, proceeding to technological simulations, and concluding with the exact mathematical analysis. We hope that your students find this activity to be fun as well as enriching, but try not to let them leave class with someone else's backpack!

Reference

Rossman, Allan; Chance, Beth; von Oehsen, J.Barr: *Workshop Statistics*: *Discovery with Data and the Graphing Calculator* (2nd ed.) Key College Publishing, 2002

Active Learning of Probability through Contrasts

Michael D. Larsen
Iowa State University • *larsen@iastate.edu*

How can we teach probability to students so that they actively participate in learning? One suggestion, the focus of this article, is to use contrasts; that is, to intentionally vary the questions and problems so that the process of solving a problem is made readily apparent. Most instructors use contrasts to some extent when presenting many topics. The first two and the last examples below are appropriate for upper elementary and above. The third and fourth examples should be accessible to high school and introductory college students. Some of these examples were presented originally in Larsen (2003). The editor was kind enough to suggest the fifth example.

Examples with Coins/Children

Probability at all levels often is introduced with exercises involving flipping coins and determining the probability of events such as obtaining three heads in a row, three heads out of five tosses, or a run of at least three heads in five tosses. Assuming the coins are fair (1/2 for heads, 1/2 for tails) and independent, the theoretical probabilities of these events occurring are 1/8, 5/16, and 1/4, respectively. After exploring these initial questions, the problem could be modified to be concerned with two heads or four heads and to involve four or six tosses. Students could be asked to compute and compare results under different scenarios.

A variation on these types of problems is to ask questions about compositions of families. Suppose a family decides to have children until it has three children or has at least one boy and one girl. What is the probability that the family has two children? That it has two or more girls? As with the coins, suppose that the chances for a female and male child are equal (1/2) and independent for each child. The answers are 1/2 and 1/4, respectively. A different 'stopping rule' produces a different sample space and solutions. A comparison family rule could be to have children until there are four children or at least one boy and one girl. An alternative rule, for a family that wants especially to have a girl, is to have at most n children but no more after a girl is born. The value of n could be, for example, two, three, or four. Students could be asked to contrast the ratio of boys to girls under various scenarios and conditions. Discussion of these and other problems appears in Larsen (2004a). This scenario may also be explored using tree diagrams. Students are able to visualize "stopping" after one boy or having at most two boys OR three children!

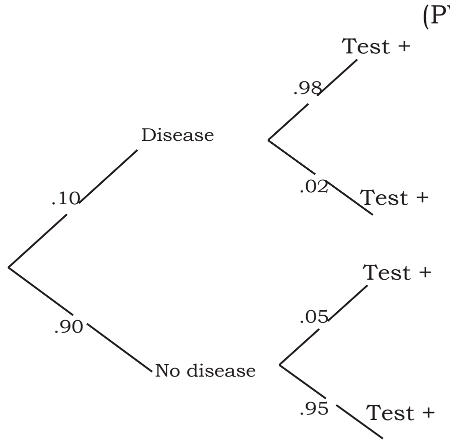
Examples with Dice

Dice experiments are another common hands-on probability demonstration at all levels of education. Murray Siegel ("*Let Them Roll, Then Show Us Your Data*" in volume 64 of *STN*) describes in detail an activity involving three dice for fifth-grade students. Contrast is used effectively when students are given individual problems appropriate for their mathematical levels. Contrast also could arise when comparing the distribution of sums when using two, three, or four dice. In general, this example gives students something to do, answers to discuss, and scenarios to contrast. Looking at sums for dice may lead to discussion of the "normal curve" and the central limit theorem well before the technical aspects of these phenomena are explored.

Examples with Probability Trees

A probability tree is a tool for studying conditional probability and Bayes' theorem as presented in high school and introductory college classes. A typical example is as follows. Suppose the prevalence (or probability) of a disease in a population is 10%. Many diseases or conditions, such as heart disease, high blood pressure, and hardness of hearing could be used as examples. A test produces a positive result 98% of the time (probability 0.98) if a patient has the disease, whereas a positive result occurs only 5% of the time if a patient is disease free. If members of the population are randomly selected and tested, what percentage tests positive? 0.10(0.98) + 0.90(0.05) = 0.098 + 0.045 = 0.143, 14.3%.

