

# Player Evaluation Activity

## Project Aim

---

This project is designed to help you understand how labor inputs are related to revenue in Major League Baseball (MLB). We will do this by analyzing the offensive skills of free-agents in baseball and the relationship of these skills to firm output (measured as team ticket sales). We will use a simplified model of player skill based on the conclusions of the article “An Economic Evaluation of the Moneyball Hypothesis” (see the Papers section). In order to complete this project, you will collect data from several sources (provided in the list below) that measure individual and team statistics, salary information for individual players, and attendance and ticket cost information for each team in the MLB.

After completing the analysis, you will write a short paper explaining what you did, why it is informative, and your conclusions. The paper should be 3-5 pages long, and specific points to address in the paper will be provided below. The following sites will be useful as you collect data.

- Player Salary Database (be sure to select the last completed season):  
<http://www.spotrac.com/mlb/free-agents/>
- Player Batting Statistics Database (be sure to select the last completed season):  
<http://www.baseball-reference.com/>
- Team Batting Statistics (be sure to select the last completed REGULAR season statistics):  
[http://www.espn.com/mlb/stats/team/\\_/stat/batting/](http://www.espn.com/mlb/stats/team/_/stat/batting/)
- Team Attendance Statistics (be sure to select the last completed season):  
<http://www.espn.com/mlb/attendance>
- Team Win Statistics (be sure to select the last completed REGULAR season):  
<http://www.espn.com/mlb/standings>
- Average Ticket Prices by Team (be sure to select the last completed season):  
<https://www.statista.com/statistics/193673/average-ticket-price-in-the-mlb-by-team/>

## The Basics

---

You will create a Google Drive folder and share access with each of your teammates and your instructor.<sup>1</sup> You will use Google Sheets so that you can work collaboratively with your teammates to collect data throughout the project. You will also use a Google Doc to complete your write-up. All elements of the project (including plotting, estimating trendlines, and graphical analysis) can be completed in Google Drive using Docs and Sheets.<sup>2</sup>

## Teams

---

The class will be divided into multiple “teams.” These teams should be made up of no more than four individuals. You will work together to operate as the general manager of an

imaginary baseball franchise, and will make personnel decisions based on a player's previous statistics.<sup>3</sup> You are expected to be a contributing member of your team throughout the process of completing this project.

## Part 1 – Understanding Firm Output

---

After reading *An economic evaluation of the Moneyball hypothesis* by Hakes and Sauer, you will collect team-level data for all 30 MLB teams, and use it to assess the impact of offensive success in baseball on the revenue generated by the firms (teams) in MLB. The following instructions describe how to collect and organize the information in a spreadsheet.

1. Place the team names in the first column of your spreadsheet, with the first cell labeled “Team.”
2. Label the second column “OBP,” and collect the **team** On-Base Percentage (OBP) for each team in MLB, and enter the values in the cells corresponding to those teams (use the information from the **regular season**).
3. Label the third column “SLG” and enter each team’s **regular season** Slugging Average (SLG) into the cells corresponding to each team.
4. The paper by Hakes and Sauer (2006) lists On-Base Percentage and Slugging Average as the most important factors in determining the number of runs (and, ultimately, wins) that a team earns. The authors find that OBP is twice as important in determining these outcomes as SLG. Create an index of offensive production in the fourth column, and label the column “Index.” The index should be calculated using the following formula:

$$\text{Index} = 100 \times (2 \times \text{OBP} + \text{SLG})$$

*Note: We multiply by 100 in order to have numbers that are easier to read.*

5. Label column 5 “WinPercentage.” Collect the **regular season** win percentage of each team, multiply by 100 (so that a team with a .500 win percent has a value of 50.0), and enter the data in the cell corresponding to the correct team.
6. Label column 6 “Attendance” and column 7 “TicketPrice.” Collect the **total** attendance for each team, as well as the **average** ticket price for each team during the most recent completed season, and enter the data in the corresponding cells of the spreadsheet.
7. Label column 8 “TotalRevenue,” and calculate the cell according to the following equation:

$$\text{Total Revenue} = \text{Attendance} \times \text{Ticket Price}$$

8. Using the information you have collected and calculated, create three scatter plots. In each scatter plot, include the linear trendline and select the option to display the equation of the trendline. The three plots you should present are:
  - (a) Team Revenue (y-axis) to Index (x-axis)
  - (b) Team Revenue (y-axis) to Win Percentage (x-axis)

(c) Win Percentage (y-axis) to Index (x-axis)

Each trendline will be formatted as a traditional linear equation of the form  $y = mx + b$ . For our purposes, we care about the **slope** of the trendline (the  $m$  in the equation above). This slope tells us the effect of an increase in our x-variable on the value of our y-variable.

9. Be sure to answer the following questions in your writeup:

- What is the effect of a one-unit increase in the Index on Total Revenue?
- How about Win Percentage?
- Is the value of the coefficient from the third plot (WinPercent vs Index) times the coefficient from the second plot (TotalRevenue vs WinPercent) close in value to the coefficient from the first plot? If true, this equality could be written  $\text{Coefficient}_2 \times \text{Coefficient}_3 = \text{Coefficient}_1$ . What might this suggest generates demand for sporting events?

## Part 2 – Estimating the Value of Player Talent

---

In this portion of the project, you will make use of the information about baseball franchises from Part 1 to estimate how valuable the contribution of free-agent baseball players should be to their new teams. This information will allow you to determine whether or not you believe that a given player is over- or under-valued according to their offensive statistics only.

