Fair Play

Contents

1 Introduction **3**

2 Restatement of Problem **3**

3 Development of Model **3**

3.1 First Model 3

3.1.1 Assumptions and Justifications 3

3.1.2 Fairness Rating System 4

3.1.3 Testing and Conclusions 6

3.2 Second Model 7

3.2.1 Modifications to Fairness Rating System 7

3.2.2 Testing and Conclusions 7

3.3 Final Model 8

3.3.1 Modifications to Fairness Rating System 8

3.3.2 Testing and Conclusions 10

4 Analysis of Final Model **10**

4.1 Strengths 10
4.2 Weaknesses and Potential Improvements 10

5 Conclusion 11

1 Introduction

Major League Baseball schedules games between thirty baseball teams at any of thirty different venues (baseball parks). The League depends on fan excitement and game attendance to make a profit. As with any profitable venture, the MLB constantly searches for ways to increase profits.

In an attempt to increase game attendance, the League decided to research the fairness of different baseball venues. However, as the League has neither the time nor the resources to perform the research directly, they have contracted our team to find and interpret fairness trends for them.

2 Restatement of Problem

Major League Baseball requested us to consider the venues of the following six teams: the Atlanta Braves, the Colorado Rockies, the New York Yankees, the Anaheim Angels, the Minnesota Twins, and the Florida Marlins. They want to know how "fair" or "unfair" each park is. Additionally, to plan for future venue remodeling and/or construction, the MLB wants to know the optimal "setting" for Major League Baseball. To aid our research, the MLB provided the outfield dimensions, wall height, and area of fair territory for each park.

3 Development of Model

3.1 First Model

3.1.1 Assumptions and Justifications

The assumptions are listed in bold font, while the justifications follow in the normal font.

1. **The factors listed below are the only determinants of fairness.** Because we are only asked about each venue's fairness, we assume that all other aspects of the game (teams, umpiring, etc) are "fair." Therefore these other aspects do not affect the overall fairness.
2. **Each factor affects the fairness equally.** We assumed this as a baseline weighting system to potentially be changed in future models.

3. **"Fairness" is defined as the equilibrium of offensive and defensive advantages.** This is explained in more detail below.

3.1.2 Fairness Rating System

The fairness rating coefficient measures the equilibrium, or lack thereof, of offensive and defensive advantages at a venue. Consequently, we created the following rules for determining and using the rating coefficient. As with the general assumptions above, we have listed the rules in bold font, while their respective explanations and justifications follow in the normal font. A more in-depth discussion of the Fairness Rating System is included as item A1 in the appendix.

Determination and Usage

- $\boldsymbol{F_c = A_{\text{defense}} - A_{\text{offense}}}$. F_c, the fairness rating coefficient, tells us how fair or unfair a venue is. A "fair" venue will create an equilibrium between A_{defense} (the defense's advantages) and A_{offense} (the offense's advantages). The magnitude of F_c allows us to determine a venue's fairness – the magnitude increases as fairness decreases – while its sign tells us whether an unfair venue favors the offense or the defense.
- **A factor which favors the offense will increase the value of $\boldsymbol{A_{\text{offense}}}$ by 1 (one point). Similarly, a factor benefitting the defense will increase $\boldsymbol{A_{\text{defense}}}$ by one point.** This implements a uniform weight for each factor, as discussed in the general assumptions for this model.
- $\boldsymbol{-10 \leq F_c \leq 10}$. This results from our use of 10 factors, which are worth plus/minus one point each.

Rules

1. **The width is the distance between foul poles. Widths greater than the average favor the offense, while those less than the average favor the defense.** A wider field creates a larger area that the defense must cover; the defense must be distributed more sparsely. As a result, the outfielders must work harder and it is less probable that they will be able to reach the ball quickly.
2. **The distance to center field is the distance from home plate to the home-run fence at center field. A short distance favors the offense; a longer distance favors the defense.** A longer center field distance generally requires the batter to hit

the ball farther for a home run. As a result, a shorter distance will allow more home runs.

