Homework 2

Case Study: Was the Mets' 2025 season "bad luck"?

In Major League Baseball (MLB), playoff spots are determined as follows: each league (American and National) sends six teams to the postseason: three division winners and three wild-card teams with the highest win percentages (PCTs) among non-division winners. Because of this format, even a small difference in winning percentage can dramatically affect playoff qualification. Although in 2025, both Cincinnati Reds and New York Mets finished with an 83–79 win-loss record, the Reds owned the head-to-head tiebreaker after going 4–2 against the Mets this season, which earned them the wildcard spot in the playoffs.

After narrowly missing the playoffs in the 2025 MLB season, New York Mets' General Manager David Stearns wants to understand what went wrong: did the Mets just have an average performance this season or did bad luck play a role in their inability to secure a spot in the playoffs?

To investigate, the GM hires you as a sport analytics consultant to conduct a pilot analysis. You decide to use the Pythagorean Expectation model that we covered in class to assess the team's performance.

Please address the questions below to determine the role of luck in the Mets' 2025 season. Before beginning your analysis, make sure to review the "Pythagorean Theorem" chapter from the *Mathletics* textbook assigned for this week to ensure you have a solid understanding of the concept.

Part I - Analysis in R

You will be working with a dataset that contains the regular season records for all teams in Major League Baseball's National League.

First, run the analysis in the provided R file. Some lines of code have already been written, while others you will need to complete. Specifically, complete the steps below following the instructions outlined in the R file:

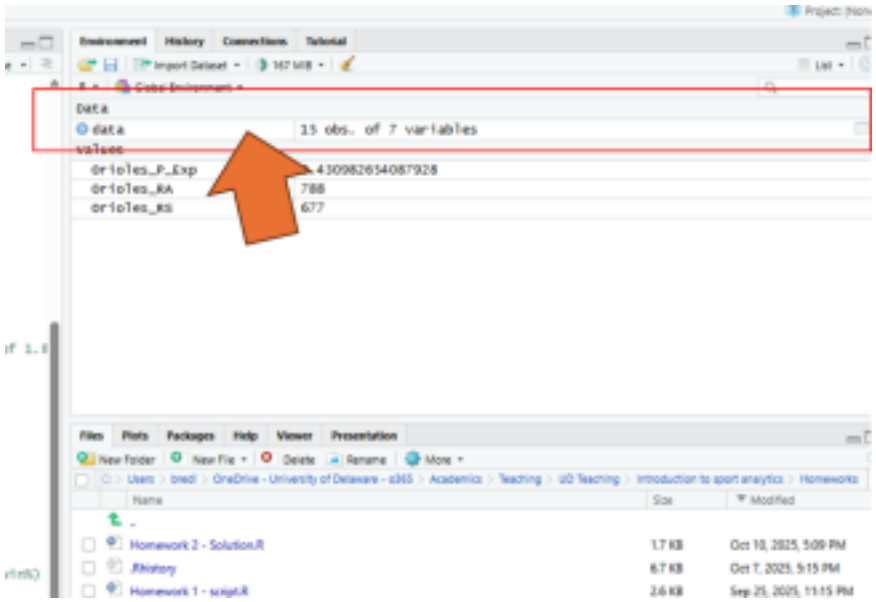
1. Create a new variable estimating the Pythagorean Expectation for 2025.
2. Create a new variable estimating the win% for 2025.
3. Create a new variable estimating the luck differential (i.e., the difference between Win% and Pythagorean Expectation).

$$\text{Luck} = \text{Actual Win\%} - \text{Pythagorean Expectation}$$

After completing your analysis, proceed to Part II to interpret your results.

Part II – Interpretation

To assist you in answering the following questions, after completing the analysis in Part I, click on your data frame object in the Global Environment tab (located in the top right corner of your RStudio screen).



You should be able to see your updated table with new columns added.

1) Definition of Pythagorean Expectation (10 points)

- Define what Pythagorean Expectation is.
- Include a brief discussion of the formula.

