USING LOTTERIES IN TEACHING A CHANCE COURSE

Written by the Chance Team for the Chance Teachers Guide
revised August 1, 1998

Probability is used in the Chance course in two ways. First, it is used to help students understand issues in the news that rely on probability concepts. These include: chances of winning at the lottery, streaks in sports, random walk and the stock market, coincidences, evaluating extra sensory perception claims, etc. Second, a knowledge of elementary probability models, such as coin tossing, are necessary to understand statistical concepts like margin of error for a poll and testing a hypothesis.

Our goal here is to show how one can use current issues in the news and various activities to make students appreciate the role that probability plays in everyday news stories and to help them understand statistical concepts.

We begin by illustrating this in terms of the many interesting probability and statistics problems involved in lotteries.   Lotteries are discussed frequently in the news, and they have a huge impact directly and indirectly on our lives. They are the most popular form of gambling and an increasingly important way that states obtain revenue. In a Chance course, we do not give the systematic account presented here, but rather discuss a number of the points made in this presentation as they come up in the news.

THE POWERBALL LOTTERY

We will discuss lotteries in terms of the Powerball Lottery. The Powerball Lottery is a multi-state lottery, a format which is gaining popularity because of the potential for large prizes. It is currently available in 20 states and Washington D.C. It is run by the Multi-State Lottery Association, and we shall use information from their web homepage, http://www.musl.com. We found their "Frequently Asked Questions," (hereafter abbreviated FAQ) to be particularly useful. These are compiled by Charles Strutt, the executive director of the Association.

A Powerball lottery ticket costs $1. For each ticket you are asked mark your choice of numbers in two boxes displayed as follows:

Pick 5								EP__
01	02	03	04	05	06	07	08	09
10	11	12	13	14	15	16	17	
18	19	20	21	22	23	24	25	
26	27	28	29	30	31	32	33	
34	35	36	37	38	39	40	41	
42	43	44	45	46	47	48	49	

cash ☐

annuity ☐

Pick 1						EP__
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	32	33	34	35
36	37	38	39	40	41	42

Table 1: Picking your numbers.

You are asked to select five numbers from the top box and one from the bottom box. The latter number is called the "Powerball". If you check EP (Easy Pick) at the top of either box, the computer will make the selections for you. You also must select "cash" or "annuity" to determine how the jackpot will be paid should you win. In what follows, we will refer to a particular selection of five plus one numbers as a "pick."

Every Wednesday and Saturday night at 10:59 p.m. Eastern Time, lottery officials draw five white balls out of a drum with 49 balls and one red ball from a drum with 42 red balls. Players win prizes when the numbers on their ticket match some or all of the numbers drawn (the order in which the numbers are drawn does not matter). There are 9 ways to win. Here are the possible prizes as presented on the back of the Powerball ticket:

You Match	You win	Odds
5 white balls and the red ball	JACKPOT[*]	1 in 80,089,128
5 white balls but not the red ball	$100,000	1 in 1,953,393
4 white balls and the red ball	$5,000	1 in 364,041
4 white balls but not the red ball	$100	1 in 8879
3 white balls and the red ball	$100	1 in 8466
3 white balls but not the red ball	$7	1 in 206
2 white balls and the red ball	$7	1 in 605
1 white ball and the red ball	$4	1 in 118
0 white balls and the red ball	$3	1 in 74

Table 2: The chance of winning.

CALCULATING THE ODDS

The first question we ask is: how are these odds determined? This is a counting problem that requires that you understand one simple counting rule: if you can do one task in n ways and, for each of these, another task in m ways, the number of ways the two tasks can be done is $n \times m$. A simple tree diagram makes this principle very clear.

When you watch the numbers being drawn on television, you see that, as the five winning white balls come out of the drum, they are lined up in a row. The first ball could be any one of 49. For each of these possibilities the next ball could be any of 48, etc. Hence the number of possibilities for the way the five white balls can come out in the order drawn is $49 \times 48 \times 47 \times 46 \times 45 = 228{,}826{,}080$.

But to win a prize, the order of these 5 white balls does not count. Thus, for a particular set of 5 balls all possible orders are considered the same. Again by our counting principle, there are $5 \times 4 \times 3 \times 2 \times 1 = 120$ possible orders. Thus, the number of possible sets of 5 white balls not counting order is $228{,}826{,}080 / 120 = 1{,}906{,}884$. This is the familiar problem of choosing a set of 5 objects out of 49, and we denote this by $C(49,5)$. Such numbers are called binomial coefficients. We can express our result as:

$$C(49,5) \;=\; \frac{49!}{5!\,44!} \;=\; \frac{49 \times 48 \times 47 \times 46 \times 45}{5 \times 4 \times 3 \times 2 \times 1}$$

Now for each pick of five white numbers there are 42 possibilities for the red Powerball, so the total number of ways the winning six numbers

[*] The official Lottery explanation for the Jackpot is: "Select the cash option and receive the full cash amount in the prize pool. Select the annuity option and we will invest the money and pay the annuity amount to you over 25 annual payments." The cash payment is typically 50-60% of the total dollar amount paid over 25 years.

can be chosen is $42 \times C(49,5) = 80,089,128$. We will need this number often and denote it by b (for big).

The lottery officials go to great pains to make sure that all b possibilities are equally likely. So, a player has one chance in 80,089,128 of winning the jackpot. Of course, the player may have to share this prize.

We note that the last column in Table 2 is labeled "odds" when it more properly describes the "probability of winning". Because the probabilities are small, there is not much difference between odds and probabilities. However, this is a good excuse to get the difference between the two concepts straightened out. The media prefers to use odds, and textbooks prefer to use probability or chance. Here the chance of winning the jackpot is 1 in 80,089,128, whereas the odds are 1 to 80,089,127 in favor (or 80,089,127 to 1 against).

To win the $100,000 second prize, the player must get the 5 white numbers correct but miss the Powerball number. How many ways can this be accomplished? There is only one way to get the set of five white numbers, but the player's Powerball pick can be any of the 41 numbers different from the red number that was drawn. Thus, the chance of winning second prize is 41 in 80,089,128; rounded to the nearest integer this is 1 in 1,953,393.

This is a good time to introduce the concept of independence. You could find the probability of winning the second prize by pointing out the probability that you get the 5 white numbers correct is $1/C(49,5)$. The chance of not getting the red ball correct is $41/42$. Since these events are independent, the chance that they both happen is the product of their individual probabilities.

We can also point out that the lottery numbers you pick are independent of those drawn to determine the winning numbers. On the other hand, your picks and those of other buyers cannot be assumed to be independent.

Discussion Question: Why not?

