Express Riddler

17 July 2020

Riddle:

This year, Major League Baseball announced it will play a shortened 60-game season, as opposed to the typical 162-game season. Baseball is a sport of numbers and statistics, and so Taylor wondered about the impact of the season's length on some famous baseball records.

Some statistics are more achievable than others in a shortened season. Suppose your true batting average is .350, meaning you have a 35 percent chance of getting a hit with every at-bat. If you have four at-bats per game, what are your chances of batting at least .400 over the course of the 60-game season? And how does this compare to your chances of batting at least .400 over the course of a 162-game season?

Extra credit: Some statistics are *less* achievable in a shortened season. What are your chances of getting a hit in at least 56 consecutive games, tying or breaking Joe DiMaggio's record, in a 60-game season? And how does this compare to your chances in a 162-game season? (Again, suppose your true batting average is .350 and you have four at-bats per game.)

Extra extra credit: In a 60-game season, what are your chances of *both* batting at least .400 and getting a hit in at least 56 consecutive games?

Solution:

Given a probability p of getting a hit (or more generally, a positive outcome) on a single at-bat, the probability P of getting exactly n hits out of N at-bats is given by the binomial distribution:

$$P = \binom{N}{n} p^n (1-p)^{N-n}.$$

For the current problem, $p = 0.35$, $N = 4 \times 60 = 240$, and the probability must be summed for all $n \geq 0.4 \times 240 = 96$. Luckily, all of this can be done in Excel. I set up the calculation in the file `Batting_average.xlsx`. First I calculated the individual probabilities for each n, and summed the answers. This I called the brute-force calculation. The result is that for 60 games, the probability is about $\boxed{\textbf{6.1\%}}$. For 162 games, the probability is $\boxed{\textbf{0.38\%}}$.

For fun, I also set up a calculator that can take in as input any number of games, number of at-bats per game, true batting average, and desired batting average. Next I decided to plot how the probability changes as a function of the number of games as well as the true batting averge (p). The plot uses 4 at-bats/game and a desired batting average of 0.4 Here is the result: