Incorporating the Effects of Designated Hitters in the Pythagorean Expectation

by

Rachel O. Gosch

Department of Mathematics
The University of Texas at Austin

Date: ____________________________

Approved:

Dr. Peter Mueller, Supervisor

Dr. David Rusin

[Department Member #3]

Thesis submitted in partial fulfillment of the requirements for graduation with the
Dean's Scholars Honors Degree in Mathematics,
The University of Texas at Austin
2015

Abstract

The Pythagorean Expectation is widely used in the field of sabermetrics to estimate a baseball team's overall season winning percentage based on the number of runs scored and allowed in its games thus far. Bill James devised the simplest version of the formula through empirical observation as $Winning\ Percentage = \frac{(RS)^2}{(RS)^2+(RA)^2}$ where RS and RA are runs scored and allowed, respectively. Statisticians later found 1.83 to be a more accurate exponent, estimating overall season wins within 3-4 games per season. Steven Miller provided a theoretical justification for the Pythagorean Expectation by modeling runs scored and allowed as independent continuous random variables drawn from Weibull distributions. This paper aims to first explain Miller's methodology using recent data and then build upon Miller's work by incorporating the effects of designated hitters, specifically on the distribution of runs scored by a team. Past studies have attempted to include other effects on run production such as ballpark factor, game state, and pitching power. The results indicate that incorporating information on designated hitters does not improve the error of the Pythagorean Expectation to better than 3-4 games per season.

I grant the Dean's Scholars Program permission to post a copy of my thesis on the University of Texas Digital Repository. For more information, visit http://repositories.lib.utexas.edu/about.

Incorporating the Effects of Designated Hitters in the Pythagorean Expectation

Department of Mathematics

Student

Signature

Date

Supervisor

Signature

Date

Contents

Acknowledgements

I'd like to thank Dr. Peter Mueller for being an awesome, supportive advisor. Many people have asked how I decided to write about sabermetrics: the idea came to me after reading *Moneyball* by Michael Lewis, one of my favorite authors.

1

Background

The field of sabermetrics, baseball statistics, has grown rapidly since its popularization by Bill James in the late 1970's and 1980s. One of James' earliest formulas, the Pythagorean Expectation, is still used today. The Pythagorean Expectation, also known as the Pythagorean Won-Loss Formula, is used to estimate a team's overall winning percentage for a season based on the number of runs scored and allowed in its games. The formula derives its name from its similarity to the well known Pythagorean Theorem, namely $a^2 + b^2 = c^2$. The original Pythagorean Expectation formula is:

$$Winning\ \% = \frac{(RS)^2}{(RS)^2 + (RA)^2} \tag{1.1}$$

where RS is the number of runs scored by a team and RA is the number of runs allowed by a team during the season. Multiplying the winning percentage by the number of games gives an estimate for the number of games a team should win for the season. In practice, the Pythagorean Expectation is often calculated mid-season to predict a team's performance for the remainder of its games.

1.1 Empirical Derivation

Originally, the formula was derived by Bill James from empirical observation of the relation between runs scored and allowed by a team and its season performance. Intuitively, the difference between runs scored and allowed is an indicator of how well a team is playing: Do they often win by just a single run, or are they consistently well outplaying their opponents? As a result, a team's actual winning percentage often converges towards its Pythagorean Expectation. This pattern can be seen with the 2005 Washington Nationals, for example. In early July, the team had a mid-season expected winning percentage of exactly .500 but actually had a record 19 games better. For the remainder of the season, they went 30-49, finishing at exactly .500 (which was actually 4 games better than their final expected winning percentage) (Baseball-Reference.com (2015)). As a result, a comparison of a team's mid-season expected and actual winning percentage can indicate if the team has been somewhat lucky or unlucky thus far.

Shortly after the creation of the Pythaogrean Expectation, James and the statistics community found 1.83 to be a more accurate exponent. Using this value typically estimates a team's winning percentage correctly within 3-4 games and is the value used by Baseball-Reference.com.

1.2 Weibull Distribution

In 2006, Steven Miller supplied a theoretical justification for the Pythagorean Expectation using Weibull distributions to model runs scored and runs allowed as independent variables (Miller (2006)). Usually, runs scored and allowed are fairly low in baseball with occasional higher scoring games. Therefore, Weibull distributions provide a good fit with their typically right skewed shape and in particular, provide

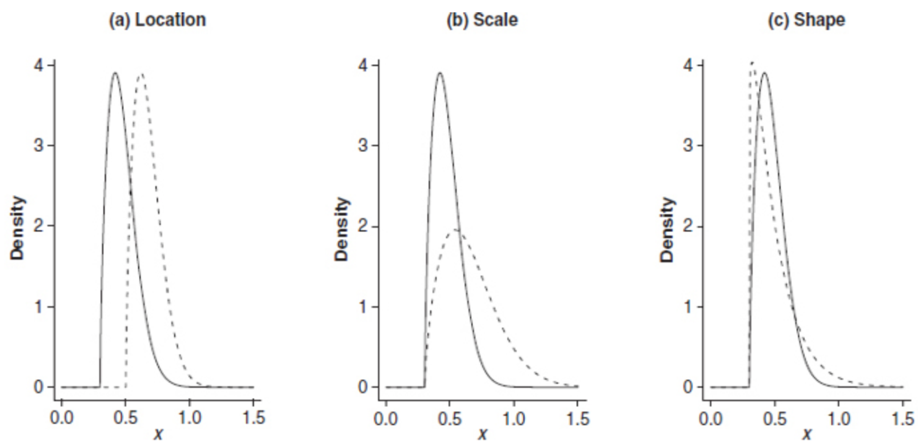

Figure 1.1: Example Weibull Distributions

a better fit than either an exponential or Rayleigh distribution (Miller (2006)). The Weibull probability density function can be written as

$$f(x;\alpha,\beta,\gamma) = \begin{cases} \frac{\gamma}{\alpha}(\frac{x-\beta}{\alpha})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha})^{\gamma}} & x \geqslant \beta \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

where α, β, and γ are given parameters. Figure 1.1 illustrates how changing the values of these parameters affects the overall distribution (Tuerlinckx (2010)). In graph (a), the location parameter β is increased; in graph (b), the scale parameter α is increased; and in graph (c), the shape parameter γ is decreased. Additional properties of Weibull distributions are described in the following lemma (Forbes et al. (2011)).

Lemma 1.1. *The mean and variance of a Weibull with parameters (α, β, γ) are*

$$\mu_{\alpha,\beta,\gamma} = \alpha\Gamma(1+\gamma^{-1}) + \beta \tag{1.3}$$

$$\sigma^2_{\alpha,\beta,\gamma} = \alpha^2\Gamma(1+2\gamma^{-1}) - \alpha^2\Gamma(1+\gamma^{-1})^2 \tag{1.4}$$

Miller's work and resulting adjustment to the Pythagorean Expectation will be further explained in the following section.

1.3 Application to Other Sports

Since its introduction, the Pythagorean Expectation has been applied to other sports including basketball, football, and lacrosse with varying exponents for each sport. For example, the NBA has been found to have an exponent of about 14 while the NFL's is around 2.50 (see Tymins (2014), Rosenfeld et al. (2010)). The size of the exponent, γ, can be interpreted as relating to the role of chance in the game: the greater the value of γ, the greater the probability that the better team will win (and hence the smaller the role of luck). A $\gamma = 0$ implies that both teams have a 50% of winning, regardless of skill level (i.e., dependent entirely on chance), since (1.1) is identically 1/2.

Rosenfeld et al. conduct an interesting study on the difference in the value of γ between regulation and overtime play in football, basketball, and baseball. Due to the brevity of overtime play, the better team's chances of winning are reduced, and luck plays a greater role. Overtime is also subject to additional factors of randomness, such as who starts with the ball (in football). Across all three sports, they find overtime γ values to be smaller than regulation play γ values (Rosenfeld et al. (2010)).

