Moneyball and sports analytics

Tool: Linear regression

The Analytics Edge: Using tools of linear regression, managers can accurately value players and minimize the risk of gut feeling evaluation of players. The hitter salaries in baseball did not accurately reflect the contribution of batting skills to winning games. Once this relationship was better evaluated, it could be exploited to gain an advantage by constructing a team carefully using more relevant baseball statistics. The Moneyball case discusses this and how a strong team was built despite the team having much lesser money to buy players.

Overview

In 2003, Michael Lewis wrote the book Moneyball which discusses how Oakland Athletics, a baseball team playing Major League Baseball in America identified a group of undervalued professional baseball players to turn themselves into one of the most successful franchises. How did Oakland Athletics, one of the poorest teams in Major League Baseball win so many games?

The story is abut Billy Beane, the Oakland Athletics general manager who was willing to discard old wisdom to get an edge over big money. In 2001, Moneyball was made into a Hollywood movie starring Brad Pitt.

Figure 4.1: The average number of wins for each team in Major League Baseball, plotted as a function of their average yearly payroll (from 1998 to 2001). The green triangle is the Oakland Athletics; they were winning a large number of games, with one of the lowest payrolls. For comparison, the red triangle is the Boston Red Sox, and the blue triangle is the New York Yankees, two of the richest teams in Major League Baseball.

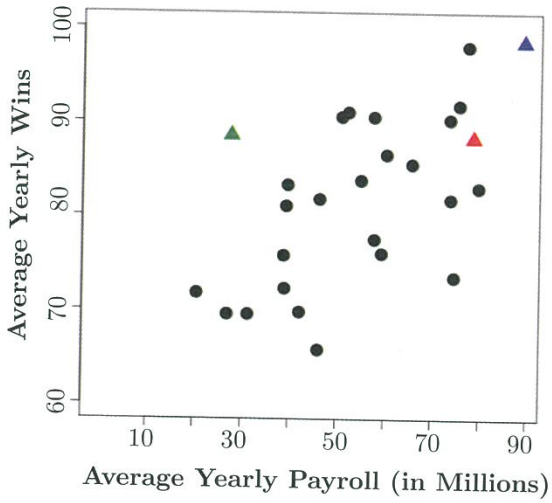

Figure 0.1: Salaries and wins of teams in 1998-2001

Baseball

Baseball is a played between two teams with nine players in the field on each team. In Major League Baseball, the game is played in nine innings in which each team gets to bat and score runs while the other team pitches and defends in the field. An innings is broken into two halves where one team bats in the first half and then the other team bats. The teams switch each time, the defending team gets three of the players from the batting team out. The winner is the team with the most runs after nine innings.

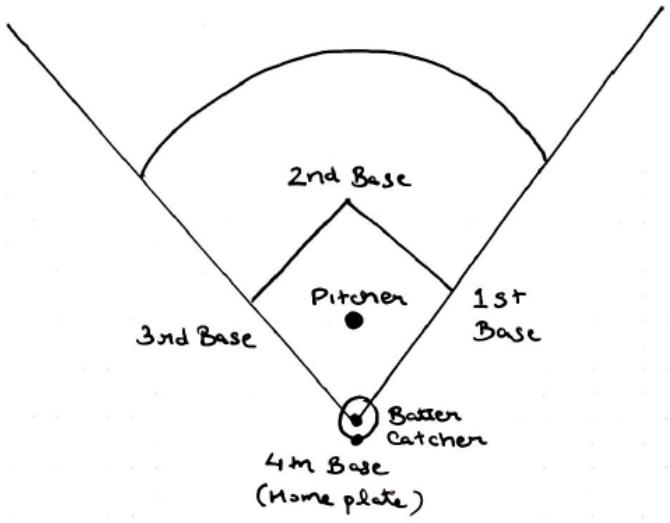

Figure 0.2: Schematic of baseball field

Sabermetrics: Empirical analysis of baseball, especially baseball statistics that measure in-game activity (the search of objective knowledge about baseball).

Key people: Bill James is considered as the father of sabermetrics as he authored a series on books on baseball statistics in the 1970s. Billy Beane was the General Manager of Oakland Athletics who played Major League Baseball for a few years in the 1980s. He was picked by talent scouts as a potential star but his own playing career failed to meet the expectations. As a General Manager, his record in the late 1990s to early 2000s was:

Table 0.1: Record as a manager. * indicates the years the Oakland Athletics made it to the playoffs

Year	Won	Lost	Played
1998	74	88	162
1999	87	75	162
2000	91	71	162*
2001	102	60	162*
2002	**103**	**59**	**162***
2003	96	66	162*
2004	91	71	162*

Paul Podesta helped Billy Beane use quantitative tools to manage Oakland Athletics. He was later the vice president of player development and scouting for the New York Mets.

