PHILLIES RESEARCH AND INFORMATION QUESTIONNAIRE

Question 1: Suppose you live in a universe where there are only two types of pitchers: Category A pitchers who strike out 30% of batters faced at the major league level, and Category B pitchers who strike out 20% of batters faced at the major league level. 90% of pitchers fall into Category B, but we don't know who they are before they play in the majors. You observe Pitcher X facing three batters in the majors, and he strikes out all three. Rounded to the nearest percentage point, what is the probability that Pitcher X's true talent is a 30% strikeout rate?

Probability of a pitcher being in category A, P(A) = 0.1
Probability of a pitcher being in category B, P(B) = 0.9
Probability of a strikeout from category A pitcher, P(K|A) = 0.3
Probability of a strikeout from category B pitcher, P(K|B) = 0.2
Probability of 3 K's from category A pitcher, P(3 K's|A) = 0.3^3 = 0.027
Probability of 3 K's from category B pitcher, P(3 K's|B) = 0.2^3 = 0.008
Plug into Baye's Theorem ->
P(A|3 K's) = (P(3K's|A) * P(A)) / (P(3K's|A) * P(A) + P(3K's|B) * P(B))
P(A|3 K's) = (0.027 * 0.1) / (0.027 * 0.1 + 0.008 * 0.9) = **27%**

Question 2: Imagine a baseball league where all players in baseball choose their teams randomly every offseason for the coming year. In this league, Team X and Team Y face each other 20 times each year. In 2022, Team X beat Team Y all 20 times. In 2023, a well-calibrated set of player projections suggests that Team Y will beat Team X 60% of the time, based on the players on each team.
How many games do you expect Team X to win vs. Team Y in 2023?

Probability of team X winning, P(X win) = 1 - 0.6 = 0.4
20 games -> 20 * P(X win) = 20 * 0.4 = **8 games**

Question 3: Rank the following events from most to least likely.
a. Flipping a fair coin four times and getting all heads
b. A team with a true talent win pct of .400 beating a team with a true talent win pct of .500
c. Blowing a 3-1 lead in a best-of-seven series to an evenly matched opponent
d. A player with a projected .400 On Base Percentage making an out in an important situation

a. 0.5^4 = 6.25%
b. 0.4 / (0.4 + 0.5) = 44%
c. Evenly matched so P(loss) = 0.5 -> 0.5^3 = 12.5%
d. 1-0.4 = 0.6

From most to least likely: d, b, c, a

Question 4: Assume that spicy food tolerance and middle finger length are uncorrelated and normally distributed in the human population. Additionally, assume that baseball teams strongly prefer players with more spicy food tolerance and longer middle fingers. What would you expect

the relationship to be between spicy food tolerance and middle finger length in the major league population?
a. Please choose from the following options: Uncorrelated, Positively correlated, Negatively correlated, Not enough information
Positively correlated
b. Explain your thinking in no more than two sentences:
Even if the traits are uncorrelated in the human population, when the MLB selects for high values in these two traits you're introducing selection bias, which induces a positive correlation.

Question 5: Imagine that baseball altered its rules such that there were 100 outs per inning. In this form of baseball, the lineup is still 9 batters long, but if a batter's spot in the lineup comes up while he is still on base, the offensive team can just skip to the next available batter without penalty.
a. How would you expect the value of baserunning would change relative to the value of hitting?
Decrease
b. Explain your thinking in one sentence.
With 100 outs per inning, advancing runners (via stealing or going 1st to 3rd, etc.) becomes less critical while maximizing total hits would become more important.

Question 6: Consider an internal projection system that rates Player M the best player in baseball. He has incentives in his contract to award him $1M if he wins the MVP.
a. How much would you be willing to pay him before the season to remove that clause from his contract?
Max pay = P(win MVP) * $1 million, so if 20% chance to win MVP then $200k
b. Explain your thinking in one sentence.
Above is a formula for calculating how much you should be willing to pay to remove the clause based on expected outcomes, but I'd argue the value of having your superstar having something to chase all season could be worth more to the team than the potential cost savings.

Question 7: Consider a game where you can pay to access an unusual lottery. Tickets are picked from a normal distribution of possible values, with mean $50 and standard deviation $50. If your ticket's value is below $1, you receive a pity prize of $1. If your ticket's value is above $100, you receive the dollar amount on your ticket, plus a bonus of $20. Otherwise, you receive the dollar amount on your ticket.
a. What is the expected value of a ticket to this lottery? Show your work by providing math, code, or a source for your thinking.

```
import numpy as np

# Parameters from problem statement
mean = 50
std_dev = 50
```

```
n_simulations = 100000000

# Simulate ticket values using a normal distribution
tickets = np.random.normal(loc=mean, scale=std_dev, size=n_simulations)

# Apply payout rules
def payout(x):
    if x < 1:
        return 1
    elif x > 100:
        return x + 20
    else:
        return x

# Vectorize the payout function
payout_vectorized = np.vectorize(payout)
single_payouts = payout_vectorized(tickets)

# Part A: Expected value of a single ticket
expected_value_single = np.mean(single_payouts)
print(f"Expected value of a single ticket: ${expected_value_single:.2f}")
```