Prior to November 2, 1997, the Powerball game was conducted by drawing 5 white balls from a drum of 45 and one red powerball from a second drum of 45. The prize for getting the red ball correct was $1, and the ticket listed the chances as 1 in 84. This often seemed wrong to players who have had elementary probability as the following exchange from the Powerball FAQ[*] illustrates:

---

[*] From the Multi-State Lottery Association web site at   http://www.musl.com/

COULD YOUR ODDS BE WRONG?

I have a simple question. You list the odds of matching only the powerball as one in 84 on the powerball "ways to win" page. From my understanding of statistics (I could be wrong, but I got an A), the odds of selecting one number out of a group is simply one over the number of choices. Since there are not 84 choices for the powerball, may I assume the balls are somehow "fixed" so that some are more common than others? Otherwise, the listed odds are somehow defying the laws of statistics. I am really very eager to hear your explanation, so please return my message. Thank you.
Susan G., via Internet.

This is one of most common questions we get about the statistics of the game. If you could play only the red Powerball, then your odds of matching it would indeed be 1 in 45. But to win the $1 prize for matching the red Powerball alone, you must do just that; match the red Powerball ALONE. When you bet a dollar and play the game, you might match one white ball and the red Powerball. You might match three white balls and the red Powerball. To determine the probability of matching the red Powerball alone, you have to factor in the chances of matching one or more of the white balls too.
C.S.

To win this last prize you must choose your six numbers so that only the Powerball number is correct. In the older version of the Powerball lottery this would be done as follows: there are $45 \times C(45,5) = 54,979,155$ ways to choose your six numbers. But here your first 5 numbers must come from the 40 numbers *not* drawn by the lottery. This can happen in $C(40,5) = 658,008$ ways. Now there is only one way to match the Powerball number, so overall you have 658,008 chances out of 54,979,155 to win this prize. This reduces to 1 chance in 83.55, or about 1 chance in 4, in agreement with the official lottery pronouncement.

The same kind of reasoning of course carries over to the present version of the game. To find the chance of winning any one of the prizes we need only count the number of ways to win the prize and divide this by the total number of possible picks b. Let $n(i)$ be the number of ways to win the ith prize. Then the values of $n(i)$ are shown in Table 3 below.

Number of ways	Match
$n(1) = 1$	all six balls
$n(2) = 41$	5 white balls but not the red ball
$n(3) = C(5,4) \times C(44,1)$	4 white balls and the red ball
$n(4) = n(3) \times 41$	4 white balls but not the red ball
$n(5) = C(5,3) \times C(44,2)$	3 white balls and the red ball
$n(6) = n(5) \times 41$	3 white balls but not the red ball
$n(7) = C(5,2) \times C(44,3)$	2 white balls and the red ball
$n(8) = C(5,1) \times C(44,4)$	1 white ball and the red ball
$n(9) = C(44,5)$	only the red ball

Table 3: How many ways can you win a particular prize?

Dividing these numbers by b, we obtain the chance of winning the corresponding prizes given in Table 2. Adding all the of $n(i)$ values gives a total of 2,303,805 ways to win something. Thus we get an overall chance of winning of $2,303,805/b = 0.02877$, which is about 1 in 35.

In a textbook, we would be apt to give the results of Table 2 as:

You Match	You win	Probability of Winning
5 white balls and the red ball	JACKPOT	0.000000012
5 white balls and not the red ball	$100,000	0.000000511
4 white balls and the red ball	$5,000	0.000002746
4 white balls and not the red ball	$100	0.000112624
3 white balls and the red ball	$100	0.000118118
3 white balls and not the red ball	$7	0.004842854
2 white balls and the red ball	$7	0.001653657
1 white ball and the red ball	$4	0.008474995
0 white balls and the red ball	$3	0.013559992

Table 4: The probabilities of winning.

As noted earlier, rounding the reciprocals of these probabilities to the nearest integer gives the numbers reported as "odds" on the lottery ticket.

Discussion Question: Which of the two methods for presenting the chances of winning, Table 2 or Table 4, do you think is best understood by the general public? Which do you prefer?

WHAT IS YOUR EXPECTED WINNING FOR A $1 TICKET?

The value of a gambling game is usually expressed in terms of the player's expected winning. If there are n prizes and $p(i)$ is the probability of winning the ith prize $w(i)$, then your expected winning is:

$$E = w(1){\times}p(1) + w(2){\times}p(2) + \cdots + w(n){\times}p(n)$$

For all prizes, except the jackpot, we can assume we know the value of the prize. However, since the size of the jackpot differs significantly from drawing to drawing, we will want to find the expected winning for different jackpot sizes. In the 508 drawings from the beginning of the lottery on April 22, 1992 through March 1, 1997 the jackpot was won 75 times. It was shared with one other winner 11 times. During this period the jackpot prize varied from 2 million dollars to $111,240,463.

If x is the amount of the jackpot and $p(i)$ the probability of winning the ith prize, the expected winning is:

$$E = x \times p(1) + 100{,}000 \times p(2) + 5000 \times p(3) + 100 \times p(4) + 100 \times p(5)$$
$$+ 7 \times p(6) + 7 \times p(7) + 4 \times p(8) + 3 \times p(9) = x/b + 0.208$$

where $b = 80{,}089{,}128$. Using this, we can find the expected winning for various values of the jackpot.

x = Jackpot ($ millions)	E = Expected Winning ($)
10	0.333
20	0.458
30	0.583
40	0.707
50	0.832
60	0.957
70	1.082
80	1.207
90	1.332
100	1.457
110	1.581
120	1.706
130	1.831
140	1.956
150	2.081

Table 5: Expected winning for different size jackpots.

A game is said to be *favorable* if the expected winning is greater than the cost of playing. Here we compare with the $1 cost of buying a ticket. Looking at Table 5, we see that the lottery appears to be a favorable game as soon as x gets up to $70 million.

The jackpot for the Powerball lottery for July 29, 1998 built up to some $295.7 million, as hordes of players lined up at ticket outlets for a shot at what had become the largest prize for any lottery in history. At first glance, it certainly looks like this was a favorable bet!

However, we need to recall that on the ticket itself, the player had to indicate his choice of cash or annuity to be used in the case he wins the jackpot. In fact, the Powerball web site regularly updates the value of the jackpot in each format. At this writing (1 August 1998), we find the report

Next Powerball Jackpot Estimate

Saturday, August 1, 1998 $10 Million ($5.5Million-cash option)

Select the cash option and receive the full cash amount in the prize pool. Select the annuity option and we will invest the money and pay the annuity amount to you over 25 annual payments.

The $10 million here is analogous to the $295.7 million from July 29, and is the number that the media likes to hype. But note that this corresponds to the annuity amount to be paid out over time, not the immediate cash value. Not only are you not going to get this money tomorrow--the lottery doesn't even have it on hand! This is explained further in an earlier excerpt from the FAQ:

> When we advertise a prize of $20 million paid over 20 years, we actually have about $12 million in cash. When someone wins the jackpot, we take bids to purchase government securities to fund the prize payout. We take the $12 million in cash and buy U.S. government-backed securities to fund these payments. We buy bonds which will mature in one year at $1 million, then bonds which will mature in two years at $1 million, etc. Generally, the longer the time to maturity, the cheaper the bonds.