2

Miller's Model

This section fully explains Miller's theoretical approach to the Pythagorean Expectation, referred to as "Miller's Model" in this paper, where runs scored and runs allowed are modelled as independent, continuous random variables from Weibull distributions. Using the notation for Weibull probability distributions in (1.2), the resulting modified Pythagorean Won-Loss Formula is

$$Winning\ \% = \frac{(RS - \beta)^{\gamma}}{(RS - \beta)^{\gamma} + (RA - \beta)^{\gamma}} \tag{2.1}$$

where RS and RA are the *average* runs scored and allowed per game, respectively. Runs scored and runs allowed are assumed to come from two separate distributions with different values of α but sharing the γ and β parameters, with $\beta = -0.5$. This choice for β is explained in the section addressing the continuity of the data.

2.1 Model Assumptions

Miller's Model makes two main assumptions about the nature or runs scored and allowed, namely, the continuity of the data and the independence of runs scored and allowed in a game.

2.1.1 *Continuity of the Data*

Clearly, runs scored and allowed are discrete, rather than continuous, data; however, the assumption of continuous variables simplifies the subsequent calculations by allowing for integration over the distributions. This technique also produces a closed-form result. The choice of $\beta = -0.5$ accounts for the discrete nature of the actual data. When fitting the Weibull distributions, the data is binned into bins of length 1; an obvious choice for these bins is

$$[0, 1), [1, 2), [2, 3), \ldots, [13, 14), [14, \infty) \tag{2.2}$$

Since baseball scores are always integers, the data would be at the left endpoint of each bin, causing the means to be skewed. Miller illustrates this issue with a simple example: suppose a team scores 0 runs in half its games and 1 run in the other half, with the corresponding bins $[0, 1)$ and $[1, 2)$. Finding the best fit constant probability function, for simplicity, yields identically $\frac{1}{2}$ over $[0, 2)$. The mean of this constant approximation is 1:

$$\int_0^2 \frac{1}{2} x \text{ d}x \;=\; \frac{1}{2}(\frac{1}{2}x^2)\bigg|_0^2 \;=\; \frac{1}{4}(4) \;=\; 1$$

However, the observed mean runs per game is $\frac{1}{2} * 0 + \frac{1}{2} * 1 = \frac{1}{2}$. The choice of $\beta = -0.5$ allows the bins to be shifted so that the data is centered in each bin, i.e.,

$$[-0.5, 0.5], [0.5, 1.5], [1.5, 2.5], \ldots, [12.5, 13.5], [13.5, \infty) \tag{2.3}$$

Using these bins, the means align:

$$\int_{-0.5}^{1.5} \frac{1}{2} x \text{ d}x \;=\; \frac{1}{2}(\frac{1}{2}x^2)\bigg|_{-0.5}^{1.5} \;=\; \frac{1}{4}(\frac{9}{4} - \frac{1}{4}) \;=\; \frac{1}{2}$$

Therefore, RS$-\beta$ (RA$-\beta$) is an estimator for the observed average runs scored (allowed) per game if the Weibull modelling runs scored (allowed) has a mean of RS (RA).

2.1.2 Independence of Runs Scored and Allowed

Another assumption under Miller's Model is the independence of runs scored and allowed. Intuitively, it seems as though these two variables would be at least somewhat dependent on each other: a team's offense and defense are comprised of mostly the same players and thus never entirely independent. If a team is good on offense (scoring runs), they are likely good on defense as well (not allowing runs).

The independence of runs scored and allowed is tested by Miller through a modified χ^2 test for an incomplete contingency table. The table is incomplete because a baseball game can never have a tie-score; therefore, the diagonal entries of a table of runs scored and allowed will be zero. Using 2004 American League data, his tests validate the assumption that, "given that runs scored and allowed cannot be equal, the runs scored and allowed per game are statistically independent events" (Miller (2006)).

Ciccolella also explores this issue more creatively (in the same issue of *By the Numbers* that Miller's condensed paper appeared). His results are inconsistent with the independence assumption, finding particular patterns in the relation between runs scored and allowed. For example,"the most common number of runs to score when losing [is] one less than the number allowed for all levels of runs scored" (Ciccolella (2006)).

To further investigate the independence of runs scored and allowed, I use a different technique than either Miller or Ciccolella with 2013 American League data. Although a baseball game cannot ultimately end in a tie, it is possible for the teams to be tied at the end of the 9th inning. When testing for independence, it is unnatural to exclude this possible outcome and restrict runs scored and allowed to being non-equal. Therefore, I look to the box-scores for all extra-inning games and record the tie score at the end of the 9th inning. The resulting distribution of runs scored and allowed can be seen in Table 2.1. Scores exceeding 12 are excluded from the table to avoid excessive zero entries, since baseball scores rarely break 10 runs. Additionally, only home games are included in the analysis to avoid duplicate recordings of runs, since the majority of games are played intraleague: for example, a game where the Boston Red Sox lose at Houston 2-4 would be recorded as a home game with RS = 4, RA = 2 (for Houston) and as an away game with RS = 2, RA = 4 (for Boston).

Runs Allowed

Runs Scored	0	1	2	3	4	5	6	7	8	9	10	11	12
0	4	8	11	12	9	4	4	11	3	3	2	1	1
1	10	13	18	26	8	15	12	12	4	3	4	2	1
2	15	24	22	24	19	20	17	11	6	5	1	1	2
3	15	23	32	29	25	9	16	8	9	3	5	2	2
4	9	16	21	50	17	15	8	7	7	6	4	3	2
5	11	15	16	9	25	9	11	7	6	8	2	3	2
6	10	12	17	16	15	18	10	8	3	1	5	2	2
7	2	7	19	11	13	15	8	6	2	0	2	1	1
8	1	6	9	11	8	5	3	6	3	3	1	1	1
9	2	2	3	2	4	3	4	4	1	0	1	0	0
10	4	2	2	4	3	3	7	1	1	1	0	1	0
11	1	3	2	5	3	1	0	2	2	0	0	1	0
12	2	1	0	0	2	1	2	0	0	1	0	0	0

Table 2.1: 2-way Contingency Table of RS and RA

The counts of runs scored and allowed can then be tested with a χ^2 test for independence. The resulting test has 144 degrees of freedom and gives a χ^2 statistic of 154.94 and a p-value of 0.2521. Therefore, I fail to reject the assumption that runs scored and runs allowed are independent.

2.2 Pythagorean Won-Loss Formula

With the assumptions of the model met, Miller's theoretical derivation of the Pythagorean Won-Loss Formula is as follows.