3. **The surface type refers to whether a venue uses grass or artificial turf on the field. Artificial turf is beneficial to the offense, while the defense prefers grass.** Artificial turf is more "springy." When the baseball hits artificial turf, the ball tends to bounce much higher and in a more unpredictable manner, making it more difficult to catch. On the other hand, a baseball bouncing off of grass will bounce lower and in a more predictable fashion; it is easier to catch [17].

4. **The humidity was measured in the city containing or nearest to each venue, and averaged over the period from 1 April to 30 September, 2002. A low measurement (below 40%) favors the offense, while a high measurement (above 60%) is preferable for the defense. A median measurement (from 40% to 60%) is considered neutral.** The MLB regular season extends from April to September. A humid environment applies friction to the baseball (in the form of air resistance; [22]). Consequently, an environment with high humidity tends to slow the ball down and shorten its travel distance [15]. On the contrary, an environment with low humidity will not slow the ball down as much, allowing it to travel farther [21].

5. **The wind speed was measured in the city housing or nearest to each venue, and averaged for the period from 1 April to 30 November, 1996. An above-average wind speed assists the offense, while a below-average speed helps the defense. The air conditioning in a dome benefits the defense.** A high wind speed affects the baseballs flight path, making it more difficult to catch. Low wind speed allows the flight path to be more predictable, making it easier to catch. Statistically, batters have hit more home runs without air conditioning than when it was on. Our team only detected the discrepancy in the months (November for wind speed but September for humidity) after the completion of the second model; we corrected the discrepancy in the third model.

6. **We distinguished between open-air stadiums and closed dome venues. The offense prefers a dome. We considered the open-air stadium to be neutral.** The white inside of a dome venue does not provide much contrast with the white baseball, making a successful catch difficult. We feel, however, that an open-air stadium confers an advantage on neither the offense nor the defense.

7. **The elevation is the vertical distance from sea level to the base of the venue. Extremely high elevations benefit the offense. Lower elevations are considered neutral.** As elevation increases, air density decreases. Consequently, at extremely high elevations, the lower-density, less resistant air allows the ball to travel farther than at lower elevations. We felt that the air density wouldn't increase enough at lower elevations to offer a significant advantage to the defense.

8. **We defined a field's coefficient of symmetry as the sum of the difference between the left field distance (lf) and right field distance (rf) and the difference between the left center field and right center field distances (lcf and rcf, respectively). Simply put, $S_c = |lf - rf| + |lcf - rcf|$. A large coefficient of symmetry benefits the offense; a small one helps the defense.** The coefficient of symmetry allows us to assign a quantitative value to an important aspect of the venue which may be difficult to describe and/or measure. A large coefficient results from a highly asymmetric field, which forces the defense to redistribute outfielders in a perhaps less efficient configuration. Additionally, an asymmetric field will generally have "pockets" which are close to the batter and therefore which reduce the home-run distance. On the other hand, a highly symmetric field allows the outfielders to position themselves efficiently and ensures that the home run distance will be as far as possible no matter where a batter aims.

9. **The wall height is the vertical distance from the base of the wall to the top. We calculated the wall height for each venue by averaging the values for left field, center field, and right field. A below-average height is beneficial to the offense, whereas an above-average height aids the defense.** A high wall requires the batter to hit the baseball higher to earn a home run. The batter, however, prefers a low wall so that less ball height is required for a home run.

10. **We considered the "playability" of the home run fence: whether the fence is conducive to helping outfielders catch high fly balls (and potential home runs). A very playable fence helps the defense. Certain fences are not conducive to use by the outfielders, and consequently aid the offense. We deemed that fences which had no distinct advantageous characteristics for either the offense or the defense were neutral.** A "playable" fence, such as a chain link fence or one with padded walls [20], allows the outfielders to climb on it, allowing them to gain the extra elevation needed to catch a ball that would otherwise be out of reach. On the other hand, if the outfielder attempts to use an "unplayable" fence, such as the brick wall at Wrigley Field, the experience will most likely be painful or disadvantageous to the outfielder in some other fashion.

3.1.3 Testing and Conclusions

We tested our model by comparing the fairness rating coefficient to historical data on fairness. Our rating system correctly classified the Angels' and Braves' venues as so-called "pitcher's parks," and both the Rockies' and Twins' venues as "hitter's parks." However, our system rated Yankee stadium as neutral and reported an overall offensive advantage for the Marlins' stadium. Both are historically "pitcher's parks."

