2022 Philadelphia Phillies Team Statistics

Applying probability and statistics on an MLB team

Cody Kimmins

School of Business, Stockton University

Probability and Applied Statistics

December 9, 2022

**Philadelphia Phillies Statistics in 2022**

Cody Kimmins

Baseball is a sport that heavily relies on statistics to give the team an edge in winning games. I have always liked watching baseball because partly because of this element. I have decided to choose the Phillies team since I have always been a fan and their World Series run this year. The odds of the team winning the National League (series right before the World Series) was at 16 to 1 or 5.9%. Even lower was the chance of winning the World Series at 4.2%. It is interesting to look at those probabilities now given that the Phillies won the National League and was 2 games away from winning it all. In baseball the managers, coaches, hitters, and pitchers all review data taken on their opponents to come up with a strategy to beat them. For example, if a hitter has a high strikeout rate when the ball is in the top of the zone then they will use this to their advantage. Another example could be observing the statistics on a pitcher like the percentage of throwing a fastball. The most common statistic for a hitter is their batting average, runs batted in, and homeruns. For a pitcher it is earned run average, number of games started in, and their win/loss percentage. The problems will be created by utilizing the statistics on the Philadelphia Phillies 2022 regular and postseason.

Creating a relative frequency histogram, we can find out the total percentage of the team’s batting average during the regular season. Based on the graph the highest frequency was between 0.230 and 0.255. A batting average of 0.300 marks a great batter while anything below 0.200 is poor. The 2022 Phillies had around 50% of good batters with their averages ranging from 0.230 to 0.280. The batters rarely had great averages with only 8% of the total ranging from 0.305 to 0.355. There were some batters that were not good enough to play regular games so this would be the bench unit. Their numbers are between the range of 0.130 and 0.180 since they either might not be MLB level ready or plainly not good.

The next section discusses a review of set notation. Let set A consist of infield position players (First base, second base, third base, shortstop, catcher, and pitcher), A = {FB, SB, TB, SS, C, SP}. Let set B consist of different types of pitchers (Starting, relief, and closing pitcher), B = {SP, RP, CP}. The intersection of A and B would be . Since a starting pitcher is not considered a position player it would be an empty set. Set A and B do not have anything in common since starting pitcher is not in both sets. Next problem will introduce one of the starting pitchers named Aaron Nola who had a good year in 2022. The proportions of pitches for Nola on the Phillies include fastball, curveball, sinker, changeup, and cutter. The percentages he uses them are approximately .33, .27, .19, .15, and .07 respectively. The sample space will be S = {FB, CB, S, CH, CR}. What is the probability of the first pitch being either a fastball or a sinker? or 52%. This result means that the chance of Aaron Nola pitching to the batter a fastball or sinker is a little over 50% which keeps them guessing. Trying to guess which pitch is thrown next is the batter’s job and having more pitches to throw will keep them off balance.

Combinations and permutations are the next important topics that were covered in the course. When the manager of the Phillies arranges the batting order it is a type of permutation problem. There are 9 batters in the order so to find how many ways they can be arranged it would be setup as ways. If we want to find the way batters can be in the lineup without any order, we will use combination. The formula will be . This means that there is only 1 combination possible for the batting lineup. In baseball the order does matter since you want someone with a good on base percentage to be hitting first second or third. Those who are not so great a batting will be sent to the bottom of the order seven, eight, or nine. Consider the following events A and B such that P(A) = .04, P(B) = .60, and P () = .02. A being the probability the team wins the world series and B is they make the playoffs. The probability that the team would win the world series given that they made the playoffs is = .033 or 3.3%. Although the team had a solid chance of making the playoffs, there was still a great chance they would not even win the World Series. In the end they did make the playoffs and the World Series and almost winning. Kyle Schwarber on the Phillies has a 7% chance of hitting a homerun at each at bat but also has the chance of striking out at 29%. What will be the probability that he either hits a homerun or strikes out? This is an example of the Additive Law of Probability and can be solved by the following formula: . P(A) = .07, P(B) = .29, = 0. Since this is a mutually exclusive event (cannot hit a homerun and strikeout), or 36%. This is a surprising statistic since Schwarber was known for swinging for the fences. When you try and swing for a homerun there is also a correlation to striking out more often.