If a test is positive, what is the chance that the patient has the disease? (PV+ = 0.098/0.143 = 0.685, 68.5%.) The probabilities can be illustrated with the tree diagram. Contrast can be introduced by changing one or more of the probabilities involved in the statement of the problem. If the prevalence is 20%, how are the probability of testing positive and the PV+ affected? If the test detects 99% of the diseased cases, but only 3% of the disease-free cases, how do they change? One way to involve students actively would be to assign one change to the problem to each of several small groups of students and then collect answers from the various groups.



Example with Binomial Probabilities

The Binomial distribution is among the most common of discrete probability distributions. Let X denote a random variable (RV) that records the number of successes in n independent trials that each have success probability p ($0<p<1$). X can have values 0, 1, ..., n. The probability that X has value k in its range is P(X=k) = nCk p^k (1-p) ^ $^{n-k}$, where nCk is the number of ways of choosing k objects from n without replacement and without regard to order. One idea for generating contrasts is to allow students to select their favorite sports figures and select a value of p appropriate for the sport. For example, the AAA baseball team for the Chicago Cubs is located in Iowa and called the Iowa Cubs: *www.iowacubs.com/asp/statistics.asp.* Looking at two current players who have played more than 100 games with the team, Trenidad Hubbard has a batting average of 0.331, whereas David Kelton's average is 0.243. These averages are the proportion of the time a player hits the ball successfully in official "at bats" or attempts. The exact definition is complex, and "at bats" can involve several swings. Some caveats are necessary to imagine that a series of attempts are independent and have this probability of success, but as a simple model for a small number of well-spaced attempts of a certain type the Binomial model is reasonable. Out of 16 such attempts, which player is more likely to be successful (to hit the ball safely) 2, 3, or 4 times? Which player is more likely to be successful 7, 8, or 9 times? Students with appropriate calculators can easily compute these and other probabilities for a selection of players and scenarios. In a computer lab, students could also use an on-line demonstration such as *http://www-stat.stanford.edu/~naras/jsm/example5.html* or *http://www.stat.sc.edu/~west/applets/binomialdemo.html.* Indeed, as in Larsen (2004b), several sports provide values of p that could be selected by students.

Examples with Guessing on Tests

If a student takes a 50-question multiple-choice quiz with four possible answers, how many answers "should" he or she get correct? Try this experiment with a class at any level using the blackline master (see pg. 5, Shoecraft, 1984). What is the mean number of correct answers? What is the distribution of number of correct answers? Are there any outliers? What would happen if each question had five possible answers? What would the mean number of correct answers theoretically be? An entire descriptive statistics unit can be built upon this one worksheet!

In conclusion, exploring contrasting situations enables students to experiment and explore. Such activities encourage students to not simply memorize an answer to a specific problem, but to understand the process of finding and rationale for a solution. Contrast can be a useful and enjoyable tool for teaching probability at all levels.

References:

Larsen, M.D. (2003). Using Technology to Connect Topics and Contrast Scenarios, presented at Beyond the Formula Conference at Monroe Community College in Rochester, New York. *http://www.public.iastate.edu/~larsen/btf03/index.html*

Larsen, M.D. (2004a). The Importance of the Sample Space for Teaching Probability, Submitted to *Journal of Statistics Education*. June 2004.

Larsen, M.D. (2004b). *Internet Companion for Statistics: A guide to websites and activities.* Duxbury: *http://larsen.duxbury.com*

Siegel, M. (2004). Let Them Roll, Then Show Us Your Data. *Statistics Teachers Newsletter*, 64, page 4. *http://www.amstat.org/education/stn/pdfs/STN64.pdf*

Shoecraft, Paul. (1984). *Math Games and Activities.* Dale Seymour Publications, an imprint of Pearson Learning Group, a division of Pearson Education, Inc. Used by permission.

Guessing on a Test

Name........ ....

A multiple-choice question with four choices can be answered in four ways: a, b, c, or d. If each choice is equally likely to be picked, the probability of answering the question correctly is ¼. Thus if we guess on, say, 50 such questions, we would expect to answer 12 or 13 of them correctly.

Investigate how well the probability of guessing the correct answer to a four-choice, multiple-choice question agrees with what actually happens when you guess on a multiple-choice test with four choices per question. Take the 50-question final exam for MA 699 QUASI MATHENOMIAL COMPLEXULUS. Since you do not have the exam, just circle the a's, b's, c's, and d's of your choice on the answer sheet below. Then check your answers using the key at the bottom of this page and record your results in the table. Then complete the table and answer the questions below.

PROBABILITY

ANSWER SHEET

1. a b c d	11. a b c d	21. a b c d	31. a b c d	41. a b c d
2. a b c d	12. a b c d	22. a b c d	32. a b c d	42. a b c d
3. a b c d	13. a b c d	23. a b c d	33. a b c d	43. a b c d
4. a b c d	14. a b c d	24. a b c d	34. a b c d	44. a b c d
5. a b c d	15. a b c d	25. a b c d	35. a b c d	45. a b c d
6. a b c d	16. a b c d	26. a b c d	36. a b c d	46. a b c d
7. a b c d	17. a b c d	27. a b c d	37. a b c d	47. a b c d
8. a b c d	18. a b c d	28. a b c d	38. a b c d	48. a b c d
9. a b c d	19. a b c d	29. a b c d	39. a b c d	49. a b c d
10. a b c d	20. a b c d	30. a b c d	40. a b c d	50. a b c d

Outcome	Tally	Fre-quency	Relative frequency	Proba-bility	Difference between relative frequency and probability
Correct				$\frac{1}{4}$	
Incorrect				$\frac{3}{4}$	
	Sum	50	$\frac{50}{50} = 1$	$\frac{4}{4} = 1$	

1. What was your score on the exam?

2. Did anyone in your class score more than 50%? Less than 10%?

3. Which would you recommend for getting a high score on a multiple-choice test, guessing or studying?

4. Charley Brown uses the following strategy for taking a multiple-choice test:

> "Let's see now. In a multiple-choice test, the answer to the first question is almost always C. Then A to sort of balance the C. Then A again to trick you. Then B and D to break the pattern. They never go too long without a D. Then three B's in a row. They always have three B's in a row someplace. Then C and A . . . . If you're smart, you can pass a multiple-choice test without being smart."

Take the 50-question final exam for MA 699 QUASI MATHEMOMIAL COMPLEXULUS again using Charley Brown's strategy for taking a multiple-choice test and see if you can improve on your score.

Key: caabdbbbca dbcccabdda bcaabbcdda ccbadcabcd badbbddcac

258

Take a Chance by Exploring the Statistics in Lotteries

Larry Lesser, Associate Professor
University of Texas at El Paso • *lesser@utep.edu*

Do you have a better chance of being struck by lightning or winning a six-number lottery? Any state or multi-state lottery offers multiple ways to connect real-world situations and mathematics. Educators who have explored this (Elliot 1993) have attracted much attention.

Numerous opportunities for critical thinking include analyzing what the prize structure is and why it is likely set up that way, and when strategy does and does not make a difference.

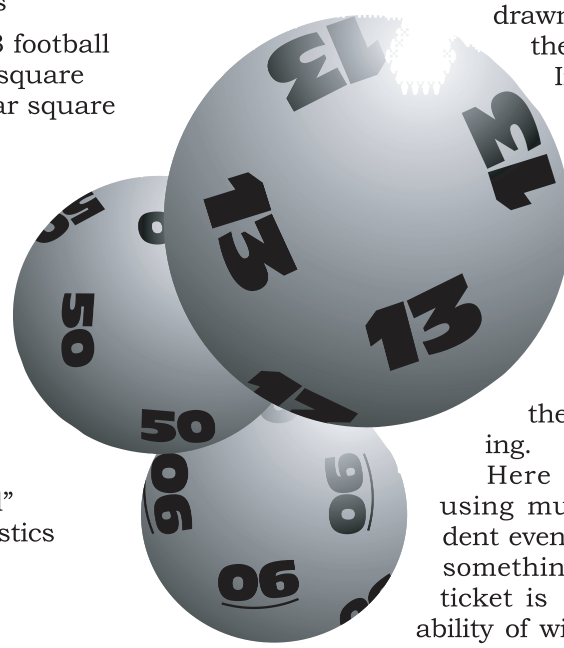
Also, understanding odds in the first place is a great opportunity for number sense, as many people of all ages seem to have a hard time visualizing very small or very large numbers.