Theorem 2.1 (Miller's Model). *Let the runs scored and runs allowed per game be two independent random variables drawn from Weibull distributions with parameters* $(\alpha_{RS}, \beta, \gamma)$ *and* $(\alpha_{RA}, \beta, \gamma)$ *respectively, where* α_{RS} *and* α_{RA} *are chosen so that the means are RS and RA. If* $\gamma > 0$*, then*

$$Won - Loss\ Percentage(RS, RA, \beta, \gamma) \;=\; \frac{(RS - \beta)^\gamma}{(RS - \beta)^\gamma + (RA - \beta)^\gamma}.$$

Proof. By (1.3), the means of Weibull distributions with parameters $(\alpha_{RS}, \beta, \gamma)$ and $(\alpha_{RA}, \beta, \gamma)$ are

$$\begin{aligned} RS &= \alpha_{RS}\Gamma(1 + \gamma^{-1}) + \beta \\ RA &= \alpha_{RA}\Gamma(1 + \gamma^{-1}) + \beta, \end{aligned}$$

respectively. These equations can be rewritten as

$$\begin{aligned} \alpha_{RS} &= \frac{RS - \beta}{\Gamma(1 + \gamma^{-1})} \\ \alpha_{RA} &= \frac{RA - \beta}{\Gamma(1 + \gamma^{-1})}. \end{aligned} \tag{2.4}$$

Let X and Y be independent random variables drawn from the Weibull distributions defined above, with X corresponding to runs scored and Y corresponding to runs allowed. A winning percentage can be found by calculating the probability that X is greater than Y. Recall the PDF of a Weibull distribution from (1.2) for the integrations below.

$$P(X > Y) \;=\; \int_{x=\beta}^{\infty} \int_{y=\beta}^{x} f(x; \alpha_{RS}, \beta, \gamma) f(y; \alpha_{RA}, \beta, \gamma) \; \mathrm{d}y \; \mathrm{d}x$$

$$= \int_{x=\beta}^{\infty}\int_{y=\beta}^{x} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha_{RS}})^{\gamma}} \frac{\gamma}{\alpha_{RA}}(\frac{y-\beta}{\alpha_{RA}})^{\gamma-1}e^{-(\frac{y-\beta}{\alpha_{RA}})^{\gamma}} \mathrm{d}y\ \mathrm{d}x$$

$$= \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha_{RS}})^{\gamma}} \left[\int_{y=\beta}^{x} \frac{\gamma}{\alpha_{RA}}(\frac{y-\beta}{\alpha_{RA}})^{\gamma-1}e^{-(\frac{y-\beta}{\alpha_{RA}})^{\gamma}} \mathrm{d}y \right] \mathrm{d}x$$

$$= \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha_{RS}})^{\gamma}} \left[-e^{-(\frac{y-\beta}{\alpha_{RA}})^{\gamma}} \Big|_{\beta}^{x} \right] \mathrm{d}x$$

$$= \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha_{RS}})^{\gamma}} \left[1 - e^{-(\frac{x-\beta}{\alpha_{RA}})^{\gamma}} \right] \mathrm{d}x$$

$$= \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha_{RS}})^{\gamma}} \mathrm{d}x - \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha_{RS}})^{\gamma}} e^{-(\frac{x-\beta}{\alpha_{RA}})^{\gamma}} \mathrm{d}x$$

The first integral in this expression reduces to 1, since the integral over the support of a probability density is 1. Therefore,

$$= 1 - \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha_{RS}})^{\gamma}} e^{-(\frac{x-\beta}{\alpha_{RA}})^{\gamma}} \mathrm{d}x$$

$$= 1 - \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha_{RS}} + \frac{x-\beta}{\alpha_{RA}})^{\gamma}} \mathrm{d}x$$

Let $\frac{1}{\alpha^{\gamma}} = \frac{1}{\alpha_{RS}^{\gamma}} = \frac{1}{\alpha_{RA}^{\gamma}} = \frac{\alpha_{RS}^{\gamma}+\alpha_{RA}^{\gamma}}{\alpha_{RS}^{\gamma}\alpha_{RA}^{\gamma}}$. Then,

$$= 1 - \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha_{RS}}(\frac{x-\beta}{\alpha_{RS}})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha})^{\gamma}} \mathrm{d}x$$

$$= 1 - \frac{\alpha^{\gamma}}{\alpha_{RS}^{\gamma}} \int_{x=\beta}^{\infty} \frac{\gamma}{\alpha}(\frac{x-\beta}{\alpha})^{\gamma-1}e^{-(\frac{x-\beta}{\alpha})^{\gamma}} \mathrm{d}x$$

Again, this integral reduces to 1. So,

$$= 1 - \frac{\alpha^{\gamma}}{\alpha_{RS}^{\gamma}}$$

$$= 1 - \frac{1}{\alpha_{RS}^{\gamma}} \frac{\alpha_{RS}^{\gamma}\alpha_{RA}^{\gamma}}{\alpha_{RS}^{\gamma}+\alpha_{RA}^{\gamma}}$$

$$= 1 - \frac{\alpha_{RA}^{\gamma}}{\alpha_{RS}^{\gamma} + \alpha_{RA}^{\gamma}}$$

$$= \frac{\alpha_{RS}^{\gamma}}{\alpha_{RS}^{\gamma} + \alpha_{RA}^{\gamma}} \tag{2.5}$$

Using (2.4) and substituting into (2.5) results in

$$P(X > Y) = \frac{(RS - \beta)^{\gamma}}{(RS - \beta)^{\gamma} + (RA - \beta)^{\gamma}}. \tag{2.6}$$

$\square$

Since X is the random variable representing runs scored and Y the random variable representing runs allowed, (2.6) is the probability of winning a game, and can be extended to estimate overall winning percentage for a season.

3

Analysis

This section explains the results of applying Miller's methodology to 2013 American League data (Miller's original analysis was conducted on 2004 American League data). These results are used as a baseline for comparison to an original model incorporating the effects of designated hitters, introduced later in this section. All of the data used in this section was analyzed in R Statistical Software and comes from the historical stats section of www.baseball-reference.com. Furthermore, all team data is restricted to the 162 regular-season games; data from postseason games is eliminated.

3.1 Multiplicity Adjustments

Statistical hypothesis testing is based on controlling the probability of Type I errors at some pre-determined level, α, such that $P(Type\ I\ error) \leqslant \alpha$. A customary choice is $\alpha = 0.05$. Recall that a Type I error is falsely rejecting a true null hypothesis. Multiplicity adjustments are a statistical adjustments used when multiple hypothesis tests are conducted simultaneously to lower the probability of making Type I errors. For example, when testing a hypothesis for each of the 15 teams in the American

League, the probability of falsely rejecting one of the 15 null hypotheses is greatly inflated. This problem can be avoided by lowering the significance level for each of the multiple comparisons. In effect, the adjustment lowers the per comparison significance level to maintain a bound on the probability of falsely rejecting any of the multiple true hypotheses. This adjustment is necessary for the analysis later in this section, so two common methods of adjustment, Bonferroni and Holm-Bonferroni, are first described briefly below.

3.1.1 Bonferroni

The Bonferroni procedure is one of the simplest methods of multiplicity adjustment. Its name stems from the use of the Bonferroni inequality.

Lemma 3.1 (Bonferroni Inequality)**.** *Let A_i, $i = 1$ to k, represent k events. Then,*

$$P\big(\bigcap_{i=1}^{k} A_i\big) \geqslant 1 - \sum_{i=1}^{k} P\big(\bar{A}_i\big),$$

where $\bar{A}_i$ is the complement of the event A_i.

Proof. Let A_i, $i = 1$ to k, represent k events. Then,

$$P(\bigcap_{i=1}^{k} A_i) = 1 - P(\bigcup_{i=1}^{k} \bar{A}_i) \geqslant 1 - \sum_{i=1}^{k} P(\bar{A}_i)$$

□

This inequality can be rewritten as

$$1 - P\big(\bigcap_{i=1}^{k} A_i\big) \leqslant \sum_{i=1}^{k} P\big(\bar{A}_i\big). \tag{3.1}$$

In the context of statistical testing, let $\bar{A}_i$ be a Type I error in the ith test of k hypothesis tests. Recall that a Type I error is the probability of falsely rejecting the null hypothesis. The probability that no Type I errors occur across the k tests is $P\big(\bigcap_{i=1}^{k} A_i\big)$, and $1 - P\big(\bigcap_{i=1}^{k} A_i\big)$ is the probability of at least one Type I error occurring. Since α is typically used to denote the level of tolerance for Type I errors in statistical tests, let $\alpha_i = P(\bar{A}_i)$. Therefore, by Lemma 3.1, the probability of at least one Type I error in k hypothesis tests is $\leqslant \sum_{i=1}^{k} \alpha_i$.