The cash option on the $295.7 million from July 29 was $161.5 million. From Table 5, we see that this still looks like a favorable bet, with expected value of about 2.

We have been assuming that the player has elected the lump sum cash payment, and treating the annuity as equivalent in present value terms. You may want to think harder about this. An article in the *Star Tribune* by Julie Trip (June 7, 1998, Metro section p. 1D) discusses the question of lump sum or annuity. The article is based on an interview with Linda

Crouse, a financial planner and certified public accountant in Portland, Oregon. From this article we read:

Crouse ran numbers to help determine whether it's better to take a windfall in payments over time - an annuity - or in a lump sum.

Crouse used the Powerball jackpot as an example to determine which pays off in the long run: the ticket that pays the winner $ 104.3 million now or pays $ 7.7 million annually for 25 years. (Both are before taxes.)

The annuity represents a 5.4 percent return. That sounds easy to beat if you take the lump sum and invest it - until you consider the huge negative effect of paying all the taxes up front instead of over 25 years. Figure 45 percent of the payout - $ 46.9 million - goes to state and federal taxes right off the bat

If you invest the remaining $ 57.4 million and receive an average return of 8 percent, you still can't beat the annuity. After all taxes are paid, you receive $ 4,235,000 annually for the annuity vs. $ 3,991,000 for the lump sum you invested at 8 percent.

Beyond about a 9 percent return, you start to beat the annuity

---

Of course, one should consider the fact that the annuity is a guaranteed payment while your investments are subject to the volatility of way you invest your money.

Well, at least with the lump sum above, we convinced ourselves that we had a favorable game. Alas, there is another rub. We have been implicitly assuming that if we hold the lucky numbers, we will get the whole prize! But if other ticket holders have selected the same numbers, the jackpot will be split. This will be a particularly important factor when large number of tickets are sold. As the jackpot grows, an increasing number of tickets are sold. For the July 29, 1998 the Jackpot was a new record 295.7 million dollars and 210,800,000 tickets were sold.

The chance of having to share the pot depends upon whether you chose the easy pick or chose your own numbers. This is because the easy pick numbers are all equally likely to occur but, as we shall see, this is not the case when people make their own choices.

We are told that about 70% of tickets sold in a typical lottery are chosen by the easy pick method. Probably this percentage is even larger when the jackpot is large since people tend to buy a number of tickets and would be more likely to use the easy pick method when they do this. We will limit our calculation to the easy pick tickets which we estimate to be 70% of 210,800,000 or 147,560,000.

The probability that a particular ticket is the winning ticket is 1/80,089,128.   The probability of k winners can be obtained from a binomial distribution with p = 1/80,089,128 and n = 147,560,000. The expected number of winning tickets among those who choose the easy pick is this np = 147,560,000/80,089,128 = 1.84. Since p is small and n is large we can use the Poisson approximation:

$$p(k) \ = \ \frac{m^k}{k!} \ e^{-m} \ , \quad k = 0,1,2,...$$

where m = 1.84. Carrying out this calculation gives:

No. Winners	Probability
0	.159
1	.292
2	.269
3	.165
4	.076
5	.028
6	.009
7	.002
8	.001
9	0

From this we find that the probability, given that there is at least one winner, that the winner is the only winner = 292/(1-.159) = .347. Thus the probability that the winner has to share the prize is .653.

Recall that the cash value of the July 29 jackpot was 161,500,000. Using the probability p(k) that we will have to share this with k others, we can find the expected amount that a player, with the winning numbers, will end up with. We need only sum the values (161,500,000)/k)*p(k) for k = 0 to 12. Carrying out this calculation we obtain the expected value of a jackpot winning $82,039,300. Now looking again at Table 5 we find that we still have a favorable game with expected value 1.2.

Unfortunately, the government isn't about to let a lucky winner just walk off with that lump sum without paying taxes. One FAQ inquirer protested the fact that 28% of that payment is withheld in federal tax. In fact, the situation is worse, since some states will take out additional money to cover state tax as well! Here in New Hampshire (at this writing at least!), they will only take out 28%. Thus we finally have to take 72% of our expected winning of $82,039,300 obtaining $59,068,296. Going one last time to Table 5, we find that, even with this huge jackpot, we are still playing a slightly unfavorable game.

Perhaps we have now explained the famous quote:

"The lottery: A tax on people who flunked math."

-- Monique Lloyd

WHAT KIND OF NUMBERS DO LOTTERY BUYERS CHOOSE?

We have suggested that we might at least be able to avoid sharing the jackpot with people who choose their own numbers if we choose our own cleverly. It is well known that people who choose their own numbers do not choose randomly. They use their child's birthday, their "lucky" numbers, arithmetic progressions such as 2-4-6-8-10-12, numbers from sports, etc.

To see what players actually do, we obtained the numbers chosen by players in the Powerball Lottery in one state on May 3, 1996. Recall that at this time the game was played by selecting five of 45 white balls and one of 45 red balls. On this day, 17,001 of the picks were chosen by the buyers, and 56,496 were chosen by the computer (Easy Pick). Thus only about 23% of the picks were chosen by the buyers.

We first compare the distribution of the individual numbers from the picks chosen by the computer and those chosen by the buyers. To make the two sets the same size, we use only the first 17,001 picks produced by the Easy Pick method. Each pick contributed 6 numbers, so in both cases we have 102,006 numbers between 1 and 45. Here is a plot of the number of times each of the numbers from 1 to 45 occurred for the picks chosen by the computer:

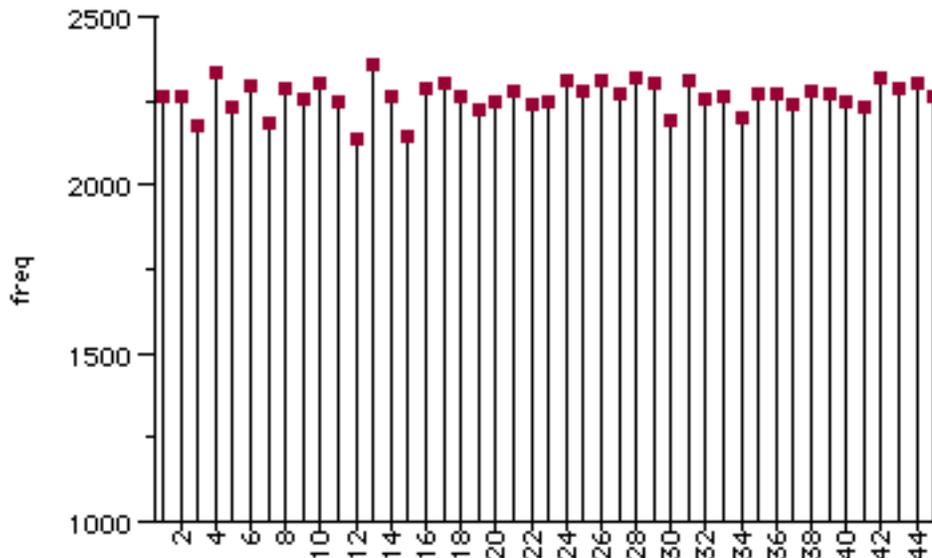## Numbers picked by computer

Figure 2.