For example, a person who purchases one ticket in the Mega Millions game (VA, GA, IL, MD, MA, MI, NJ, NY, OH, WA) has a 1 in 135,145,920 chance of winning (at least a share of) the jackpot!

That same tiny probability is represented by each of these events

(Can you make up a new one?):

- predicting 27 out of 27 coin flips
- dividing a region the size of 16.3 football fields (including endzones) into square inches, and guessing a particular square inch from the region
- guessing a particular piece of typing paper from a stack of paper one-and-a-half times as high as Mount Everest!
- guessing a particular second in a 4.3 year period (or from 30,000 compact disks' worth of music!)



The lottery scenario can be "mined" throughout the usual topics of a statistics course:

Univariate Statistics

Explore the types of lotteries available and analyze them for the size of the jackpot, for the numbers on the balls drawn, for the number of drawings until the jackpot is won, for the number of tickets sold, etc. Many lotteries have such information accessible via their websites or we can simulate drawings.

Bivariate Statistics

Create scatterplots or lines of fit (and later, correlations) for the variables of "jackpot size" and "tickets sold". Or do this for "current winning numbers" and "how long ago those numbers were last picked".

Probability

The experimental (empirical) probabilities of each number drawn (actual drawing or simulated drawing) can be compared to the theoretical probability. We can simulate a single drawing (i.e., our own "Quick Pick"!) with the TI-83 by the sequence: MATH--> PRB--> randInt(1,50,6); ENTER. With some calculators, try APPS --> ProbSim --> Random Numbers; Set (Numbers: 6; Range: 1-50; Repeat: No; Draw). Students can do many 6-ball simulated drawings and summarize the results. Often they will note and focus on whatever number appears the most often (which motivates some explorations using distributions, as described in the next section).

The concept of dependent and independent events comes up very naturally. In some lottery games, 6 balls are drawn without replacement from balls numbered 1 to, say, 50. Thus, within a drawing, the probability that the number 13 is the next ball drawn IS affected by whether or not the preceding ball drawn was 13. In other lottery games, five balls are drawn from one bucket, but the sixth ball is drawn from a separate "bonus ball" bucket, and the result of the bonus ball bucket is independent from the results of the bucket that yielded the other five balls. Also, from week to week, one 6-ball drawing does not affect the results of another 6-ball drawing.

Here are some sample questions using multiplication rule for independent events: If the probability of winning something with an instant scratch-off ticket is 1 in 5, then what's the probability of winning at least one prize if you buy 5 scratch-off tickets? If the probability of winning something (even if just $3) in a 6-ball lottery drawing is 1/60, then what's probability you play 10 times and win nothing? What's the probability of playing Mega Millions 10,000 times (enough money to buy a car!) without ever winning the jackpot? A

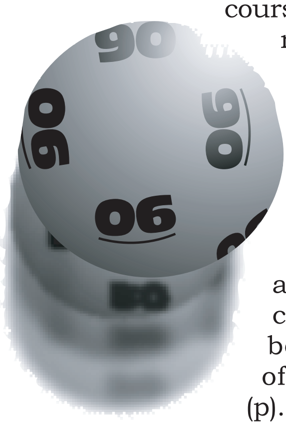
CRITICAL THINKING: Why might state lotteries (especially those with large or fast-growing populations) prefer a larger number of balls to draw from? How is this related to the likelihood of jackpots being unclaimed and rolling over into the next drawing's jackpot?

typical state lottery drawing draws 6 balls without replacement (and without order mattering) from balls numbered 1 to, say, 50. The combination coefficient "50 choose 6" = 50!/(6!(50-6)!)can be evaluated by the TI-83 by entering the following sequence: 50; MATH --> PRB --> nCr; 6; ENTER. We obtain 15,890,700 and the probability a ticket would match the jackpot set of numbers was therefore 1/15890700. If we made the game more challenging by choosing 6 balls from a set numbered 1 to 56, show that even though the number of balls increased only 12%, the number of combinations increased by over 100%!

