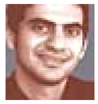
THE WALL STREET JOURNAL. ONLINE

March 23, 2006

THE NUMBERS GUY
By CARL BIALIK



Picking the Perfect NCAA Bracket

March 23, 2006

Some big prizes are being offered for anyone who fills out a perfect NCAA men's college basketball tournament bracket. Papa John's is dangling one million pizzas. America Online is offering $1 million. Gambling site Sportsbook.com has put up $10 million, and says it may offer even more money next year.

But it's unlikely anyone will pay out a single penny -- or slice of pepperoni. "You're much more likely to get hit by lightning attending the game than to win the pizza," says Brad Carlin, a professor of biostatistics at the University of Minnesota who has studied the NCAA tournament.

A look at the odds of winning shows why companies are willing to risk such valuable loot. Filling out a perfect bracket means predicting the outcome of 63 games. If each game were a true toss-up, that would mean your chance of perfection is a mere one in two to the 63rd power, or one in nine million trillion (yes, million trillion -- there are no tidy terms for numbers this large). Put another way, you are about 60 billion times more likely to win the multistate Powerball lottery[1].

Of the half-dozen Web sites I contacted, representing millions of entries over the past decade, not one spokesman remembered an entrant coming close to perfection. Through this year's upset-rich first two rounds -- the tournament resumes Thursday -- every entrant to those sites has already been eliminated from the perfection contests. Most sites' current leaders missed at least three games in the first round. Making that free pizza even more elusive: More college players are leaving top hoops programs before graduating, making upperclassmen-heavy second-tier schools more competitive and leveling the playing field.

Papa John's promised to split the one million pepperoni pizzas, which it valued at $15 million, among all entrants at Sporting News's Web site, if just one achieved perfection. "Even though the odds are long, it would be a pretty cool giveaway for everyone to win a share of the million pizzas," Jason Kint, vice president and general manager of sportingnews.com, told me. But none of the roughly 90,000 entrants lasted past the first round. Northwestern State's upset of the University of Iowa tripped up almost everyone. Only three entrants were still in contention after that game, and all three predicted, incorrectly, that Northern Iowa would upset Georgetown. (All is not lost: The best forecaster still gets free pizza for a year.)

Until last year, Sportsbook.com, owned by the U.K.'s Sportingbet PLC, offered $1 million for a perfect bracket. But then marketing director Alex Cjakowski decided to raise the stakes to $10 million. "I looked at it and said, this is never going to happen," he told me. "So let's have a big headline number." He added, "Next year, maybe I'll offer $100 million."

Improving the Odds

Of course, you can do better than just flipping a coin for each game. Some teams are better than others. I spoke with a half-dozen statisticians and mathematicians to get their best guesses about how well an informed picker could theoretically do. The most generous estimate for the chance of a perfect bracket: about one in 150 million.

If millions of people enter a particular contest, it might seem like the chance of someone winning is suddenly in the realm of possibility. But there's a catch: This scenario assumes everyone maximized their chances by picking mostly favorites, so those with the best shot at winning are likely to have identical entries. These contests generally protect themselves from big losses by stating they'll divvy up the loot if there are multiple perfect brackets.

These favorable conditions make insuring these prize offers a good business, as the Dallas company SCA Promotions has discovered. SCA, founded by 11-time world bridge champion Robert D. Hamman, has taken on the insurance risk for roughly 50 perfect-bracket prizes -- including a Sporting News offer of $1 million in 2001, according to vice president Chris Hamman, the founder's son. In the 12 years it has been doing so, SCA has never had to pay out a claim. Mr. Hamman declines to disclose the fee structure for these risk-transfer contracts, but says, "It's pretty normal for sponsors to spend about 2.5% of the prize value. If it's a large enough prize, [the percentage] could easily go up." (Sportsbook, for its part, says it hasn't bought any insurance to hedge its $10 million offer.)

Mr. Hamman says he has had inquiries from companies eager to launch similar contests for this summer's World Cup. The soccer tournament features 48 games and three possible outcomes -- win, lose or tie -- for the first 32 games, making it even harder to perfectly predict than the

PROBABILITY OF PERFECTION

Here are a few other methods for getting an estimate of the probability of picking a perfect bracket:

A No. 1 seed has never lost a first-round game to a No. 16 seed, though several came close this year; and No. 2 seeds rarely lose to No. 15 seeds -- just four times in 88 games. So the University of Minnesota's Brad Carlin suggests, as a simple model, assuming those games are locks, and counting the other 55 games as toss-ups. That yields a probability of one in two to the 55th power, or one in 36 million billion.