There does not seem to be very much variation, but it is worth checking how much variation we would expect if the numbers were, in fact, randomly chosen. If they were, the numbers of occurrences of a particular number, say 17, would have a binomial distribution with $n =$ 102,006 and $p = 1/45$. Such a distribution has mean np and standard deviation $\sqrt{npq}$ where $q = 1-p$. This gives mean 2267 and standard deviation 47. It would be unusual to see differences of more than three standard deviations, or 141, from the mean. It is hard to tell the actual differences from the graph, so we looked at the actual data. We found that, for all but two numbers, the results were within two standard deviations of the mean. For the other 2 numbers the results were within 3 standard deviations of the mean. Thus the picks chosen by the computer do not appear to be inconsistent with the random model. A ChiSquare test would give a way to proceed with a formal test of this hypothesis.

We look next at the picks chosen by the players. Recall that we have the same number 17,001 of picks so we again have 102,006 individual numbers.
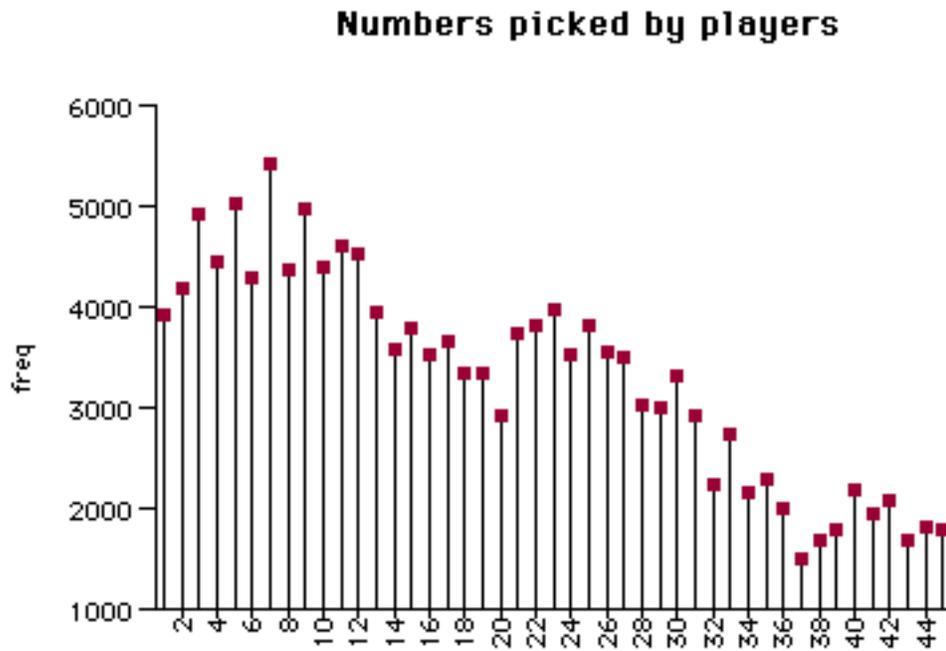
Numbers picked by players

Figure 3.

Here the numbers are in increasing order of frequency of occurance:

```
37     0.010
38     0.011
43     0.012
45     0.012
39     0.012
44     0.012
41     0.013
36     0.013
42     0.014
34     0.014
40     0.015
32     0.015
35     0.016
33     0.018
20     0.019
29     0.020
28     0.020
31     0.020
18     0.022
30     0.023
19     0.023
```

```
27     0.023
24     0.024
14     0.024
26     0.024
16     0.024
17     0.024
21     0.025
15     0.025
25     0.026
1      0.026
22     0.026
13     0.026
23     0.027
6      0.028
2      0.029
10     0.029
4      0.029
8      0.030
12     0.030
11     0.031
3      0.033
5      0.033
9      0.033
7      0.036
```

You don't have to do any fancy tests to see that these are not randomly chosen numbers. The most popular number 7 was chosen 3,176 times which would be 19 standard deviations above the mean if the numbers were randomly chosen!

It is often observed that people use birthdays to choose their numbers. If they did, we would expect numbers from 1 to 12 to be most likely to be picked since such numbers can occur both in the month and the day. The next most likely numbers to be picked would be those from 13 to 31 where the remaining days of the months could occur. The least likely numbers would then be those from 32 to 45 where the year of the birthday could occur but only for those at least 52 years old. Note that this is indeed what appears to be happening.

Finally, we look at the winning numbers to see if they could be considered to be randomly chosen. Recall that the lottery officials try to ensure that they are. Here we have many fewer numbers so we expect more variation even if randomly chosen.
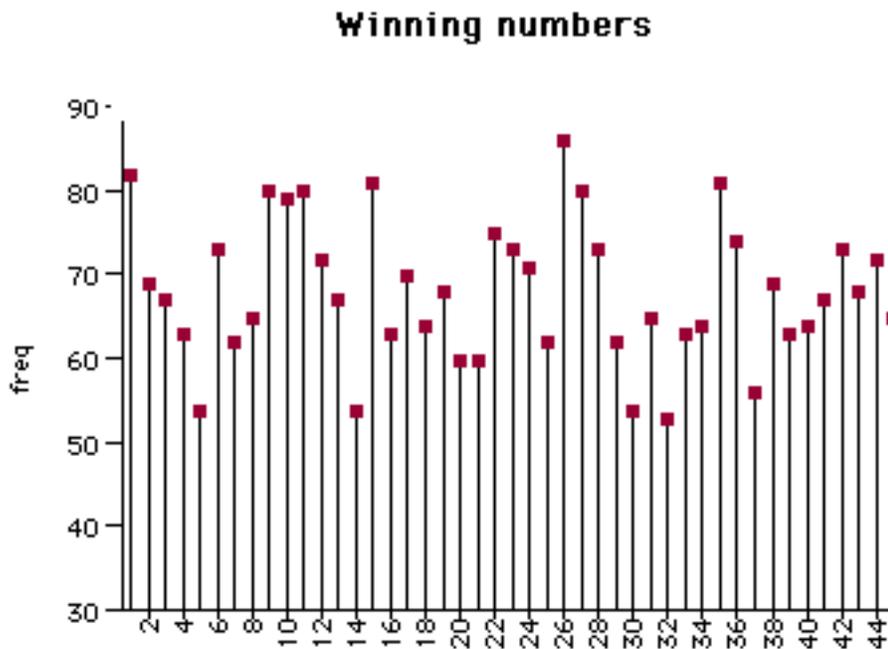
Figure 4.