Discrete Random Variables

Lesser (1997) shows how classes can explore lottery probabilities and expected values using spreadsheet technology. Using strategies such as those in Henze and Riedwyl (1998) allows us to show that even though one cannot increase the probability of a jackpot win, one CAN (through avoiding choosing combinations of numbers that are always popular) increase the expected value of a win.

Also, recall that for discrete probability distributions, the pdf (probability density function) gives you the probability that a random variable equals a specific value, while the cdf (cumulative distribution function) gives the probability a random variable is less than or equal to a specific value. While these concepts can be very abstract to students in their first statistics course, lottery drawings offer a concrete real-world application for most of the most commonly encountered discrete random variables. (Actually, drawing one number from a set of numbered balls has already illustrated the discrete uniform distribution!)



A binomial random variable is the number (x) of successes when there is a fixed number (n) of independent trials, each of which has a success probability (p). Entering values for n, p, and x (in that order) into the TI-83's binomcdf or binompdf commands (via the 2nd DISTR keys) can be used to investigate how rare it is for a particular lottery number to occur so frequently out of a fixed number of drawings. For example, in a "draw 6 balls from a set of balls numbered 1 to 50" game, the probability of the number 17 occurring exactly 3 times in 20 6-ball drawings would be binompdf (20, .12, 3), or about 22.4%.

A geometric random variable is the number x of trials (each of which is independent and has probability p of success) before the first success occurs. The geometpdf and geometcdf commands (inserting numbers for the values of p and x) can therefore be used to explore the probability that it takes a particular number of drawings before a drawing includes a particular number (or, for that matter, to finally win a jackpot). For the game mentioned in the above paragraph, the probability of no more than 50 drawings needed for the first occurrence of 17 is: geometcdf (.12, 50), which is about 99.8%.

Continuous Random Variables

The most common continuous probability distribution is the normal distribution. Beyond simply approximating the binomial distribution (which is less crucial when technology can deliver exact values easily), the normal distribution can be used to illustrate the central limit theorem. Suppose 6 numbered balls are drawn from a group of balls numbered 1 to n. The mean of those 6 balls will be approximately normal with mean $(n+1)/2$ and standard error $s/\sqrt{6} = \sqrt{[(n^2-1)/12]}/\sqrt{6} = (\sqrt{(n^2-1)})/(6\sqrt{2})$.

Inferential Statistics

Point and interval estimation and hypothesis testing can be done on variables previously listed under descriptive statistics. And the chi-square distribution can be used to explore the randomness of all the past Lotto numbers chosen, as did Lamb et.al.(1994).

You can bet that students will enjoy exploring the chance of becoming a millionaire. They may discover that they can more easily earn money in other ways and save themselves a few bucks in the process!

References:

Elliot, D. (1993, August 28), Professor seeks to even odds of lottery. Austin *American-Statesman,* B1, B3.

Henze, N. and Riedwyl, H. (1998). *How to win more: Strategies for increasing a lottery win.* Natick, MA: A.K. Peters.

Lamb, J.; Huffstutler, R.; Brock, A.; Aslan, F. (January 1994). A statistical analysis of the Texas Lottery. *Texas Mathematics Teacher,* 41(1), 4-17.

Lesser, L. (1997). Exploring lotteries with Excel. *Spreadsheet User*, 4(2) 4-7.

Dear Readers,

From the Editors:

We hope that you find this issue of interest. Please pass it or a copy on to a fellow teacher as we hope that everyone can discover something of interest.

The next full issue will center around the themes of inference and hypothesis testing. We would like to hear from you if you have a good idea and want to write a brief article or contribute an activity. Please let us know.

Sincerely,
Beth Lazerick, *stneditor@aol.com*
Murray Siegel, *siegel@gssm.k12.sc.us*

Send mailing address corrections and additions to: Madge Haven: *madge@amstat.org*

Let us know your Zip +4!

Statistics Teacher Network
c/o American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3415
USA

Non-Profit
Organization
U.S. Postage
PAID
Permit No. 361
Alexandria, VA