Assuming the same α level across all k tests, this result implies an α level k times larger than intended. For example, suppose 10 hypothesis tests are being conducted, each with $\alpha = 0.05$. Without multiplicity adjustment, the implied upper-bound for probability of Type I error is 0.50! Bonferroni accounts for this issue by dividing the α level by the number of tests before assessing the significance of p-values (or equivalently, multiplying the p-values by the number of tests being run and using the original α).

3.1.2 Holm-Bonferroni

The Bonferroni procedure is often criticized as being overly conservative in its adjustments to avoid Type I errors. A similar procedure, the Holm-Bonferroni, adjusts p-values in a "sequentially rejective" method (Holm (1979)). Compared to the classical Bonferroni, the probability of rejecting any set of false hypotheses is greater or equal with the Holm-Bonferroni.

The sequentially rejective procedure is as follows: supppose n hypothesis tests are being conducted. The n p-values are ordered such that $p^{(1)} \leqslant p^{(2)} \leqslant \ldots \leqslant p^{(n)}$, and let $H^{(1)}$ be the hypothesis test corresponding to $p^{(1)}$. Recall that the classical procedure compares all n p-values to the adjusted significance level $\frac{\alpha}{n}$. In the Holm method, $p^{(1)}$ is compared to $\frac{\alpha}{n}$; if the p-value is not significant, i.e., $p^{(1)} > \frac{\alpha}{n}$, then the procedure stops and all null hypotheses are accepted (fail to be rejected). However, if $p^{(1)} \leqslant \frac{\alpha}{n}$, then $H^{(1)}$ is rejected, and $p^{(2)}$ is compared to $\frac{\alpha}{n-1}$. The process continues with each rejected hypothesis until the first failure to reject, at which point all remaining hypotheses are also accepted. The last p-value, $p^{(n)}$, is compared to α (if necessary). Similar to the classic Bonferroni, an equivalent technique involves multiplying the p-values by progressively smaller integers, i.e. $n, n-1, n-2, \ldots, 1$ and comparing to the original α.

3.2 Miller's Model 2013 Results

Before applying Miller's Model to 2013 season data, I first validated my program by using 2004 American League data and comparing the results to those in Miller's paper. After confirming the accuracy of the code, I analyzed the 2013 data.

The first step requires creating the 15 bins in (2.3) and binning the runs scored and allowed in each game. Next, the number of occurrences in each bin is tabulated, e.g., a team scored 0 runs 15 times, 1 run 19 times, 2 runs 17 times, etc. Using this information, the best fit Weibulls are then fitted to each team for runs scored and runs allowed by minimizing the squared difference between fitted and empirical frequencies. Figure 3.1 shows the fitted Weibull distributions for the New York Yankees. Instances of 14 or more runs (scored or allowed) are grouped together in the last bin of the histogram, $[13.5, \infty)$. The fitted distributions for every American League team can be found in Appendix A.

With the γ exponent value from each team's fitted Weibull distributions, the Pythagorean Expectation can be computed for each team. Table 3.1 illustrates these results. For each team, the Pythagorean Expectation is computed mid-season (using data from the first 81 games) for a prediction of the team's final winning percentage, as well as at the end of the season (using data from all 162 games) to evaluate the accuracy of

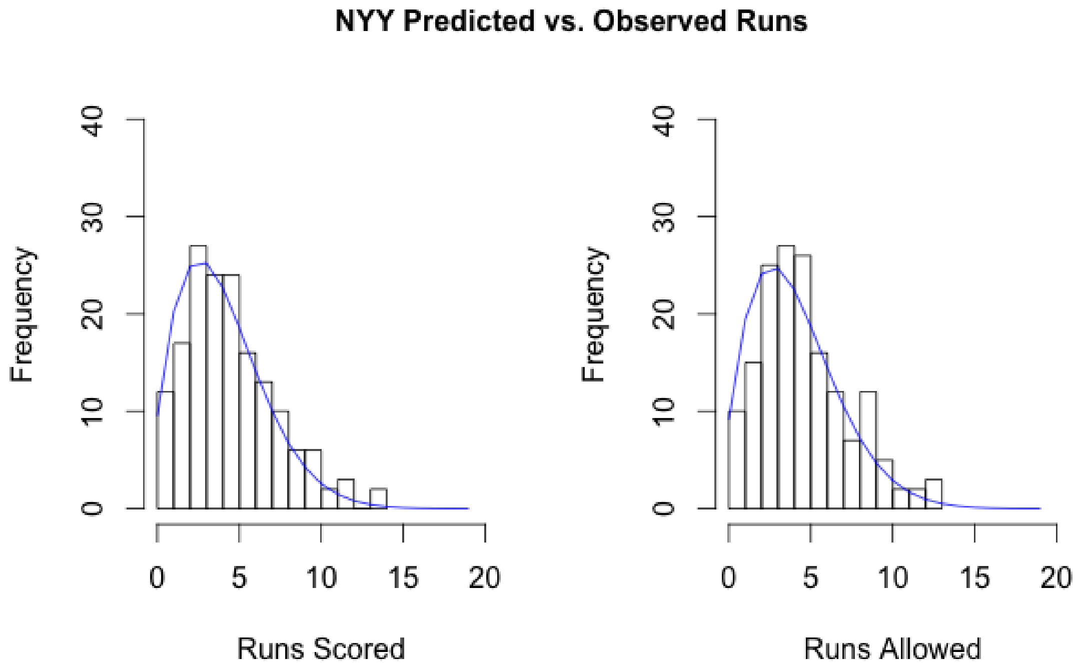

FIGURE 3.1: Fitted Weibull Distributions for the New York Yankees

the model. The difference in the observed and final estimate of the number of games won is recorded under "Error" (these numbers are found by multiplying the percentages in the 4th and 5th columns by 162). As previously mentioned, the Pythagorean Expectation consistently has an error of about 3-4 games per season; the 2013 data follow this pattern with a mean absolute error of 3.12. In terms of predictive value, the mean absolute difference in mid-season predicted number of wins and final wins is 5.21. Additionally, the teams have an average γ of 1.73 with a standard error of 0.13. Although this value is slightly less than the established best value of 1.83, it is still within 1 standard error.

Team	Obs. Mid-Season Win %	Mid-Season Pred. of Final Win %	Obs. Final Win %	Final Estimate of Win %	Error
Baltimore Orioles	55.56%	52.16%	52.47%	52.28%	0.31
Boston Red Sox	59.26%	58.49%	59.88%	60.28%	-0.65
Chicago White Sox	40.74%	43.60%	38.89%	42.17%	-5.31
Cleveland Indians	53.09%	52.24%	56.79%	54.15%	4.28
Detroit Tigers	53.09%	56.63%	57.41%	58.86%	-2.35
Houston Astros	37.04%	38.78%	31.48%	37.90%	-10.40
Kansas City Royals	48.15%	51.08%	53.09%	52.89%	0.32
Los Angeles Angels	46.91%	50.41%	48.15%	49.79%	-2.66
Minnesota Twins	44.44%	45.65%	40.74%	40.19%	0.89
New York Yankees	51.85%	48.06%	52.47%	48.77%	6.00
Oakland Athletics	58.02%	55.98%	59.26%	57.23%	3.29
Seattle Mariners	43.21%	42.92%	43.83%	43.30%	0.85
Tampa Bay Rays	51.85%	51.59%	56.17%	52.76%	5.54
Texas Rangers	58.02%	53.62%	56.17%	55.43%	1.21
Toronto Blue Jays	49.38%	50.00%	45.68%	47.41%	-2.81

Table 3.1: Pythagorean Expectations vs. Observed Winning Percentages at Mid-Season and End-of-Season

After finding the distributions, the fit is tested with a χ^2 goodness of fit test for each team. The χ^2 statistic is defined as

$$\chi^2 \;=\; \sum_{i=0}^{14} \frac{(RS_{obs}(k) - Games * A_{RS}(k))^2}{Games * A_{RS}(k)} + \sum_{i=0}^{14} \frac{(RA_{obs}(k) - Games * A_{RA}(k))^2}{Games * A_{RA}(k)}$$

where $RS_{obs}(k)$ is the occurrences in the kth bin, i.e., the number of games in which k runs were scored (with the exception of $k = 14$, where the bin is the number of games in which at least 14 runs were scored), $Games$ is the number of games played, and $A_{RS}(k)$ is the area under the Weibull distribution for runs scored corresponding to the kth bin (Miller (2006)). Each test has 24 degrees of freedom. Since 15 tests are being run simultaneously, a multiplicity adjustment must be used on the resulting p-values. Both the Bonferroni and Holm-Bonferroni techniques are included in Table 3.2. From these results, the Weibull distributions appear to be a good fit for all but the Baltimore Orioles and Detroit Tigers, both of which are significant at $\alpha = 0.05$.