Since there were 508 drawings in the period we are considering, we have
$6 \times 508 = 3048$ numbers. Now, if the numbers are randomly chosen, the
number of times a particular numbers occurs has a binomial distribution
with $n = 3048$ and $p = 1/45$. Such a distribution has a mean of 67.7 and
standard deviation 8.14. The biggest deviations from the mean are about
2 standard deviations so this looks consistent with the hypothesis that
the numbers were randomly chosen. Again, we could check this with a
ChiSquare test.

FINDING PATTERNS

Recall that players choose their first five numbers to match the white
balls separately from their choice of the Powerball number to match the
red ball. Thus, if we are looking for patterns in the way people choose
their numbers, it is best to consider the first five numbers by themselves.

For comparison, we again consider the 17,001 picks chosen by the Easy
Pick method and by the players. For the Easy Picks, we found that 136 of
these were represented twice and 2 were represented 3 times.

To see if we should have expected to find 3 picks the same, we again
apply a result from the appendix to estimate the number of people
required to have probability p of finding k or more with the same

birthday. We choose as the number of possible "birthdays" the number of possible first five numbers: C(45,5) = 1,221,759. We find that, for this number of possible birthdays and 17,001 people, the probability of finding 3 or more picks the same is .42. We also find that there is only a probability of .002 of finding 4 or more the same birthday. Thus we should not be surprised at finding 3 picks the same and should not expect to find 4 the same. Again, the computer picks seem to conform to random choices.

Not surprisingly, we found no obvious patterns in the sets of numbers that were repeated more than once in the sets of numbers chosen by the computer.

We look next at the 17,001 sets of 5 numbers chosen by the lottery players. We found 966 sets of numbers that were represented more than once (compared to 138 for the Easy Pick numbers). The largest number of times a particular set of numbers was chosen was 24. This occurred for the pick 02-14-18-21-39.   Looking at the order in which the numbers were given to us, we noticed that these occurred consecutively in blocks of 5, with the blocks themselves close together. The ticket on which you mark your numbers allowed room for 5 sets of numbers. We concluded that one player had made 24 picks all with the same five numbers. He at least chose different Powerball numbers. The same explanation applied to the next most popular pick 08-12-24-25-27, which occurred 16 times.

The third most popular set 03-13-23-33-43 was picked by 13 people and was more typical of the patterns that people chose. In this version of Powerball, the numbers were arranged on the ticket as shown below:

Pick 5								EP__
01	02	03	04	05	06	07	08	09
10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	27
28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45

Note that the pick 03-13-23-33-43 is an arithmetic sequence obtained by going down a diagonal starting with 03.   Similarly, the set of numbers 01-11-21-31-41, which was chosen 10 times, corresponds to going down a diagonal starting with 01 and the set 06-15-24-33-42, chosen 9 times, corresponds to going down a column starting with 06. The most interesting pattern noticed was 01-09-23-37-45, occurring 8 times, which results from choosing the corner points and the middle point. Since we do not expect repetitions of 4 or more to occur by chance, we looked at all those that occurred 4 or more times. We could explain all but three such sets of 5 numbers. These were:

01-03-09-30-34 (occurred 5 times, always with Powerball number 40)
05-06-16-18-23 (occurred 4 times, always with Powerball number 31)
02-05-20-26-43 (occurred 4 times with different Powerball numbers)

Here is a Chance News item related to the problem of people choosing popular numbers. The letters followed an article in *The Times* stating that the inaugural drawing of the new British Lottery had five times the number of winners expected, including seven people who had to share the jackpot. They blamed this on the fact that the six winning numbers 03-05-14-22-30-44 had five numbers under 31 and most people chose low numbers. In this lottery, you choose 6 numbers between 1 and 49 and have to get them all correct to win the jackpot. If you get three numbers correct you win £10. The amount you win for any other prize depends on the number of other people who win this prize.

The Times, 24 November 1994, letters to the editor.
Slim pickings in National Lottery
From Mr George Coggan

Sir, With random choices, the odds against there being seven or more jackpot winners in the National Lottery when only 44 million tickets have been sold are 23-1. This suggests that those who predicted that low numbers would be popular were right as the slightly disproportionate number of single digits (3 and 5 came up) would combine to produce more winners than would be produced by entirely random selections.

Mildly interesting, one might think, but then one suddenly realizes that there is a lurking danger that the rules create the possibility that when (as will happen sooner or later) three single digit numbers come up the prize fund may not be enough to cover the Pounds 10 guaranteed minimum prize, never mind a jackpot. I estimate that if the number 7 had come up instead of say 44 the prize fund in this first lottery would have been about Pounds 5 million short of the guarantee. What then panic?

Yours sincerely,
GEORGE COGGAN,
14 Cavendish Crescent North,
The Park, Nottingham.
November 21. 1994 Tuesday

The Times, 29 November 1994, letter to the editor.
No need to fear a lottery shortfall
From the Director General of the National Lottery

Sir, Mr George Coggan (letter, November 24) raises concerns about the National Lottery Prize Fund's ability to pay winners when "popular" numbers are selected in the weekly draw.

We are aware that players do not choose numbers randomly but use birthdays, sequences or other lucky numbers. This causes the number of winners to deviate each week from the number predicted by statistical theory. Experience from other lotteries shows that the number of winners of the lower prizes can vary by up to 30 per cent from the theoretical expectation.

In the first National Lottery game there were many more Pounds 10 prize-winners than theory predicted. It is just as likely that future draws will produce fewer than expected winners and, because each higher prize pool is shared between the winners, prize values will rise accordingly.

Mr Coggan suggests a pessimistic scenario in which the cost of paying the fixed Pounds 10 prizes to those who choose three correct numbers exceeds the prize fund. Best advice, and observations from other lotteries around the world, is that, even after allowing for the concentration on "popular" numbers, the chances of this happening are extremely remote.

Your readers will be reassured to know, however, that I have not relied totally upon statistics or evidence from other lotteries. Camelot's license to operate the National Lottery also requires them to provide substantial additional funds by way of deposit in trust and by guarantee to protect the interests of the prize-winners in unexpected circumstances.

	Yours faithfully,
	PETER A. DAVIS,
Director General,	National Lottery,
	PO Box 4465,
	London SW1Y 5XL.
	November 25.

Of course, it is interesting to look at this problem for the Powerball lottery.   We noted that, in our sample of 17,000 numbers where players picked their own numbers, there were particular sets of five numbers for the white balls that were chosen as many as 10 times. For example the set of numbers 01-11-21-31-41 obtained by going down a diagonal starting

at 1 in the box where you mark your numbers was chosen 10 times in our sample of 17,000.