Pythagorean Expectation is a sports analytics formula that attempts to predict the percentage of games that a baseball team "should" win. Originally developed by Bill James, the result is considered a win percentage prediction. It is based on the team's runs scored and runs allowed for the entire 162-game season. The formula is very simple. The numerator is the team's number of runs scored raised to a specific exponent, typically 2, 1.83, or 1.79. The denominator is the addition of this term and the team's number of runs allowed raised to the exact same exponent.

2) Assessing 2025 Pythagorean Expectation (10 points)

Inspect the column of the table that displays the Pythagorean Expectation. You can sort this column in ascending or descending order by clicking on the column name.

- Were the Mets *expected* to make the playoffs?

After sorting the Pythagorean Expectation column, the Mets had the 6th best predicted win percentage in the National League from the formula with 1.83 as the common exponent. Because there are 6 teams that make the playoffs, the Mets were actually expected to make the playoffs based on Pythagorean Expectation.

- Which team(s) would lose their playoff spot(s) if the playoff spots were based on the Pythagorean Expectation?

The Cincinnati Reds would lose their playoff spot to the New York Mets if spots were based on the Pythagorean Expectation formula with 1.83 as the exponent.

3) Quantifying Luck (10 points)

Inspect the column that displays the luck differential. You can sort the column by clicking on the column name.

- Which two National League teams most overperformed expectations?

The Miami Marlins and Washington Nationals most overperformed expectations. Both teams had positive luck differentials, which means their actual win percentages were larger than their predicted win percentages from the Pythagorean Expectation formula with 1.83 as the exponent.

- Which two National League teams most underperformed expectations?

The Atlanta Braves and Chicago Cubs most underperformed expectations. Both teams had negative luck differentials, which means their actual win percentages were smaller than their predicted win percentages from the Pythagorean Expectation formula with 1.83 as the exponent.

4) Acknowledging the Limitations of Pythagorean Expectation (10 points).

As part of any analytics project, you need to disclose limitations associated with the methodology used. What limitations does the Pythagorean Expectation model have when trying to predict outcomes?

The first obvious limitation is that it only uses runs scored and runs allowed, which do not paint the full picture of what happens across a full 162 game season. Next, the exponent used is up for debate. Some models use 2, while others use 1.79 or 1.83. The exponent is very dependent on the sport and situation, so it can be hard to perfectly calculate. Third, the formula does not perfectly consider the context of the season, especially injuries. Like other formulas and statistics, Pythagorean Expectation only considers the outputted runs scored and runs allowed without considering how they happened.

5) Discuss Other Applications of Pythagorean Expectation (10 points)

After you present your results to the GM, he becomes very interested in learning more about the Pythagorean Expectation model and the insights it can generate. In approximately 100 words, name and discuss at least two other ways the Pythagorean Expectation can be applied in sport analytics (i.e., other than assessing “luck”).

Pythagorean Expectation can also be used for both outcome predictions and player contribution. Generally speaking, Pythagorean Expectation is more accurate than actual wins when it comes to predicting future performance. This metric can be calculated by using prior wins and the number of remaining games. It can be used for player contribution when a player gets traded to a different team. The runs "created" by the incoming player(s) can be compared to the runs "created" by the outgoing player(s), and the difference can then be considered in a win percentage formula for the remaining number of games.

Extra Credit (10 points)

The accuracy of the Pythagorean Expectation formula depends on the exponent used in the equation.

Design a strategy to compare the accuracy of formulas with exponents of 2, 1.83, and 1.79 using the data you used in this assignment. Which exponent yields the most accurate prediction for our 2025 regular season results, and why might that be the case?

This problem can be solved by first creating predictor variables for each of the three different Pythagorean Expectation formulas. In my R script, the original one with 1.83 as the exponent was already created. Therefore, two variables were created for the Pythagorean Expectation formulas with 2 and 1.79 as the exponents. Second, correlation values can be created for the three relationships that the predicted win percentages have with the actual win percentages for the 2025 National League standings. 5 decimal places were used for all three correlations, as the three values were incredibly close. The highest correlation was 0.97614 between actual 2025 win percentage and the Pythagorean Expectation win percentage with 1.79 as the exponent.