Team	χ^2	P-value	Bonferroni	Holm-Bonferroni
Baltimore Orioles	64.15	<0.0001	0.0002	0.0002
Boston Red Sox	22.44	0.5532	1.0	1.0
Chicago White Sox	27.96	0.2617	1.0	1.0
Cleveland Indians	26.00	0.3533	1.0	1.0
Detroit Tigers	51.94	0.0008	0.0119	0.0112
Houston Astros	24.89	0.4119	1.0	1.0
Kansas City Royals	27.11	0.2992	1.0	1.0
Los Angeles Angels	20.27	0.6812	1.0	1.0
Minnesota Twins	24.70	0.4224	1.0	1.0
New York Yankees	21.88	0.5866	1.0	1.0
Oakland Athletics	14.91	0.9234	1.0	1.0
Seattle Mariners	25.89	0.3589	1.0	1.0
Tampa Bay Rays	42.49	0.0114	0.1704	0.1476
Texas Rangers	24.77	0.4184	1.0	1.0
Toronto Blue Jays	27.19	0.2958	1.0	1.0

Table 3.2: Results of χ^2 Goodness of Fit Tests

3.3 Designated Hitter Model

Two leagues exist in Major League Baseball (MLB), the American League (AL) and the National League (NL). Each league consists of 15 teams, divided equally into three divisions. Both leagues follow generally the same rules, with a major exception being the position of Designated Hitter (DH). In the AL, designated hitters are players who bat in place of their team's pitcher; the NL has no such position, and pitchers bat for themselves. When interleague games are played, the rules of the home team are respected with regards to designated hitters: if the game is played in an AL park, a DH may be used. This difference in a team's offense could potentially affect its run production and thus alter the probability distribution describing its runs scored. In particular, for 2013, the average runs scored in games with a DH is 4.36 while the average in games without a DH is 3.91. Therefore, I attempt to improve upon Miller's Model by accounting for the use of designated hitters.

The use or non-use of designated hitters is likely only to affect the runs scored by a team, since DH is an offensive position. Therefore, my model assumes runs scored in games with a DH and runs scored in games without a DH are drawn from two separate Weibull distributions. Teams in the AL play 142 intraleague games and 20 interleague games each season; interleague games are equally divided between home and away parks. Therefore, each team plays 152 games with a DH and 10 games without.

Theorem 3.2 (Designated Hitter Model)**.** *Let the runs scored using a DH, runs scored without a DH, and runs allowed per game be three independent random variables drawn from Weibull distributions with parameters* $(\alpha_{RS_{DH}}, \beta, \gamma)$, $(\alpha_{RS_{NDH}}, \beta, \gamma)$, *and* $(\alpha_{RA}, \beta, \gamma)$ *respectively, where* $\alpha_{RS_{DH}}$, $\alpha_{RS_{NDH}}$, *and* α_{RA} *are chosen so that the means are* RS_{DH}, RS_{NDH}, *and* RA. *If* $\gamma > 0$, *then*

$$\begin{aligned} Winning\ \% \ = \ & \left(\frac{152}{162}\right)\frac{(RS_{DH}-\beta)^{\gamma}}{(RS_{DH}-\beta)^{\gamma}+(RA-\beta)^{\gamma}} \quad + \\ & \left(\frac{10}{162}\right)\frac{(RS_{NDH}-\beta)^{\gamma}}{(RS_{NDH}-\beta)^{\gamma}+(RA-\beta)^{\gamma}} \end{aligned} \tag{3.2}$$

3.3.1 2013 Results

Using this method, Weibull distributions are fitted again for each team. Figure 3.2 shows the new fitted Weibull distributions for the New York Yankees; the fitted distribution for all teams can be found in Appendix A. These distributions also result in an average γ value of 1.73 with a standard error of 0.12. The mean absolute error for number of games won is 3.19, almost identical to that of Miller's Model; Figure 3.3 shows the end-of-season estimated number of wins for each model against actual wins. Clearly, the two models yield highly similar estimations. Furthermore, the mean absolute error for mid-season predictions of games won is 5.39, also very similar to the previous model; Figure 3.4 illustrates the mid-season predicted winning percentage for each model and the observed winning percentage.

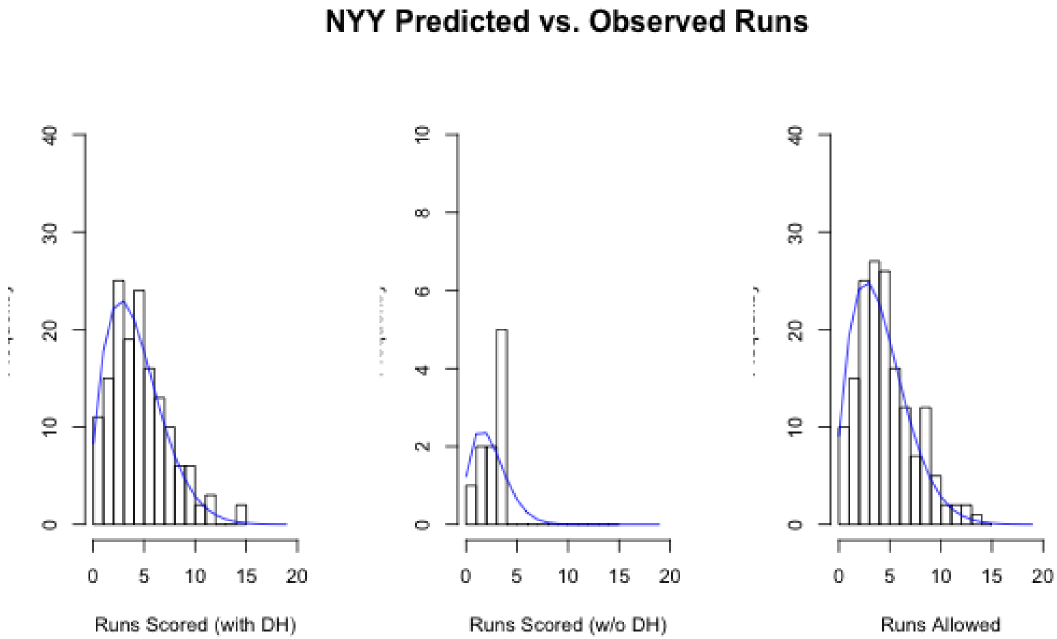

FIGURE 3.2: Fitted Weibull Distributions for the New York Yankees, Incorporating Designated Hitters

Running χ^2 goodness of fit tests for the DH models, with 37 degrees of freedom, shows that the Baltimore Orioles, Cleveland Indians, Detroit Tigers, and Kansas City Royals do not appear to have a good fit and can be rejected at $\alpha = 0.05$ (see Table 3.3) In particular, the Indians and Tigers have especially large χ^2 values. These poor fits likely stem from the fact that the Weibull distributions for runs scored without DH are fit based on only 10 observations per team. Overall, these results combined with the mean absolute errors indicate that the new model incorporating the effects of designated hitters does not offer much improvement in terms of fit or predictive value.