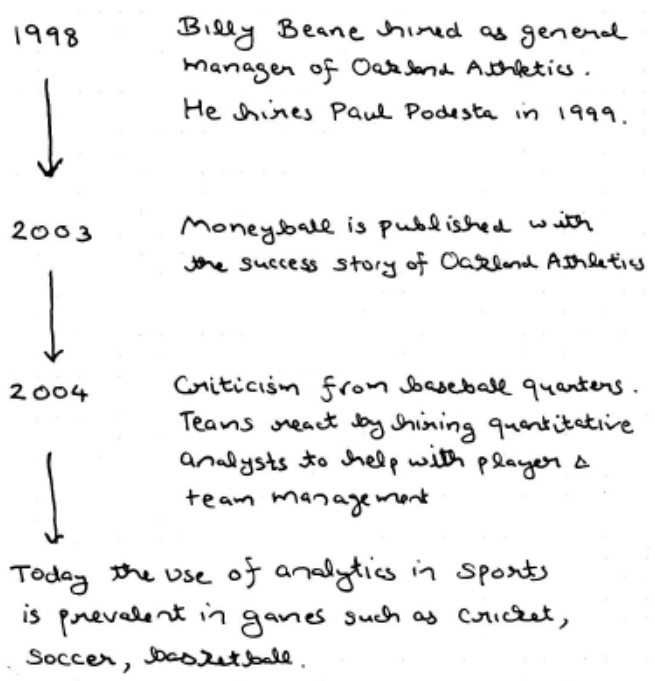

Figure 0.3: Timeline of events

Baseball statistics

While baseball is a team sport, success is a function of the achievements of individual players which can be easily observed. To measure batting skills, some of the commonly used statistics are:

1. Batting average is a measure of how often a batter reaches base by hitting safely. This is measured as:

$$\text{Batting average } BA = \frac{\text{Number of hits}}{\text{Number of at bats}}$$

   Batting average is however a crude measure of skill since it ignores the added productivity of hits that allow you to run more than a single base (singles and home runs are counted in a similar manner).

2. Slugging percentage is a more refined measure:

$$\text{Slugging percentage } SLG = \frac{\text{Total bases}}{\text{Number of at bats}}$$

   This counts doubles twice as much as singles, homeruns four times as much as singles and so on.

3. On base percentage (OBP) is the fraction of plate appearances (includes at bats and walks) in which the player reached base successfully through either a hit, walk or otherwise. Unlike batting averages, this statistic does not care about how the player gets on base.

4. On base plus slugging (OPS) is a statistic that simply is defined as the sum of on base percentage and slugging percentage:

$$OPS = OBP + SLG$$

Note that while OPS adds the two statistics together, it is still unclear what the relative importance of these two statistics are.

For example, suppose a team has OBP = 1. This means that every player who comes to bat, gets to first base and thus the team never gets out. Thus the team would score an infinite number of runs in this case.

On the other hand, suppose a team has SLG = 1. Then it is possible in this case, for each player to get to first base and the team to score an infinite number of runs. It is however also possible that, the first player gets a home run (4 bases) while the remaining three players do not get to base at all. In this case, the team scores only 1 run. Even the, SLG = 1. Thus, it is seems intuitive that OBP should be more important than SLG, but the question is how much more?

Players

In 2002, Oakland Athletics won 103 games and made it to playoffs. This was despite losing one of their star players Jason Giambi to the New York Yankees. He was one of the stars winning the Most Valuable Players in the American League in 2000. To replace him and two other player stars Johnny Damon and Jason Isringhausen, Oakland Athletics hired an injured Scott Hatteberg, unorthodox pitcher Chad Bradford, past his prime David Justice for a smaller prices. Scott Hatteberg was player with a good OBP but was valued much lesser when the Oakland Athletics signed him.

Summary

Data: Data on player performance in terms of statistics on batting, fielding, pitching is readily available from sites such as http://www.baseball-reference.com

Model: Linear regression is used to determine the relative importance of various statistics on the team performance.

Decision: The model identifies the key performance statistics that are correlated to team wins. Based on this, it is possible to build teams by picking the right players.

Value: A mathematical model provides a new insight that can identify undervalued players. This helps to build better teams. Many sports teams now have quantitative analysts who aid in team management.