Of the 2022 starting pitchers on the Phillies, 67% were right-handed and 33% were left-handed. Among the right-handed pitchers 75% had an ERA under 4 and 50% for the left-handed. ERA or earned run average represents the number of earned runs a pitcher allows per nine innings. The lower this number is the better. If a pitcher is selected at random, what is the probability that he will have an ERA under 4.00? Let E denote the event “under 4.00 era”, R the event “a right hander is selected”, and L the event "a left hander is selected”. Then . , so and = .19. Taking these values and adding it will be so the probability of randomly picking a pitcher with an ERA under 4.00 is 69%. This makes sense because the team consists mostly of great pitchers with only a few who were not good during the 2022 season. The next topic is Bayes Rule which is the probability of event B given that event A occurred. Using the previous question but this time we want to find the conditional probability that it is a left-handed pitcher. The events will be labelled as the same being R for right-handed pitchers, L for left-handed pitchers, and E for under 4 ERA. We can solve this by using the formula . This means that there is a 25% chance of randomly picking a starting pitcher that has an ERA under 4.00 that is also left-handed.

In the next section the different distribution probabilities are discussed starting with Binomial. This distribution can calculate the x amount of success on n trials being that the trials are independent of each other which also have two outcomes: success and failure. The Phillies’ fielding statistics were not great for 2022 and especially for the first baseman Rhys Hoskins. The probability of Hoskins making an error was 25% every game. What is the probability that Hoskins made exactly 2 errors in one game? Let Y = errors, n = 2, and p = .25. . I am surprised by this statistic since Hoskins is known for a lot of errors during the season. Thinking about more makes sense because most players do not even commit a single error during most games. The next section is the topic of the Geometric distribution which is like binomial. The probability of success is equal to p and is constant from trial to trial. However, instead of the number of successes that occur in n trials, the geometric random variable Y is the number of the trail on which the first success occurs. In 2022 Zack Wheeler, who was one of the Phillies’ best pitchers, had a strikeout rate of 27%. We want to find the probability of Wheeler striking out the fifth batter he faces in a game. The Geometric distribution formula is as follows = p(5) . We can also calculate the average or the expected value which is = . = 3.7%. Now for the variance now if we plug in which this measures the spread between values in our data set. Now for the standard deviation .

The next distribution is the Hypergeometric distribution which describes the number of successes in a sequence of n trials from a finite population without replacement. The 2022 Phillies team consisted of 27 relief pitchers in the bullpen but 13 of them are not used. The coach selects 4 of the relief pitchers at random. What is the probability that all 4 are used? Using the formula = = , 57%. The next section is the Poisson probability distribution gives the probability of a number of independent events occurring in a fixed time. As an example, a typical baseball game lasts 3 hours and the Phillies pitcher Ranger Suarez averages 1.33 runs given up to the opposite team per hour. During any given hour, what is the probability that he gives up exactly 1 run? . The next section is about Tchebysheff’s Theorem which helps to estimate least proportion of operations that are within a specific number of standard deviations away from the mean. As an example, the Phillies pitcher Zack Wheeler has 5 pitches and with his average pitch speed being 90mph with a standard deviation of 10. What would be the minimum percentage of mph for Wheeler to throw between 75 and 105mph? To solve this, we can use the formula = = 1.5 and = -1.5. For the percentage, = .55, it would be 55%. The uniform distribution is another distribution we learned the formula is as shown here . We can look at the batting averages of the team specifically within the interval 130 and 255 and find the probability that the next batter will have a batting average more than 200.. This means that there is a 44% the next batter will have a batting average more than 200.

The next section is about bivariate and multivariate probability distributions where random variables can be defined over the same sample space. We can look at the number of hits and runs by the Phillies’ batters. For the 9 players, 7 of them had 100 or more hits and 50 or more runs. Let Y1 denote the players that had 100 or more hits and Y2 denote the players whose hits equated to 50 or more runs for the team.

Y1 = and Y2 =

We can put this data into the following table format.

Y1

|  |  |  |  |
| --- | --- | --- | --- |
|  | 0 | 1 | Total |
| 0 | 2/9 | 0 | 2/9 |
| 1 | 0 | 6/9 | 6/9 |
| 2 | 0 | 1/9 | 1/9 |
| Total | 2/9 | 7/9 | 1 |

Y2

With this table we can see that most players had 100 hits or more that equated to at least 99 runs for the team. This makes sense because the more hits a player gets the better the chance, they will have more runs. There are no players that have less than 100 hits with 50 or more runs which also makes sense.

References

<https://www.pitcherlist.com/player/aaron-nola/>

<https://baseballsavant.mlb.com/savant-player/zack-wheeler-554430?stats=statcast-r-pitching-mlb>

<https://www.baseball-reference.com/teams/PHI/2022.shtml>

<https://www.mlb.com/glossary>

<https://www.fangraphs.com/standings/playoff-odds-graphs?lg=A&div=A&stat=ws&year=2022>

<https://www.baseball-almanac.com/teamstats/fielding.php?y=2022&t=PHI>