Recall that in the July 29 drawing there were 210,800,000 tickets sold. This suggests that in this drawing there 63240000/17000 = 3720 players might choose this same set of five white numbers. If the lottery officials had the bad luck to also choose it this would cost them 372 million dollars! The new boxes are not as symmetric as the old ones that our data applied to. This may help them with this potential problem. Of course, the real thing that will save them is that they are very unlikely to choose a popular set of numbers.

HOW OFTEN IS THE JACKPOT WON?

The size of the jackpot changes from one drawing to the next. If, on a given drawing, no one chooses the winning numbers, the jackpot is increased several million dollars for the next drawing. When there is a winner, the jackpot goes back to the minimum amount, which currently is 5 million dollars. The size of the increase when there is no winner depends upon the number of tickets sold for the previous drawing.   We investigate the size of the jackpots through the years of the original rules.

We find, on the Powerball homepage, the amounts of the jackpot in all the drawings under the original rules. The jackpots, from the beginning of the Powerball lottery on April 22, 1992 until March 1, 1997, range from 2 million to 111,240,463. The jackpot was won in 75 of these 805 drawings. 11 of these times there were two winners and never more than two winners. The total of all these jackpots was $2,206,159,204 with an average of $29,415,456. The average number of drawings between jackpots being won was 6.72 or, since there are two drawing a week, about 3 weeks. Here is the distribution of times between jackpots:
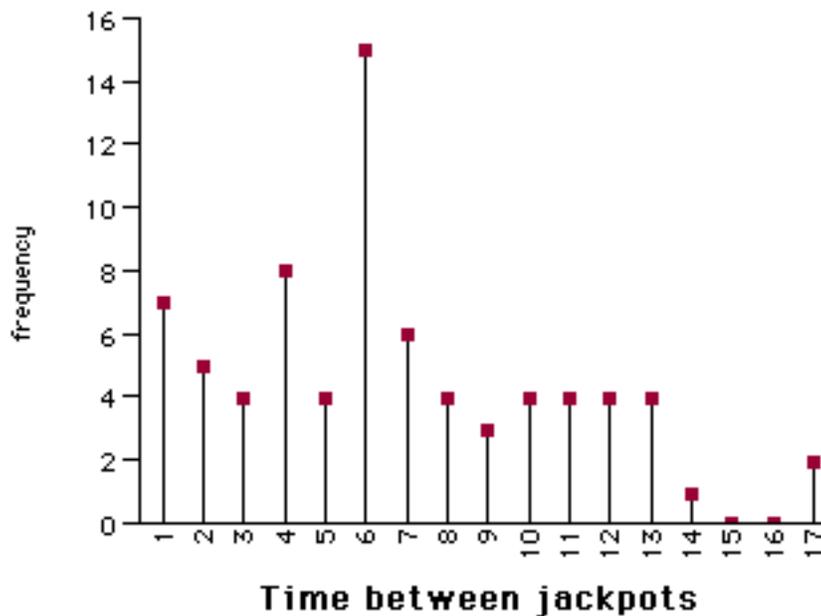
Figure 5.

DISCUSSION QUESTION: The mode 6 seems rather over represented. Can you think of any explanation for this, or is it just chance.

It is interesting to ask what kind of a probability model would be appropriate to describe the time between winning the jackpot. The probability of the jackpot being won depends on the number of tickets sold. (Actually, on the number of different picks chosen.) If the same number of tickets were sold on each drawing then the appropriate model would be: toss a coin with probability p for heads until the first time heads turns up where $1/p$ is the average time between heads. Unfortunately, it is not at all reasonable to assume the same number of tickets will be sold. Here is what the Powerball FAQ says about this:

> For a $10 million jackpot draw we sell about $11 million. For a $20 million jackpot we sell about $13 million. With a $100 million jackpot we sell $50 to $70 million for the draw (depending on time of year and other factors).

Let's assume that, for a particular jackpot, n million tickets are sold. Then the probability that a particular person does not win the jackpot is $\frac{b-1}{b}$ where, for the old version of the game, $b = 45 \times C(45,5) = 54{,}979{,}155$.

The probability that none of the tickets sold wins the jackpot is $\left(\frac{b-1}{b}\right)^{n}$.

Here is a table of these probabilities for the different values of the number of tickets sold.

Millions of tickets sold	Probability no one wins
10	.834
20	.695
30	.579
40	.483
50	.403
60	.336
70	.280

Table 8: Probability no one wins the jackpot.

From these we might be able to make a reasonable model where the probability that a head turns up each time increases after each tail.

ESTIMATING THE TOTAL NUMBER OF TICKETS SOLD

From the Multi-State Lottery homepage, we find the number of winners for all the prizes during this time period. (There are times when the prizes are doubled as a promotional feature.)

Match	Prize	Winners	Double winners	Total winners	Amount won
5+1	millions	86	0	86	$2,206,159,204
5	$100,000	3540	11	3,551	356,200,000
4+1	$5000	15,463	50	15,513	77,815,000
4	$100	706,839	2,182	709,021	71,120,300
3+1	$100	618,832	1,916	620,748	62,266,400
3	$5	27,334,931	85,375	27,420,306	137,528,405
2+1	$5	7,825,514	24,119	7,849,633	39,368,760
1+1	$2	36,107,707	109,877	36,217,584	72,654,922
0+1	$1	51,928,026	157,755	52,085,781	52243536
Total				124,922,137	$3,075,356,527

Table 9: The prizes won up to March 1, 1997.

An important bit of information we do not find on the Lottery's web page is the total number of tickets sold. However, we can estimate this from the information we have. We know the probability of winning each prize from Table 4. Adding these we find the probability of winning at least one prize to be $p = .0286$. From Table 9 we see that the total number of people who won a prize in the first 508 drawings was $w = 124{,}922{,}137$. If n tickets were sold then we would expect w to be approximately np. Thus we can estimate n by w/p. This gives an

estimate of 4,364,023,663 for the number of tickets sold during this period.

As a check on our estimate for n, we can calculate the expected number of each type of prize and see if these are reasonably close to the actual number of winners for each of the 9 prizes. The results of doing this are shown in Table 10.

Prize	Expected	Observed
5+1	79	86
5	3493	3551
4+1	15875	15513
4	698509	709,021
3+1	619133	620,748
3	27,241,832	27,420,306
2+1	7,842,346	7,849,633
1+1	36,270,849	36,217,584
1	52,230,022	52,085,781

Table 10: Observed vs. expected number of prizes won.

We see that the observed numbers fit the expected numbers quite well.