3.2 Second Model

3.2.1 Modifications to Fairness Rating System

After comparing the results of our first model to real data, we adjusted some of the scoring methods to help our model to more accurately fit the data. The changes are enumerated below in bold, followed by their respective justifications in normal font.

1. **Only extreme outfield widths earn advantage points. Median values are now considered neutral. Extremely wide parks favor the offense, while very narrow parks aid the defense.** The width of a field is a relatively long characteristic. We decided that the relatively small difference between the Twins' field width (473.9 ft) and the Braves' field width (470.2 ft) is not significant. However, the difference between the Rockies' and Yankees' field widths (492.9 ft and 446.9 ft, respectively) is very significant.

2. **The relative humidity at the Twins' stadium is not considered.** Because the Twins play in a closed dome, the relative humidity (which is measured in the open air) is inconsequential.

3. **In addition to the point for extremely high elevations, elevations below 100 ft earn an advantage point for the defense.** Though we did not at first consider the increased air density at very low elevations significant, introducing this rule helped our model to match our historical data. This may, indeed, signify some low-elevation effect, or it may control for some factor which we did not take into account.

4. **When scoring symmetry, we account for the geometric shape of the outfield. A more awkward shape confers an advantage to the offense.** Our heuristic ranked the Angels' park as very symmetric. However, the heuristic did not account for the off-center placement of "center field." This change remedies the shortcoming.

3.2.2 Testing and Conclusions

Our modifications reaped a significant increase in the overall accuracy of our model. The venues of the Marlins and the Yankees, which the first model classified incorrectly, are now both correctly reported as "pitcher's parks." All the other venues, which the first model accurately classified, are still identified correctly by this model.

3.3 Final Model

3.3.1 Modifications to Fairness Rating System

The uniform-weight scoring system utilized for the first and second models worked fairly well, but it was also rather inaccurate: it did not account for the varying degrees to which the ten factors affect offensive and defensive performance. To remedy this inaccuracy, we developed a weighted scoring system that assigns a "weight," or importance, to each of the ten factors. To help balance the system, we ensured that an ideal field for the offense would result in $F_c = -36$. Similarly, an ideal defensive field will result in $F_c = +36$. An optimal field still receives a score of zero (0).

We first ranked the factors in order of importance. We then assigned a score for different extreme (and, in certain cases, intermediate) values of the factor. Factors which aid the offense receive a negative score, while those that help the offense receive a positive score (remember that $F_c = A_{\text{defense}} - A_{\text{offense}}$). We list the weighted scores below in bold, each accompanied by either a new justification in normal text, or a short summary of our previous justification(s) in italic text.

1. $\textbf{Surface type :} \begin{cases} \textbf{Natural Grass} & \mathbf{+9} \\ \textbf{Artificial Turf} & \mathbf{-8} \end{cases}$
   A baseball bouncing off of turf is less predictable than one bouncing off of grass.

2. $\textbf{Elevation } \boldsymbol{h} \textbf{ :} \begin{cases} \boldsymbol{h} < \textbf{100 ft} & \mathbf{+3} \\ \textbf{100 ft} < \boldsymbol{h} < \textbf{1000 ft} & \mathbf{0} \\ \textbf{1000 ft} < \boldsymbol{h} < \textbf{3000 ft} & \mathbf{-3} \\ \textbf{3000 ft} < \boldsymbol{h} & \mathbf{-7} \end{cases}$
   Air at higher elevations is less dense, and therefore provides less resistance to the baseball.

3. $\textbf{Width } \boldsymbol{w} \textbf{ :} \begin{cases} \boldsymbol{w} < \textbf{460 ft} & \mathbf{+7} \\ \textbf{460 ft} < \boldsymbol{w} < \textbf{480 ft} & \mathbf{0} \\ \textbf{480 ft} < \boldsymbol{w} & \mathbf{-6} \end{cases}$
   A wide field forces outfielders to cover more area, reducing the probability that they will get to the ball quickly.