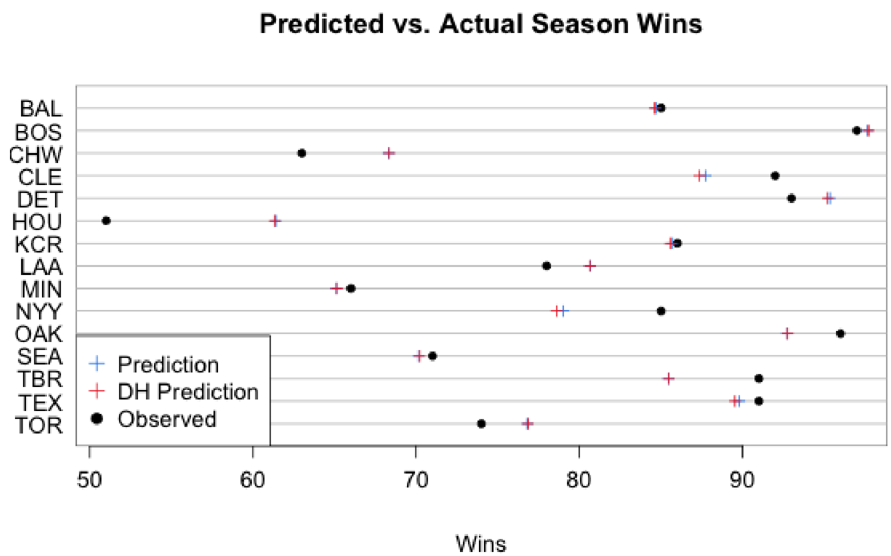

FIGURE 3.3: End-of-Season Pythagorean Expectations vs. Observed End-of-Season Wins

3.3.2 *Divisional Analysis*

Since the lack of improvement in the DH model is likely due to the fact that the Weibull distributions are fit with only 10 observations per team, a larger sample size is preferred. Therefore, I apply the models on a divisional basis: using the 3 divisions of the AL (West, East, and Central), the total number of games increases to 810, 760 with DH and 50 without. Although calculating Pythagorean Expectations for divisions may not be as practical as for individual teams, this analysis can still be insightful as to whether the lack of improvement from incorporating DH is due solely to the inadequate sample size. Since the proportion of games played with and without DH remains the same at the team and division level, (3.2) can still be used for the Pythagorean Expectations.

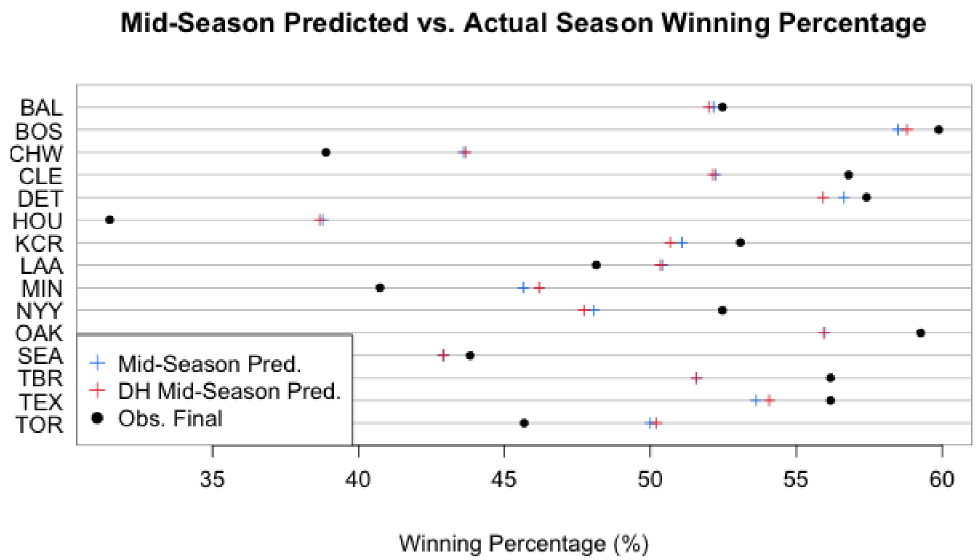

FIGURE 3.4: Mid-Season Pythagorean Expectation Predictions vs. Observed End-of-Season Winning Percentages

Team	χ^2	P-value	Bonferroni	Holm-Bonferroni
Baltimore Orioles	72.75	0.0004	0.0061	0.0053
Boston Red Sox	32.22	0.6924	1.0	1.0
Chicago White Sox	30.79	0.7544	1.0	1.0
Cleveland Indians	329.00	<0.0001	<0.0001	<0.0001
Detroit Tigers	241.37	<0.0001	<0.0001	<0.0001
Houston Astros	46.46	0.1370	1.0	1.0
Kansas City Royals	72.57	0.0004	0.0064	0.0053
Los Angeles Angels	34.45	0.5891	1.0	1.0
Minnesota Twins	30.39	0.7706	1.0	1.0
New York Yankees	28.06	0.8551	1.0	1.0
Oakland Athletics	29.55	0.8032	1.0	1.0
Seattle Mariners	32.72	0.6701	1.0	1.0
Tampa Bay Rays	48.42	0.0991	1.0	0.9909
Texas Rangers	32.82	0.6652	1.0	1.0
Toronto Blue Jays	50.90	0.0637	0.9559	0.7010

Table 3.3: Results of χ^2 Goodness of Fit Tests for DH Model

The fitted Weibull distributions for the 3 divisions using each model are shown in Appendix B. Using Miller's Model, the mean absolute error for number of games won is 6.17, and the mean absolute error for mid-season predictions is 6.61. The same values for the DH model are 6.20 and 7.43, respectively. Again, the DH model shows little improvement over the original model in terms of fit or predictive value. With χ^2 goodness of fit tests, the AL Central and East are rejected with Miller's Model at $\alpha = 0.05$, using the less conservative Holm-Bonferroni values (see Table 3.4). Additionally, the AL Central and East can be rejected at $\alpha = 0.05$ significance for the DH model, providing evidence that the increased sample size does not necessarily lead to better fitting distributions under this model.

Division	χ^2	P-value	Bonferroni	Holm-Bonferroni
Miller's Model				
AL Central	42.82	0.0104	0.0313	0.0313
AL East	39.72	0.0229	0.0688	0.0458
AL West	22.05	0.5766	1.0	0.5766
DH Model				
AL Central	85.10	<0.0001	<0.0001	<0.0001
AL East	66.36	0.0021	0.0064	0.0043
AL West	32.28	0.6990	1.0	0.6898

Table 3.4: Results of χ^2 Goodness of Fit Tests for Divisional Models

3.4 Conclusion

Overall, the models incorporating designated hitters did not improve upon the existing Pythagorean Expectation. Both the DH model and Miller's Model estimate season wins within about 3 games per season at the team level, and within around 6 games at the divisional level. The lack of improvement likely stems from the fact that American League teams play so few games each season without a DH, resulting in a small sample size with which to find the optimal Weibull distribution. These results indicate that the difference in run production between games with and without designated hitters cannot be accounted for in the construction of the Weibull distributions. However, future work could attempt to find a more nuanced way to incorporate the effects of designated hitters. For now, it appears as though James' simpler formula for Pythagorean Expectation from the 1980's still provides the best estimator of a team's winning percentage.

3.4.1 Predicting 2015 UT Baseball

As a fun exercise, I now apply the Pythagorean Expectation to UT baseball to estimate the team's performance for the remainder of the 2015 season. For these estimations, the classic Pythagorean Expectation formula is used with the average gamma found for the 2013 American League data, 1.73.