Another check would be to see if the amount paid out is the same as the amount taken in. We have the following statement from the Powerball FAQ

The Multi-State Lottery Association (which administers the PowerBall game) is a non-profit government-benefit association entirely owned and operated by the member state lotteries. PowerBall is a 50% prize payout game which means that 50 cents of every one dollar ticket is paid out in prizes. The state lottery keeps 50% as its share and then pays the rest out in prizes. The state lottery pays the cash prizes directly to the players in its state and then sends the percentage share for the jackpot prize back to the association where we hold it until there is a winner. The state that sells a ticket keeps ALL of the profit from that ticket. The only money that is sent to the central organization is the money to pay the jackpot prize.

From this we see that the amount paid out should be 50% of the amount taken in. To find the amount paid out we have to remember that, for the jackpot, the lottery officials do not need the total amount of the jackpot. They indicate, for example, that for a 20 million dollar jackpot they only need 12 million dollars. Based on this, we have assumed that they need about 60% for any jackpot. If we use this, we find that the total prizes have cost the lottery a total of 2,192,892,845 dollars. Our estimate for the amount taken in (number of tickets sold) was

4,364,023,663. Thus our estimate would say that the lottery has paid out 50.2% of the amount they have taken in. This is pleasantly close to the 50% that the lottery officials aim for. It appears that the whims of fate even suggest that the percentage should be slightly larger than 50%.

Here is what the lottery officials themselves say on their FAQ about how they have done.

> Have you had to pay more or less than what you expected statistically since the start of Powerball?
> Darin H., via Internet.
>
> The players have been luckier than the game design says they should be. To date (April 24, 1996), we have paid out $14,050,085 in "extra" prize money to winners - that is, prizes above what we would expect to pay out if people would pay more attention to the statistics. The total prizes paid out to date is over $722 million, so the difference is not significant (unless it is your $14 million!).

OTHER LOTTERIES POSE NEW QUESTIONS

There are many other interesting questions that can be explored about lotteries. The questions that one asks depend, to some extent, on the nature of the lottery. For example, in September 1996 the Multi-State lottery introduced a new lottery called Daily Millions where the amount of the jackpot is always $1 million dollars and, if you win, you don't have to share it with another person who also has the same winning pick (actually, if there are more than 10 such winners they share a 10 million dollar prize.) Here is an article from Chance News raising problems about this lottery.

> **Daily Millions beats odds: no one wins -- 5-month losing streak puzzles even statisticians.**
> *Star Tribune*, 7 Feb. 1997, 1B
> Pat Doyle
>
> The Daily Millions lottery was started nearly five months ago and at the time of this article, 34 million tickets had been sold without a single $1 million jackpot. It is stated that one would expect 3 or 4 winners by now and the probability of having no winners in this period is put at 1/38.
>
> The Daily Millions lottery is run by Multi-State Lottery Association which also runs the Powerball lottery. The Daily Millions was invented to give a lottery where you are do not

have to share the jackpot with other winners. You also are about 6 times more likely to win the Jackpot in the Daily Millions lottery than in the Powerball lottery.

However, the fact that no one has won the jackpot yet has hurt the sales. The Daily Millions has slumped from $3.75 million in the first week of the lottery to $1.23 million in the week ending Feb. 1.

Despite not having to pay out any jackpots, the participating states are required to set aside 11 percent of the ticket sales to be put in a pool for future winners. Charles Strutt, executive director of the Multi-State Lottery Association, said the money piling up in the jackpot pool will come in handy if players beat the odds in a different way, winning more than expected.

P.S. On Saturday February 8, the Daily Millions lottery had its first winner.

USING LOTTERY STORIES TO DISCUSS COINCIDENCES

James Hanley[*] has discussed how stories about lottery winners provide good examples to discuss the meaning of apparent coincidences. Here is his first example:

From the *Montreal Gazette* on September 10, 1981

Same Number 2-state Winner

Boston (UPI) -- Lottery officials say that there is 1 chance in 100 million that the same four-digit lottery numbers would be drawn in Massachusetts and New Hampshire on the same night. That's just what happened Tuesday.

The number 8092 came up, paying #5,482 in Massachusetts and $4,500 in New Hampshire. "There is a 1-in-10,000 chance of any four digit number being drawn at any given time," Massachusetts Lottery Commission official David Ellis said. "But the odds of it happening with two states at any one time are just fantastic," he said.

---

[*] Hanley, James A. "**Jumping to Coincidences,** "*The American Statistician*, **Vol 46, No. 3, pp 197-201.**

To assess the likelihood of this happening, should we find the probability that some two such lotteries have the same two numbers during a given period of time? Is this different from a reporter noticing that the number that turned up in the lottery in New Hampshire on Wednesday happened also to occur in the Massachusetts lottery on Saturday?

Here is another of his examples:

From the *New York Times* of February 14, 1986.

**Odds-Defying Jersey Women Hits
Lottery Jackpot 2d Time**

Defying odds in the realm of the preposterous--1 in 17 million--a women who won $3.9 million in the New Jersey state lottery last October has hit the jackpot again and yesterday laid claim to an addition $1.5 million prize...

She was the first two time million-dollar winner in the history of New Jersey's lottery, state officials said. They added that they had never before heard of a person winning two million-dollar prizes in any of the nation's 22 state lotteries.

For aficionados of miraculous odds, the numbers were mind boggling: In winning her first prize last Oct. 24, Mrs. Adams was up against odds of 1 in 3.2 million. The odds of winning last Monday, when numbers were drawn in a somewhat modified game, were 1 in 5.2 million.

And after due consultation with a professor of statistics at Rutgers University, lottery officials concluded that the odds of winning the top lottery prize twice in a lifetime were 1 in about 17.3 trillion--that is, 17,300,000,000,000.

Does it matter that she played the lottery many times often buying more than one ticket? Again, are we talking about this happening somewhere, sometime? Should we ever believe that something with these odds has happened?

LOTTERY SYSTEMS

Richard Paulson[*] observes that claims made about systems for improving your chances at lotteries illustrate important statistical concepts. For example, people claim that, by analyzing the historical data of winning numbers, it is possible to predict future winners. Indeed,

---

[*] Paulson, Richard A. "Using Lottery Games to Illustrate Statistical Concepts and Abuses", *The American Statistician* , **Vol. 45, No. 3, pp. 202-204.**

lottery sites encourage this by making this historical data available. Sometimes the argument is simply that, when a particular number has not turned up as often as would be expected, then this number is more likely to come up in the future. This is often called the "gambler's fallacy" and all too many people believe it. The fact that it is not true is the basis for many beautiful applications of chance processes called Martingales.

Paulson remarks that he particularly enjoys discussing the following system. Consider, the six winning numbers in the Powerball Lottery. If they occur randomly their sum should be approximately normally distributed with mean $6(1+45)/2 = 138$ and standard deviation approximately 32. Thus, sets of six numbers whose sum is more than 64 away from the mean 138, are not likely to occur as winning sets and should be avoided. It is better to pick six numbers whose sum is near 138.