4. $\textbf{Field symmetry } \boldsymbol{S_c} \textbf{ :} \begin{cases} \boldsymbol{S_c} \leq \textbf{15 ft} & \mathbf{+3} \\ \textbf{15 ft} < \boldsymbol{S_c} < \textbf{25 ft} & \mathbf{0} \\ \textbf{25 ft} < \boldsymbol{S_c} & \mathbf{-5} \\ \textbf{Angels' stadium} & \mathbf{-2} \end{cases}$
   An asymmetric field causes outfielders to assume a less efficient configuration, making a timely catch and/or recovery throw less probable. Our manual adjustment for the Angels' field remains.

5. **Wind speed v :** $\begin{cases} v < 3 \text{ mph} & +4 \\ 3 \text{ mph} < v < 9 \text{ mph} & 0 \\ 9 \text{ mph} < v & -5 \\ \text{AC (Dome)} & -4 \end{cases}$

   We obtained and averaged wind speed data from the city housing or closest to each venue during the 1996 MLB season (1 April to 30 September; this corrects our earlier discrepancy). In the dome, the air conditioning dramatically reduces humidity.

6. **Dome/Stadium design :** $\begin{cases} \text{Open Stadium} & +3 \\ \text{Enclosed Dome} & -5 \end{cases}$

   As stated before, the inside of a dome provides little contrast with the baseball, making it difficult to see and catch. On the contrary, the dark sky of an open stadium (most games are played in the evening or at night) provides very good contrast with the baseball, making it much easier to see.

7. **Center field distance d :** $\begin{cases} 410 \text{ ft} < d & +3 \\ 400 \text{ ft} < d < 410 \text{ ft} & 0 \\ d < 400 \text{ ft} & -1 \end{cases}$

   A long center field distance increases the required distance for the batter to complete a home run. A shorter distance makes it easier for the batter to hit a home run.

8. **Wall height h :** $\begin{cases} 15 \text{ ft} < h & +6 \\ 8 \text{ ft} < h < 15 \text{ ft} & 0 \\ h \leq 8 \text{ ft} & -2 \end{cases}$

   A low wall decreases the required height for a hit to become a home run, making home runs easier.

9. **Wall playability :** $\begin{cases} \text{Very Playable} & +4 \\ \text{No distinct qualities} & 0 \\ \text{Not Very Playable} & -2 \end{cases}$

   A playable wall allows the outfielder to obtain the extra elevation needed to catch high fly balls and potential home runs. An unplayable wall strongly deters the defense from obtaining this advantage.

10. **Climate :** $\begin{cases} \text{Cold and Dry} & +1 \\ \text{Cold and Humid} & 0 \\ \text{Hot and Dry} & 0 \\ \text{No distinct qualities} & 0 \\ \text{Hot and Humid} & -1 \end{cases}$

    Further research showed that humidity actually helps the ball travel farther; it doesn't resist it as we believed during the first two models. Additionally, high temperature lowers air density, helping the ball travel farther. We combined temperature with our previous humidity factor because together they influence the distance the baseball will travel. However, neither influences the distance to a great degree.

3.3.2 Testing and Conclusions

As before, we compared the rating coefficients we obtained to historical data about each park. With the different weighting system, we obtained the following scores and classifications. A tabular itemization and discussion of the scores for each venue are included as items A3 and A4 in the appendix. As is visible, all of the classifications from our model match their historical counterparts.

Team	Score	Classification	Historical Classification
Angels	+10	Pitcher's Park	Pitcher's Park [18]
Braves	+ 5	Pitcher's Park	Pitcher's Park [5]
Rockies	− 2	Hitter's Park	Hitter's Park [5]
Yankees	+17	Pitcher's Park	Pitcher's Park [5]
Twins	−16	Hitter's Park	Hitter's Park [19]
Marlins	+13	Pitcher's Park	Pitcher's Park [12]

4 Analysis of Final Model

4.1 Strengths

- Employs an accurate and realistic rating system to answer the question
- Easily adaptable to unforeseen factors
- Assigns quantifiable values to specific venue attributes
- Balances offensive advantages and defensive advantages
- Allows us to create a number of optimal venue designs yet maintain diversity among them

4.2 Weaknesses and Potential Improvements

Though we felt our model provided an excellent approach to the problem, there were some improvements which we would have investigated had we been given more time. Additionally, there were some concepts which we did research but were unable to find data for. Our potential improvements include the following.