To explore whether the average gamma found for professional baseball is appropriate for college, the mid-season and end-of-season estimations are calculated for the 2014 UT season and compared to the observed results. As Table 3.5 illustrates, using $\gamma = 1.73$ yields nearly highly accurate estimations for the 2014 season; the mid-season prediction is only 3% off from the final winning percentage, and UT's record declined corresponding with this mid-season estimation from 78.79% to a final 68.66%. Therefore, predictions for the 2015 should be fairly accurate as well. Based on data from the first 32 games of the 2015 season, the UT team currently has a winning record of 53.15% and a predicted record of 62.11% for the season. This estimation implies that UT is currently under-performing expectations and should win a large proportion of its remaining games to approach the Pythagorean Expectation.

Year	Obs. Mid-Season Win %	Mid-Season Pred. of Final Win %	Obs. Final Win %	Final Estimate of Win %
2014	78.79%	71.53%	68.66%	67.05%
2015	53.13%	62.11%	?	?

Table 3.5: Pythagorean Expectations and Predictions for UT Baseball

Appendix A

Fitted Weibull Distributions: Team Analysis

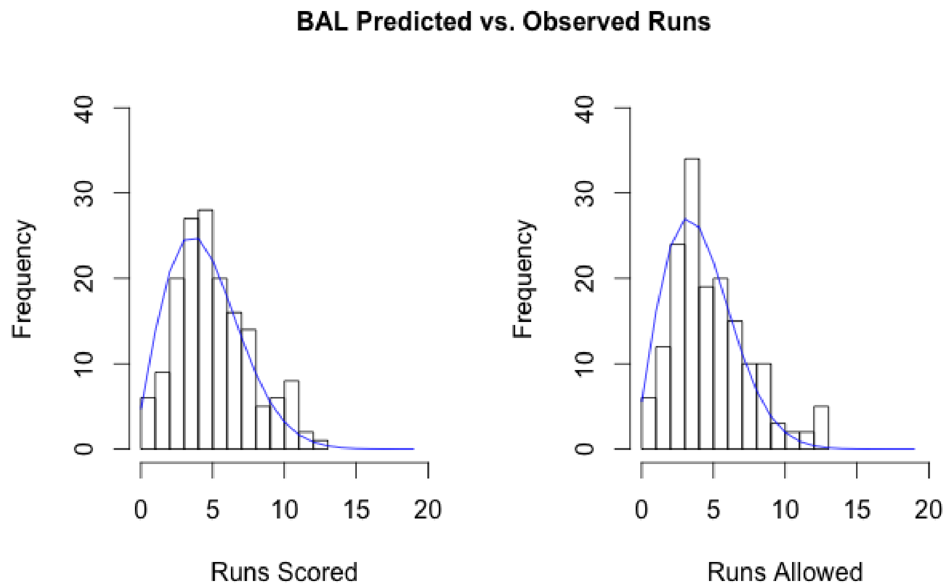

FIGURE A.1: Fitted Weibull Distributions for the Baltimore Orioles

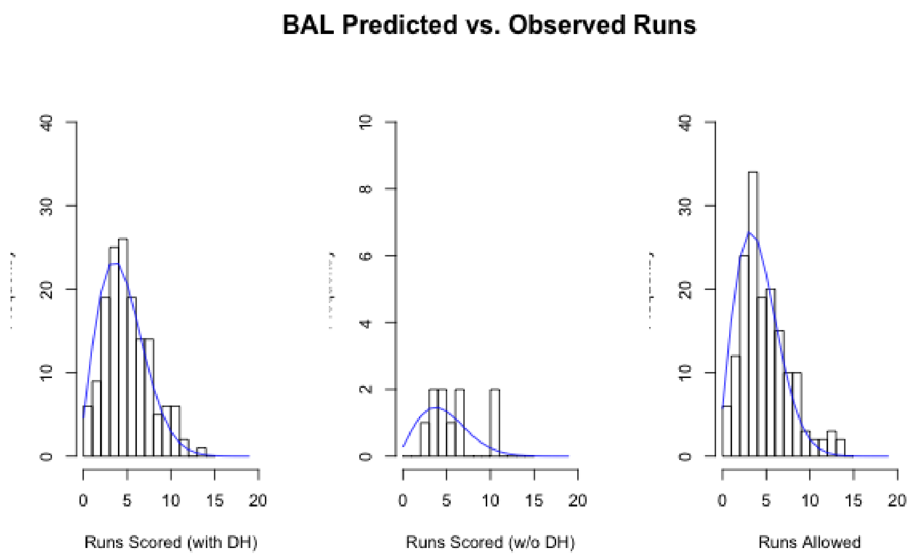

FIGURE A.2: Fitted Weibull Distributions for the Baltimore Orioles, Incorporating Designated Hitters

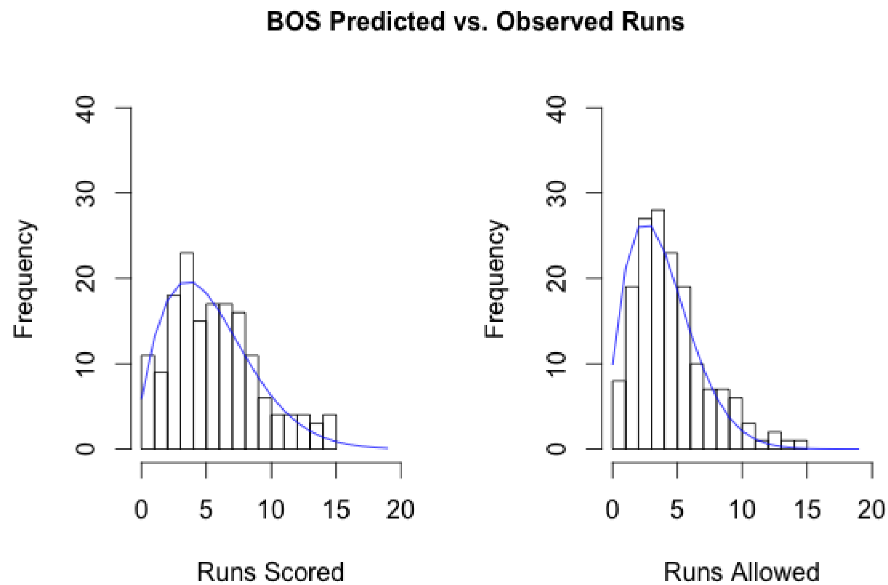

FIGURE A.3: Fitted Weibull Distributions for the Boston Red Sox

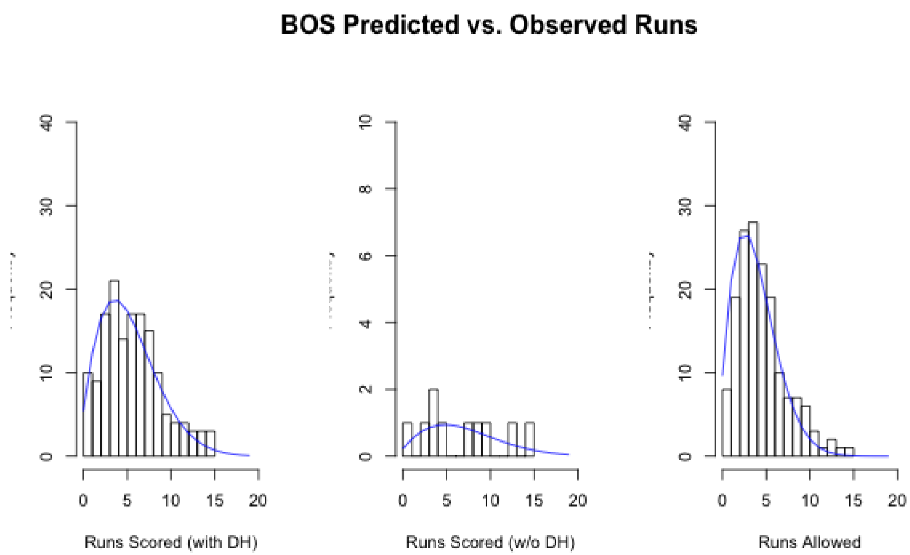

FIGURE A.4: Fitted Weibull Distributions for the Boston Red Sox, Incorporating Designated Hitters

