Thank you for applying to join the Milwaukee Brewers Baseball Research & Development team. We'd like to get a feel for your problem-solving abilities through a series of small exercises. Please answer the following questions, showing all your work and code, and submit to r_and_d_hiring@brewers.com.

The primary languages we use are R and Python, but feel free to answer these using any programming language. Code is not required for any question, however attached code should include enough information to allow us to successfully run the program – please identify any needed libraries and include any support files you've used. Questions with numerical answers do not require exact answers, however, please be reasonable with any approximations or rounding.

If you have any questions, please feel free to send an email to r_and_d_hiring@brewers.com and we'll get back to you as soon as possible. If you are sending Python files, please compress all your files together into a single zip file. Please also watch your spam folder as replies may occasionally be filtered.

1. The probability that Hitter A draws a walk in a plate appearance is 7% and the probability that Hitter B draws a walk in a plate appearance is 10%. In a given month they each have 100 plate appearances. What is the probability that Hitter A draws more walks than Hitter B in that month?

2. A pitcher pitches the ball into the strike zone 48% of the time. When a pitch is thrown in the strike zone, the batter swings at the pitch 64% of the time, and when the pitch is thrown out of the strike zone the batter swings at the pitch 29% of the time. In a matchup of this batter and pitcher, a pitch is thrown and the batter swings. What is the probability that the pitch was in the strike zone?

3. A linear regression model is fit to determine which characteristics predict fastball success in the majors. The dependent variable is run expectancy change per 2000 pitches, a measure of pitch effectiveness where lower values indicate that the pitcher is more effective. The estimated coefficients on the linear model, as well as the estimated standard error on the estimates, are as follows:

   Intercept: 138
   Velocity (mph): -1.77
   Induced Vertical Break (in): -1.08
   Release Height (ft): 8.62
   Residual standard error/standard deviation of error on estimates: 9.0

   a) Given the table of pitch characteristics below, what is the probability that Pitcher A has a lower run expectancy change than Pitcher B under this model?

	Pitcher A	Pitcher B
Velocity (mph)	92	95
Induced Vertical Break (in)	18	12
Release Height (ft)	6	5.5

b) Running a linear regression on the same data set with Induced Vertical Break as the dependent variable gives the following coefficients:

Intercept: -0.48
Velocity (mph): 0.0126
Release Height (ft): 2.60

What would the coefficients (intercept and two slopes) be in a regression of run expectancy change on Velocity and Release Height only?

c) You hypothesize that taller pitchers are more effective than shorter pitchers with the same velocity and release point. If this were true, how would you expect the magnitude of the coefficient on Release Height in the final regression model from Part b (model predicting run expectancy change as a function of Velocity and Release Height) to change if pitchers' heights were added into the regression?

d) The original model assumes the variables were linearly related to each other. What is one way you would expect this assumption to be violated? How could you model fastball effectiveness using these features without that assumption?

4. At the beginning of the day on Sept 25, 2024, the National League Wild Card standings were the following:

Team	Wins	Losses
San Diego Padres	91	66
New York Mets	87	70
Arizona Diamondbacks	87	71
Atlanta Braves	86	71

The Mets and Braves have 2 remaining games against each other. The Padres and Diamondbacks have 3 remaining games against each other. Assume that all other games are against other teams. Assume that the probability of each team winning any of the remaining games is 50%.

The Top 3 of these teams, ordered by winning percentage, will make the playoffs. In the event of ties between teams, assume the tiebreaker order is determined by the table above, i.e., whichever team was highest in the standings on Sept 25, 2024 wins the tiebreaker.

What is the probability that the Mets finish the season as the third ranked of these four teams?

5. Assume that in extra innings the road team scores runs in each of their half innings according to the following distribution:

0 Runs	1 Run	2 Runs	3 Runs
35%	40%	15%	10%

When the home team bats in the bottom half of each extra inning, they will adjust their strategy depending on how many runs the road team scored in the top half of each inning. Assume that the home team scores according to the following distributions:

Road Team Half Inning Results	0 Runs	1 Run	2 Runs	3 Runs
2 or 3 Runs	35%	40%	15%	10%
1 Run	35%	20%	45%	
0 Runs	1 - *p*	*p*		

For what value of *p* will the home team be expected to win 60% of the extra inning games?

6. You are tasked with constructing a model to determine the probability that a pitcher will throw a pickoff throw over to first base instead of pitching the ball in situations where there is a runner on first base. How would you construct the model given that you had access to the following:

Identities of all players on the field (fielders, pitcher, batter, runners), game state (number of outs, runners on base, inning, score, count), ball and player time series tracking data (positions for given timestamps), crowd size, pickoff statistics from the previous season, fielding percentage for all fielders, temperature, day/night game indicator, average seasonal sprint speed for all players, and pitch usage statistics for the pitcher.

Note: There is no need to find data and generate an actual model. Please just describe how you would proceed assuming you had the data listed. Please limit your response to 300 words.

7. A baseball team is considering offering an extension to one of their young players. The following table has the player's average WAR (wins above replacement) projections for the next 8 seasons as of today.

The team has two possible options for contracts as of today:

a) The team can go year-to-year on the contract. At the end of each year the team can sign the player to a 1-year guaranteed deal for the next season. The expected salaries for each of those 6 years are shown in the table, however based on the player's performance in each year those salaries may go up or down. If the team chooses not to offer the next year's contract, the player is released and becomes a Free Agent. The player is not allowed to opt out after any of the first 5 years, but becomes a Free Agent after the 6^{th} season.

b) The team can offer an 8-year guaranteed contract at the terms in the table below. The salaries for all 8 seasons are guaranteed immediately and the player becomes a Free Agent after the 8th season.

What value of *X* would you guarantee for the final year of the 8-year guaranteed contract to have a break-even value of 0 between the two contracts (i.e. you would be indifferent to which of the two contracts the team offers the player)? Explain any simplifying assumptions you make and how you might improve your estimate with more time. Please limit your answer to 300 words.

Season	WAR	Year-to-Year ($M)	Guarantee ($M)
2025	1.4	0.76	2.00
2026	2.2	0.78	2.50
2027	2.9	0.80	4.50
2028	3.3	6.00	7.50
2029	3.4	10.00	10.50
2030	3.1	15.00	14.50
2031	2.8		19.00
2032	2.5		*X*

8. If MLB expanded the current playoff system to include 8 teams in each league, with a 3-game series in the first round, a 5-game series in the second round, and 7-game series for the Championship Series and World Series, how should organizations adjust their team-building approach? Please limit your answer to 300 words.