Chuck Newman, director of New York University's Courant Institute of Mathematical Sciences, suggests making the more generous assumption that the top two seeds in each region are better than the rest of the field, so any games they play against lower seeds are locks. Only games between No. 1 and 2 seeds, or amongst the other teams in the rest of the field, are toss-ups in this model. That makes 24 games locks, yielding a probability of one in 550 billion. Prof. Newman also suggests dividing up the bracket between seeds one through four, and then the rest, which makes 32 games locks and yields a probability of one in 2.1 billion. Tournaments never quite play out like this -- at least one lower seed always has scored an upset -- but again, this is a model.

I also asked Gary Lorden, head of the math department at Cal Tech and an advisor to the CBS show "Numbers," to crunch some numbers for me. He created a simulation that assumed the accuracy rate of the person filling out a bracket would vary from game to game, but would generally fall in a narrow range. One simulation, that gave pickers an overall accuracy rate of about 67%, suggested a probability of perfect-picking of one in 240 billion.

Running similar simulations with forecaster accuracy of 70% gives a probability of one in 13 billion, while getting a whopping 75% of games right, on average, yields a probability of one in 150 million -- impressive, but not amounting to much for a $1 million prize. "Even if you think you're hot stuff, the expected value of this is almost a penny," Prof. Lorden wryly notes.

Dennis DeTurck, professor of mathematics at the University of Pennsylvania, suggests using the seed numbers to estimate the probability of the favorite winning each game. In his scheme, the odds of the favorite are arbitrarily set at (higher seed number-lower seed number)/30 + 0.5. So a No. 1 seed playing a No. 16 seed has a probability of (16-1)/30+0.5=1 to win. Extending these odds across the whole tournament yields a probability of one in 150 billion.

(Besides those quoted, thanks to Stan Garstka and Ed Kaplan of Yale, Jerry Reiter of Duke University and Robert L. Wardrop of the University of Wisconsin for their help.)

college-basketball tourney. He compares the odds of having to pay out for the basketball or soccer tournaments to a local bar's promotion that SCA once insured, offering cash prizes if, on a given day, aliens landed in the bar's parking lot or Elvis showed up.

Crunching the Numbers

My ad hoc advisory committee of mathematicians and statisticians proposed several ways to estimate the odds of picking a perfect bracket, which provide a window into the thinking of probabilists[2]. First, a word on the structure of the single-elimination tournament: It's divided into four regions. In each region, the 16 teams are assigned ranking numbers, or seeds, from one to 16, better to worse, and paired off in first-round games such that the seed numbers add up to 17. Assuming favorites advance, second-round games pit teams whose seeds add to nine; regional semifinals, five; and regional finals, three. The four regional champions play a pair of semifinals, with the winners playing for the championship. Teams with the same seed number from different regions are considered to be roughly equal in strength. (A picture of the bracket -- or a PDF[3] -- is worth a thousand of my words describing it.)

The simplest calculation, mentioned earlier, assumes that each team has an equal chance at winning each game. If you assume instead that the favorite -- either by seed number, sports betting lines or various power ratings -- has a two-thirds chance of winning each game, the probability of a perfect bracket rises to two in three to the 63rd power, or one in 124 billion. Mr. Hamman of SCA Promotions, the insurer, says that favorites historically win about 72% of the time, which would yield a probability of perfection of one in 970 million.

I run through some other schemes that can be applied to any year's tournament in the box at left. To measure the probability for this year's tournament, Jay Emerson, assistant professor of statistics at Yale, suggests using power ratings developed by Ken Pomeroy, a 32-year-old meteorologist from Cheyenne, Wyo. These ratings are based on team's records, margin of victory, strength of schedule and other factors, and are expressed in units of points. For example, through last weekend's games Villanova has a rating[4] of 65.64 and Boston College has a rating of 61.99, so Villanova is expected to beat Boston by about four points -- the difference in their ratings -- when they play Friday.

A forecaster could use the ratings from before the tournament (which Mr. Pomeroy sent to me) to predict who would win each matchup. Mr. Pomeroy says the ratings chose a winner in about 71.3% of games this year before the tournament. "There's so much variation in performance from game to game, that even if you had a perfect system of ranking teams by how good they are, you'd still have significant errors," he told me.

Based on Mr. Pomeroy's stats, I computed the probability that teams would win in all 63 matchups -- I don't recommend you try this at home -- and found that if I had relied on power ratings, I would have had a one in 722 billion chance of a perfect bracket. (I'd also have chosen Kansas, a first-round loser, to make the Final Four.)

Of course, none of these models account for forecaster psychology. The great satisfaction of picking an upset, and the lure of picking one's own favorite team to win, combine to make picking all favorites more unpalatable than pizza is palatable.

These forces conspired to make me, a writer of both a sports column and numbers column, pick first-round loser Syracuse to win the championship in our office pool. I'm tied for last place.