- We looked but were unable to find data for wind direction.

- We would have liked to consider all thirty major league ballparks, but that would have required too much time per model.

5 Conclusion

The thirty venues of Major League Baseball encompass a plethora of environments, playing conditions, and design goals. Because of this wide diversity, it is difficult to classify them all in an efficient, quantitative manner. However, we have created a model which does exactly that: it provides a numerical coefficient for each venue. This score conveys how fair or unfair a field is, and whether a field that is not completely fair favors the offense or the defense. Additionally, our model allowed us to draft a memo to Mr. Allan Selig, the Commisioner of Major League Baseball, as item A5 in the appendix.

References

[1] Aber, Susan W. and James S. High Altitude KAP. November 19, 2003. <http://www.geospectra.net/kite/weather/h_altit.htm>

[2] Atlanta Weather History - CityRating.com. November 19, 2003. <http://www.cityrating.com/cityweather.asp?City=Atlanta>

[3] Ballparks of Baseball-The Fields of Major League Baseball. November 19, 2003. <http://www.ballparksofbaseball.com/>

[4] Baseball Field Design and Maintenance. November 19, 2003. <http://www.baseballfielddesign.com/>

[5] Baseball-statistics.com. November 19, 2003. <http://www.baseball-statistics.com/>

[6] Brazeal, David. Are Dynasties Bad for Baseball? - Baseball Primer. November 19, 2003. <http://www.baseballprimer.com/articles/david_brazeal_2002-09-03_0.shtml

[7] Denver Weather History - CityRating.com. November 19, 2003. <http://www.cityrating.com/cityweather.asp?City=Denver>

[8] IRI/LDEO Climate Data Library. November 20, 2003. <http://iridl.ldeo.columbia.edu/>

[9] Los Angeles Relative Humidity - CityRating.com. November 19, 2003. <http://www.cityrating.com/cityhumidity.asp?City=Los+Angeles>

[10] Miami Weather History - CityRating.com. November 19, 2003.
<http://www.cityrating.com/cityweather.asp?city=Miami>

[11] Minneapolis - St. Paul Relative Humidity - CityRating.com. November 19, 2003.
<http://www.cityrating.com/cityhumidity.asp?City=Minneapolis+-+St.+Paul>

[12] Mock, Joe. Baseballparks.com, for fans of baseball parks. November 19, 2003.
<http://baseballparks.com/>

[13] Munsey & Suppes. Ballparks by Munsey and Suppes. November 19, 2003.
<http://www.ballparks.com/baseball/index.htm>

[14] New York Weather History - CityRating.com. November 19, 2003.
<http://www.cityrating.com/cityweather.asp?city=New+York>

[15] Palmer, Chad. USATODAY.com. November 20, 2003.
<http://www.usatoday.com/weather/tg/wbasebal/wbasebal.htm>

[16] Schlatter, Thomas. Golfballs. November 19, 2003.
<http://www.weatherwise.org/qr/qry.golfball.html>

[17] Science at the Stadium. November 19, 2003.
<http://www.turnerlearning.com/efts/bball/science.htm>

[18] Shea, Dennis. Clutch Hits - Baseball Primer. November 20, 2003.
<http://www.baseballprimer.com/clutch/archives/00008790.shtml>

[19] Sheldon, Mark. New Ballpark: Updates and Information. November 19, 2003.
<http://minnesota.twins.mlb.com/NASApp/mlb/min/ballpark/min_ballpark_news.jsp>

[20] The Savvy Traveller - Baseball Parks. November 19, 2003.
<http://www.thesavvytraveller.com/news/winter/north_america/baseball.htm>

[21] Wanner, Noel. Science of Baseball: How Far Can You Hit one?. November 20, 2003.
<http://www.exploratorium.edu/baseball/howfar3.html>

[22] Williams, Jack. USATODAY.com. November 20, 2003.
<http://www.usatoday.com/weather/wdensity.htm