1)

*Therefore, the probability of a pitcher being a Category A pitcher given they struck out three batters in a row is* **27%.**

2)

*Team X is expected to win* **8** *games versus team y this year.*

3)

*The ranking of most likely to least likely, respectively, is* **d, b, c, a.**

4) Even though spicy food tolerance and middle finger length are uncorrelated in the general population, the selective bias of baseball teams preferring players with both higher spicy food tolerance and longer middle fingers would result in a positive correlation in the MLB population. Due to this selection process, the MLB population becomes a non-random sample of the general population, and players who make it into the MLB tend to have both traits more often than would be expected by chance alone.

5) With 100 outs per inning and the ability to skip batters still on base, the marginal increase in run expectancy from taking an extra base is less significant compared to the increased opportunities to generate runs through hitting, making aggressive baserunning less valuable.

6) The price of removing the clause from the contract would be the expected value of the player’s bonus which would be the probability “p” that the player will win the MVP, derived possibly from the prior probability of how often the top player in the projection system has won the MVP, multiplied by the $1M amount i.e.

7)

phi\_value = norm.cdf(-0.98)

print(phi\_value)  # Output is approximately 0.1635



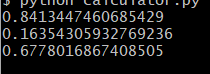
phi\_value\_upper = norm.cdf(1)

phi\_value\_lower = norm.cdf(-.98)

print(phi\_value\_upper)

print(phi\_value\_lower)

print(phi\_value\_upper-phi\_value\_lower)



from scipy.integrate import quad

# Parameters

mu = 50  # Mean of the normal distribution

sigma = 50  # Standard deviation of the normal distribution

lower\_bound = 1

upper\_bound = 100

# Define the PDF function

def pdf(x, mu, sigma):

    return norm.pdf(x, mu, sigma)

# Define the integrand function

def integrand(x, mu, sigma):

    return x \* pdf(x, mu, sigma)

# Perform numerical integration

integral, \_ = quad(integrand, lower\_bound, upper\_bound, args=(mu, sigma))

# Calculate the probability of the range

probability\_range = norm.cdf(upper\_bound, mu, sigma) - norm.cdf(lower\_bound, mu, sigma)

# Calculate the conditional expectation

expected\_value = integral / probability\_range

print("Expected value E[X | 1  X  100]:", expected\_value)



lc\_phi\_val = norm.pdf(1)

prob\_above\_100 = 1- norm.cdf(1)

print(lc\_phi\_val)

print(prob\_above\_100)

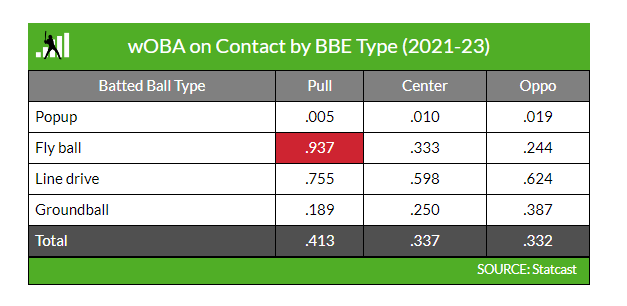
print("Expected Value E[X+20|X>100]: ",50\*lc\_phi\_val/prob\_above\_100+70)



8) If the Phillies have a 3-0 lead in the NLCS, and we assume the probability of winning each game is 50% (as if the opponent is evenly matched), the probability of sweeping the series is therefore, just **50%** ignoring the prior probability which is irrelevant since all the NLCS are independent events.

9) Research has shown that different ballparks, and hence different environments, change movement profiles on pitches. For example, pitcher’s stuff+ has been calculated as worse at Coors Field due to effects such as seam-shifted wake not having the same effect, and conversely, stuff+ sometimes improves in a dome setting such as at Tropicana. Therefore, it would be interesting to test out a park-adjusted Stuff+. This would theoretically reward pitches thrown at Coors and penalize (even if ever so slightly) pitches in Tampa to try and achieve a true talent/pitch profile stripped of environmental effects i.e. If Garrett Crochet primarily plays in a park where he doesn’t get as much ride as he normally would on his four-seamer, it would behoove an analytics department to examine this, particularly if they work for an organization in a “ride-friendlier” park. Additionally, pitchers have well defined and predictive handedness splits in much smaller sample sizes than hitters do and that’s directly a result of splits occurring at the individual pitch classification level. Therefore, if Stuff+ was also trained on the batter’s handedness when assessing run values of pitches, and a handedness metric could be observed for splits to encapsulate/ consider how effective certain pitches are against each handedness, it could be more predictive for individual matchups. Lastly, secondary pitches’ features are defined off of the fastballs, but I believe there could be more research done into how the fastballs themselves play off of each other. For example, as stated in the primer, the mean for sinkers is lower than four seamers, and there has been a massive surge in the number of pitchers with three fastballs now, likely to keep the hitters off balance with different looks. This strategy has appeared to be effective for many starters (Hunter Brown added a sinker this season and went from a possible demotion if the rotation was healthier to being one of the best starters over the last couple of months), yet the models seem to underrate the array of fastballs these pitchers possess, despite the apparent effectiveness of the approach**.** This leads me to wonder if the model should be adjusted or extended for multiple frequent use of different fastballs. Currently, all fastballs (four-seamers, sinkers, and cutters) are lumped in together according to Eno Sarris, as even though they are classified as different pitches according to Statcast and have their own model outputs, they all are under the “fastball” model.

P.S. Sorry I obviously exceeded the three sentence recommendation, but I wanted to mention each of these topics as I have considered each of them since researching Stuff+. I’m also incredibly interested in Stuff+ and the predictive power it can have of a pitcher’s true talent level, and I could have an indefinitely long conversation on Stuff+, so at the very least I had to throw out a few ideas I had already been considering.

10) There are a multitude of reasons why hitters that have identical batted ball profiles in terms of EV and Launch Angle could have different results. Arguably, the most notable being due to park factors. In this hypothetical scenario, Batter A most likely played in a hitter friendlier environment for his handedness than his counterpart. Additionally, the spray angles weren’t specified and almost certainly differed among the two hitters’ batted balls. Specifically, pulled fly balls and pulled line drives outperform batted balls hit either to center or to the opposite field, and its by a rather large margin:  


Intuitively, this makes sense as the shortest distance is to the foul poles, but research has also confirmed that hitters have higher bat speeds during pull-side contact and the 75th percentile exit velocities are much higher for negative spray angles (pulled batted balls) as opposed to balls hit to the opposite field. However, since in this scenario all the batted balls have the same exit velocities and should theoretically have the same distance due to the combination of launch angle and exit velocity, the results would differ due to the distance being shorter to pull side than it is to center, meaning hitter B probably pulled their fly balls much more infrequently. Lastly, it could be due to simply just luck and random variation. Hitter A already probably hit in a friendlier park, as calculated by the 3 year rolling park factor, but there also could have been ideal conditions on several days conducive to home runs (warmer weather, higher humidity, random favorable wind days, etc) that might not be fully encapsulated by a long running average factor. Furthermore, oddities such as the way balls are handled pre-game or if there are slight characteristic differences on the baseballs themselves, it could create varying drag which would affect the distance traveled on balls hit with the same velocity off the bat and at the same launch angle.