1. Using the list of free agents for the current (or most recent) offseason, select 30 players *randomly*.<sup>4</sup> Note that all players should be position players, not pitchers. Please be sure to explain the process that you use to select the 30 players at random. Place these 30 players' names in the first column of a new spreadsheet (You may also use a column for first name and another for last name if you choose).
2. Using the Player Batting Statistics Database listed at the top of the assignment, obtain the most recent OBP and SLG statistics for each of the 30 players on your list. Label the second column "OBP" and the third column "SLG," and enter the values in the appropriate cells. *Note: For some players, you may have to use minor league statistics to obtain recent OBP and SLG values. We will pretend for the purposes of this project that these numbers accurately reflect how the player will perform in the major leagues, even though this is a VERY strong assumption.*
3. Collect each player's salary from the Player Salary Database. Use the salary for each player in the *upcoming* season, not their past salary. If a player has no salary, assign them a salary of \$535,000 (approximate league minimum salary). Enter these values in the fourth column, and label the column "Salary."
4. In the fifth column, calculate the player's offensive index using the same equation as you used for the team offensive index. (See Part 1 Step 4).
5. Label the sixth column "Difference." Subtract the offensive index value for a Mendoza Player (one who has an "OBP" of .250 and an "SLG" of .300) from the index value calculated for each player in the previous step.

$$\text{Difference} = \text{Player's Index} - \text{Mendoza Index} \quad (1)$$

6. Label the seventh column “Marginal Product.” Divide the “Difference” value of each player by nine, because a starter on an MLB team represents 1 in 9 at-bats for his team, and so his index value will represent  $1/9^{\text{th}}$  of the team index value.<sup>5</sup>

$$\text{Marginal Product} = \text{Difference}/9 \quad (2)$$

7. In the eighth column, multiply the “Marginal Product” value for each player by revenue generated at the team level for each additional unit of offensive output. Revenue generated is equal to \$535,000 plus the **slope** of the trendline from the Index vs. Revenue plot in Part 1 of the project multiplied by the actual output.<sup>6</sup> Label this column “MRP” (Marginal Revenue Product). This value is an estimate of the revenue generated by a team that signs a particular player, and in a competitive market should be the amount paid to the player for signing with a team.

$$\text{Marginal Revenue Product} = \$535,000 + \text{Marginal Product} \times \text{Marginal Revenue} \quad (3)$$

8. In the ninth column, use an “if” statement to assign a value of “undervalued” to players whose MRP is greater than their Salary, and a value of “overvalued” to players whose MRP is less than their Salary. Remember, these values are stored in columns 4 and 7.
9. Make a scatterplot of the true salary (x-axis) and the predicted salary (y-axis) of each player in your sample.
10. Using the plotting tools in Google Sheets, determine the  $R^2$  value using projected salary and actual salary. Using the CORRELL function, calculate the correlation coefficient for these two variables.
11. Be sure to answer the following questions in your writeup:
- How close are these estimates to the true salary of each player? What do you think might cause the differences?
  - What do the  $R^2$  and correlation coefficient tell us about how players are valued in Major League Baseball compared to our projected values for free agents?

## What to Turn In

---

Write out, in three to five double-spaced, typewritten pages:

1. What you did, including how you calculated the cells in your spreadsheets.
2. Why this process might be valuable to a professional baseball team, and how the information that you calculated should be used in order to make a baseball team more profitable.

3. Conclusions. Explain what this project tells you about the market for baseball players, and how this information could be applied to other markets.

## Papers

---

Hakes, Jahn K., and Raymond D. Sauer. "An economic evaluation of the Moneyball hypothesis." *The Journal of Economic Perspectives* 20.3 (2006): 173-185.

## Notes

<sup>1</sup>Creating Files and Folders: [https://support.google.com/docs/answer/49114?hl=en&ref\\_topic=4722291](https://support.google.com/docs/answer/49114?hl=en&ref_topic=4722291)

Sharing Files and Folders: [https://support.google.com/docs/answer/2494822?hl=en&ref\\_topic=4722291](https://support.google.com/docs/answer/2494822?hl=en&ref_topic=4722291)

<sup>2</sup>Help with any problems using Google Docs (or sheets) throughout the project can be found by searching the help documentation available at <https://support.google.com/docs/?hl=en\#topic=1382883>

<sup>3</sup>For information on why using past performance to model future pay and performance in sports is frequently a dangerous idea, see *Stumbling on Wins* (by Berri and Schmit) and *Scorecasting* (by Moscovitz and Wertheim).

<sup>4</sup>Google Sheets has a built in "RANDOM" function that is useful in selecting players randomly.

<sup>5</sup>Note that if you were to calculate a team's index value from the composite of all players' MRP values, you should start with a baseline index of 80. This index represents a team composed only of Mendoza players (the worst team that should ever be fielded in our imaginary world, since another Mendoza player can, in our theoretical world, always be found). Adding the MRP values of all players to this base of 80 would then generate an accurate team-level index.

<sup>6</sup>We add the league minimum wage to a player's MRP because this is the amount that must be paid to players whose Marginal Product is equal to 0. All valuation should take this into account, since we are trying to determine how much value a player generates **beyond** the value generated by a Mendoza Player.