* * *

Several readers responded to my last column[5], which covered several topics. First up are excerpts from letters about methamphetamine. I wrote that a stat from the state of Tennessee overstated the problems posed by the illegal drug:

There may have been some exaggeration from that Tennessee group, but I live in the Ozarks, an area long identified with a meth problem. Perhaps meth users aren't literally dead in five years, but from what I have been informed, they might as well be. Their brains are irreparably damaged and they are not going to come back to who they were -- they're gone. And in their condition, they cause others to die either from their irrational behavior or criminal behavior. These are numbers on which that I would like to see you do a follow-up piece, because I am sure I don't have a complete picture. And the complete picture, I suspect, is scarier than the exaggerated one.

--Sidney Ewer

While you may be correct in your numbers of addicts, deaths, etc., I feel that your article minimizes the meth problem. I guarantee you that if you were to have a family member or close friend addicted to this evil you would not be so dismissive of the statistics or the problems associated with addiction.

--Kevin Maloney

Incidentally, Willamette Week, an alternative weekly newspaper in Oregon, on Wednesday published an article[6] asserting that the Oregonian, which has covered[7] meth addiction extensively, "relied on bad statistics and a rhetoric of crisis, ultimately misleading its readers into believing they face a far greater scourge than the facts support." The Oregonian defended its reporting.

Many readers also wrote in about an item on a poll comparing Americans' knowledge about Simpsons characters with their ability to identify freedoms from the First Amendment:

I glanced through the McCormick Tribune survey results and the thing that struck me as odd was the need for survey respondents to specifically know that the right/freedom granted was in the First Amendment. It seemed to me that a very large percentage of people knew what their rights/freedoms were, though not specifically where the right/freedom is granted. It is a little like knowing which section of your state's penal code prohibits speeding or knowing the Social Security numbers of the

Simpsons family.

--Shawn Nelson

Why do we care so much about ignorance of the contents of the First Amendment? The beauty of our country is that it is so free that we take it for granted and can take it for granted. I don't notice the air I breathe. It's just there.

--Bob Barnes

Why is it "troubling" to learn that 22% of Americans can name all five members of the Simpson family? This series has aired for, what, 12? 13? 14? years [actually, it's in its 17th season --Carl] and some networks broadcast episodes every weekday. American history has nothing on "The Simpsons" for media domination.

--Adam

Other readers responded to my criticism of companies' misuse of the statistical concept "law of large numbers" in explaining slowing earnings growth:

Thank you for calling Google and eBay on their misuse of the "law of large numbers." If it is indeed corporate vernacular, it must be stopped immediately. I'm trying to imagine how scientists could misuse business terms as part of their vernacular. Perhaps I should start using phrases in the lab like, "the electron is doing due diligence on the proton." Or, "glutamate really adds value to the synapse."

--Casimir Wierzynski

Several readers also responded to a New York Times article[8] I linked to, about the proliferation of arbitrary and ever-growing numbers on magazine covers:

Who wouldn't want to read about 25 sexy little secrets? Especially with a picture of Jennifer Alba on the cover! Thank God for the seventh freedom of the First Amendment: freedom to publish "sexy little secrets!"

--Shad Sletto

Write to Carl Bialik at numbersguy@wsj.com[9]

URL for this article:
http://online.wsj.com/article/SB114304314046605235.html

Hyperlinks in this Article:
(1) http://www.powerball.com/powerball/pb_prizes.asp
(2) http://en.wikipedia.org/wiki/Probabilist
(3) http://online.wsj.com/documents/ncaamen06.pdf
(4) http://kenpom.com/rate.php
(5) http://online.wsj.com/public/article/SB114183387483792619.html
(6) http://www.wweek.com/editorial/3220/7368/

(7) http://www.oregonlive.com/special/oregonian/meth/
(8) http://www.nytimes.com/2006/02/10/business/media/10numbers.html?ex=1297227600&en=2a6303527fa61666&ei=5090&partner=rssuserland&emc=rss&pagewanted=all

(9) mailto:numbersguy@wsj.com
(10) http://online.wsj.com/public/page/0,,2_1125,00.html
(11) http://online.wsj.com/public/page/0,,2_1125,00.html
(12) mailto:numbersguy@wsj.com
(13) http://online.wsj.com/email
(14) http://WSJ.com/NumbersGuy
(15) http://online.wsj.com/xml/rss/0,,3_7028,00.xml
(16) https://users2.wsj.com/WebIntegration/WebIntegrationServlet?call=A_EC
(17) https://users2.wsj.com/WebIntegration/WebIntegrationServlet?call=A_EC
(18) http://online.wsj.com/reg/page/0,,5_3017,00.html
(19) http://online.wsj.com/article/0,,the_daily_fix,00.html
(20) mailto:numbersguy@wsj.com