We leave the reader to ponder this last system and with the following advice which paraphrases the advice of our teacher Joe Doob. Play the lottery once. Then wait until there has been a drawing without a winner and play again. Then wait until there have been two drawings without a winner and play again. Continue in this manner. You will enjoy playing and not lose too much.

APPENDIX

THE BIRTHDAY PROBLEM

Note: We wrote this when we thought the birthday problem would help settle useful computations relating to the lottery. We did not use it as much as we thought we would so we leave it as an appendix.

The birthday problem asks: what is the necessary number of people in a room to make it a favorable bet that two people have the same birthday? The surprising answer is 23. To show this, we compute the probability that there is no duplication of birthdays in a room with 23 people. Since there are 365 possibilities for each person's birthday, our familiar counting principle shows that there are 365^{23} possible assignments of birthdays for the 23 people. How many of these assignments give all different birthdays? For this to happen, the first person can have any of 365 possible birthdays; but, for each of these, the second person must have one of 364 possible birthdays; and then the third person one of 363; etc. Hence the probability of no match is

$$\frac{365 \times 364 \times 363 \times \quad \times 343}{365^{23}} = 0.4927.$$

Thus, there is a slightly better than 50% chance for a match. To be sure that 22 would not work, we should calculate, in the same manner, the probability of duplicate birthdays in a group of 22 people. If we do this, we find the probability of finding no matches to be .5243 showing that, with 22 people, there is less than a 50% chance of a match.

This problem is often confused with the following related problem: if you are in a room with 23 other people, what is the chance that at least one other person in the room has your birthday? The chance that any one person had a different birthday than yours is 364/365. Hence the probability that all 23 other people have a different birthday than you do is

$$(364/365)^{23} = 0.939.$$

Thus, while there is a better than 50% chance that two people in the room have the same birthday, there is only about a 6% chance that someone else has your birthday. We note that the birthday provides a nice point of departure for a study of coincidences in general. An excellent introduction to techniques in this area is given by Diaconis and Mosteller[*].

We have assumed that all birthdays are equally likely to occur. This is not quite true. Here is a graph the summarizing the frequency of birthdays in the United States (1978-1987) given by Thomas Nunnikhoven.[**] For each month, we have plotted the average daily frequency for days in that month. If the distribution for birthdays were uniform, then (ignoring leap years) each frequency would be $1/365 = 0.00274$.

---

[*] Methods for studying coincidences," *Journal of the American Statistical Association*, **84**, 853-861.

[**] A birthday problem solution for non-uniform birth frequencies, *The American Statistician*, **Vol. 46, No. 4 pp. 270-273.**
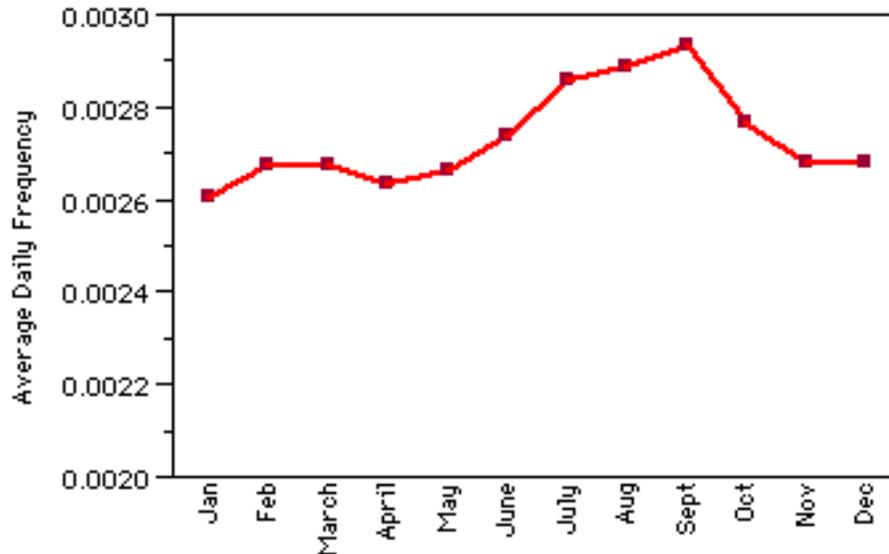
Birthday Distribution

Figure 1.

Nunnikhoven finds that the probability of a match with 23 people, using this actual distribution, agrees with value found for the uniform distribution to three decimal places. It can be shown that a non-uniform distribution of birthdays can only increase the probability for a match with 23 people.

To apply results on the birthday problem to the lottery, we should think of living on a planet where there are $b = 80,089,128$ possible birthdays. We must live a very long time! Here birthdays correspond to picks of the six numbers. Assuming all birthdays equally likely corresponds to assuming that all picks are equally likely. This will be true for the group who use the Easy Pick method (Powerball officials say that this is about 70% of the buyers) but not for the other 30% who pick their own numbers.

There are a number of approximate methods for answering birthday problems when the number of possible birthdays m is large. In the appendix, we illustrate a Poisson approximation showing that the number of people necessary to make it a fair bet that at least two people have the same birthday is approximately $1.2 \sqrt{m}$. Choosing $m = 80,089,128$ we find that only 10,740 Easy Pick tickets have to be sold before there is a better than even chance that at least two of these picks are the same, even though there are more than 80 million possible picks!

Consider the birthday problem with m equally likely dates and k people. Here is a way to approximate the probability that there is no birthday match among the k people. There are $C(k,2)$ pairs that might match, and any such pair has probability $1/m$ of actually matching. If k is a small fraction of m, we can approximate the distribution for the number of matches by a Poisson distribution with parameter $= C(k,2)/m$. Then the probability p that there are no matches is approximated by e^{-}. Thus we have the following approximate relationship among m, k and p:

$$\exp\{-C(k,2)/m\} \quad p.$$

For example, we estimated that $p < 1/2$ for $C(k,2) > m \log(2)$. Approximating $C(k,2)$ with $k^2/2$, the last inequality gives $k > \sqrt{2m\log(2)}$ $1.2\sqrt{m}$. Choosing $m = 80,089,128$ we found in the text that only 10,740 Easy Pick tickets have to be sold to give a better than even chance that at least one pair would match.

Now consider birthday triples. There are $C(k,3)$ triples that might match, and any such triple has probability $1/m^2$ of actually matching. Thus, letting q denote the probability that there is no birthday triple among the m people, we have the Poisson approximation

$$\exp\{-C(k,3)/m^2\} \quad q.$$

Similarly, if r is the probability that there are no birthday quadruples, we obtain

$$\exp\{-C(k,4)/m^3\} \quad r.$$

In the text we considered the case of $m = C(45,5) = 1,221,759$ sets of 5 white numbers chosen from 45 (the old Powerball format) and $k = 17,001$ tickets. The approximations gives q .58 and r .998. The complementary probabilities were of interest in our earlier discussion.