Expected value of a single ticket: $57.51 (may fluctuate +/- $0.01 with each simulation)

b. You now have the opportunity to receive two tickets from that lottery (an "entry"), and can redeem only the higher value ticket as the prize for your entry. What is the new expected value of an entry to this lottery? Show your work by providing math, code, or a source for your thinking.

```
# Part B: Two tickets per entry, redeem higher payout
tickets_1 = np.random.normal(loc=mean, scale=std_dev, size=n_simulations)
tickets_2 = np.random.normal(loc=mean, scale=std_dev, size=n_simulations)

payouts_1 = payout_vectorized(tickets_1)
payouts_2 = payout_vectorized(tickets_2)

# Take the higher of the two payouts
entry_payouts = np.maximum(payouts_1, payouts_2)

# Expected value of an entry
```

```
expected_value_entry = np.mean(entry_payouts)
print(f"Expected value of an entry (two tickets, redeem higher):
${expected_value_entry:.2f}")
```

Expected value of an entry (two tickets, redeem higher): $84.44 (may fluctuate +/- $0.01 with each simulation)

Question 8: Historically, sweeping the NLCS has been quite difficult; since switching to a 7 game format, only 4 of 37 NLCS have ended in a sweep. Imagine that the Phillies take a 3-0 lead in the NLCS.
a. What is the approximate probability, rounded to the nearest 10%, that the Phillies sweep the series?
80%
b. Explain your thinking in one sentence.
Historically, across all best of 7 series, the team up 3 to 0 has swept 31/41 times (75.6%), in the NLCS they've swept 4/7 times, but that sample is too small to draw much of a conclusion from.
Source: https://champsorchumps.us/records/mlb-series-odds-up-3-0

Question 9: Read this primer on Stuff+. What are some other features you might want to add to a Stuff+ model in order to make it more predictive of future performance? Limit your answer to three sentences.

To make Stuff+ more predictive of future performance, you could incorporate pitcher fatigue metrics, such as pitch count trends or recovery time between outings, which influence pitch quality over time. Adding environmental factors like ballpark dimensions and weather conditions could help contextualize pitch effectiveness. Finally, integrating batter quality faced—such as average exit velocity or chase rate—would allow the model to account for the difficulty of opposition and better isolate pitcher skill.

Question 10: Imagine there are two batters (Batter A and Batter B) that both happen to hit the exact same number of fly balls at the exact same combinations of Exit Velocity (EV) and Launch Angle (LA) during the 2022 season. In other words, every hypothetical fly ball hit by both batters was exactly the same according to EV and LA in 2022. However, Batter A happened to finish the season with 30 HRs, while Batter B finished the season with only 20 HRs. What factors could potentially explain this discrepancy in overall player production?
Please list your top 3 reasons.

1. **Park factors (stadium dimensions, altitude)**
2. **Pull vs opposite field (at Yankee Stadium, 390 ft to right-center is a HR but 390 ft to left-center is in the field of play)**
3. **Luck/randomness (wind, temperature, defensive positioning)**

Question 11: Using the resources available in the BaseballCV repository, leverage transfer learning to improve upon the glove-tracking portion of this pretrained object detection model by using images in the
baseball_rubber_home_glove.zip file.

a. Produce a writeup or notebook detailing the entire predictive modeling pipeline, including the reasoning behind any assumptions, feature engineering, or modeling choices along with a section evaluating theperformance of your model compared to the pretrained model
b. Please return all files (data, code, documentation) necessary to run the predictive modeling pipeline in a private GitHub repository. When you're ready to submit your questionnaire, please share the privaterepository with pmcfarlane@phillies.com
c. Your submission will be judged primarily on the quality of your predictive modeling pipeline and justification for your decision-making process in constructing the model
d. Using online resources such as Stack Overflow is encouraged! Just please include a citation (by including a comment in your code with a link to the resource)

For this project, I started by gaining a baseline understanding of how the pretrained model was performing by training it as-is. On my first run, I noticed that some labels in the dataset included an unexpected fifth class. To address this, I added a section to my script to remove the extra label, ensuring the data aligned with the model's expected format.

Due to time and computational constraints, I limited training to 20 epochs. Ideally, I would have trained for somewhere between 50 and 100 epochs. Even with the shorter run, the model showed promising results. I noticed that a blur-based data augmentation was being applied, but only with a probability of 0.01. I attempted to increase this to somewhere between 0.2 and 0.5, but couldn't get the YOLO model to accept the new blur parameters. While I was thinking about augmentation, I also visually inspected the training set to confirm a good mix of lighting conditions—day games, night games, and afternoon games with varied sun/shade coverage across the mound and home plate.

After reviewing the pretrained model's performance (see figure above), I felt it was already training well, so I opted for minor modifications to start. My first change was reducing the learning rate from 0.01 to 0.005. The idea was that a smaller learning rate would allow the model to fine-tune more precisely, especially since it was starting from a pretrained checkpoint.

As expected, this model (see figure above) trained a bit slower than the original, but I was happy with the results. My next adjustment was increasing the number of warmup epochs from 2 to 4. I made this change to give the training process a smoother start and reduce the risk of unstable updates early on.

These results were encouraging as well (see figure above). Looking ahead, I have a few directions I'd like to explore. First, I'd obviously like to train for more epochs—the plots in this report clearly show that the metrics hadn't plateaued and would likely improve with continued training. I'd also like to investigate failure cases to identify any recurring patterns. Another

avenue is experimenting with different optimizers. AdamW has performed well, but I'm curious whether alternatives like SGD or RAdam could outperform it. I've also experimented with composite optimizers in the past, and I'd be interested to see whether combinations like AdamW + Lookahead or SAM + SGD could push performance even further.