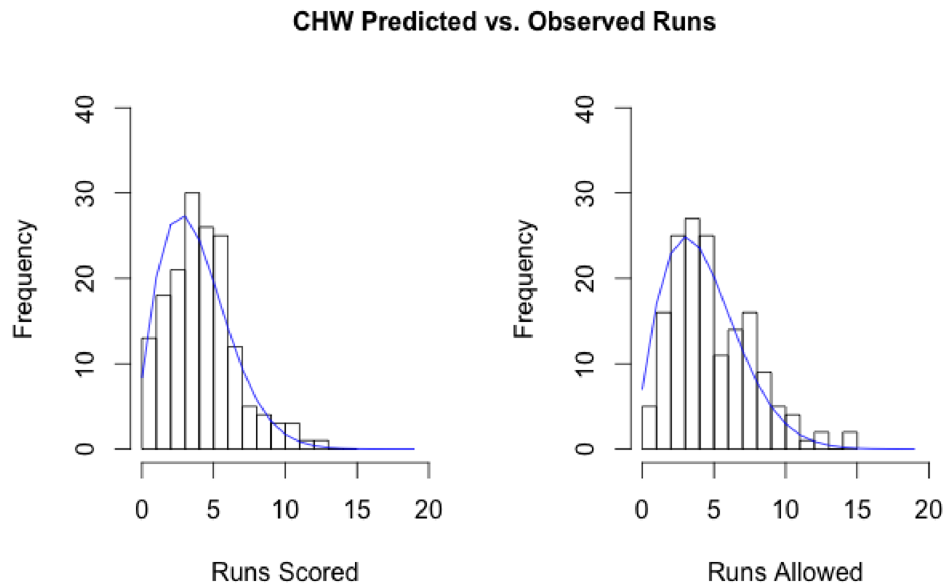

FIGURE A.5: Fitted Weibull Distributions for the Chicago White Sox

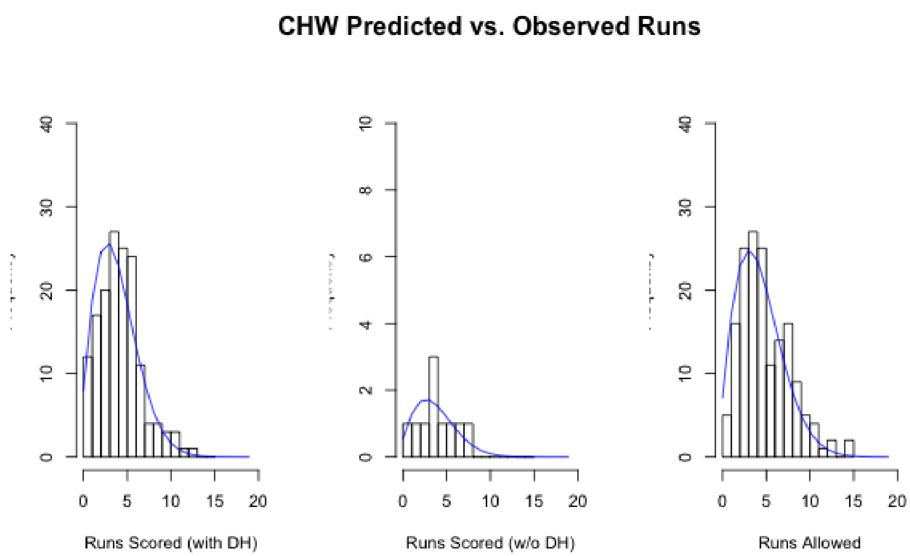

FIGURE A.6: Fitted Weibull Distributions for the Chicago White Sox, Incorporating Designated Hitters

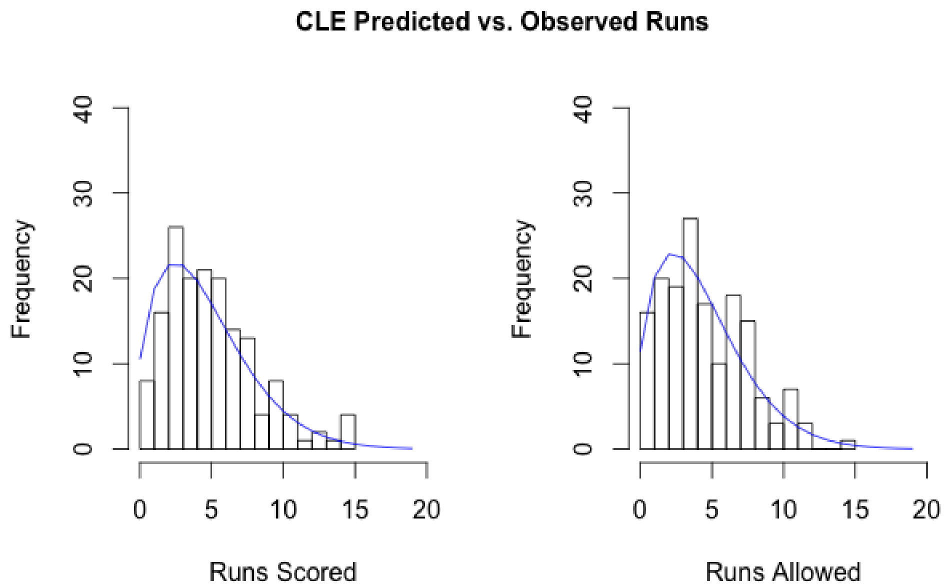

Figure A.7: Fitted Weibull Distributions for the Cleveland Indians

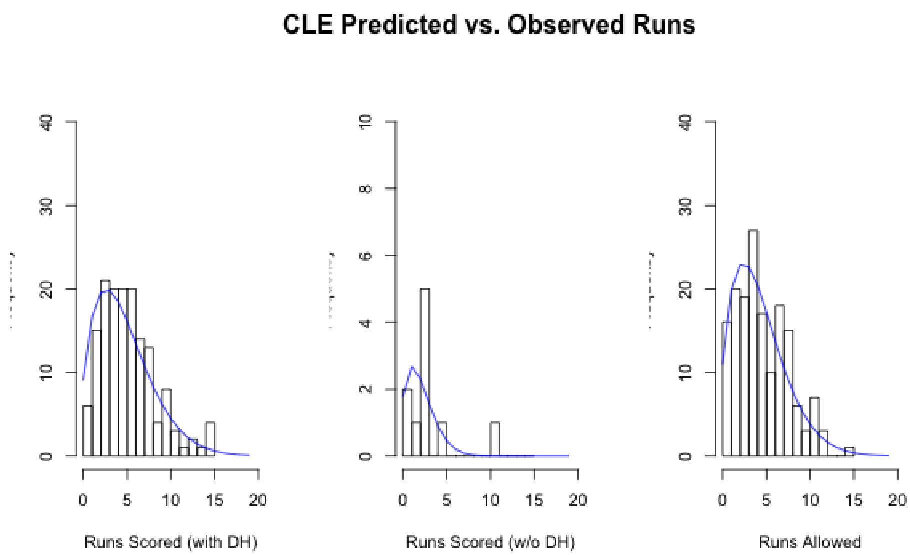

Figure A.8: Fitted Weibull Distributions for the Cleveland Indians, Incorporating Designated Hitters

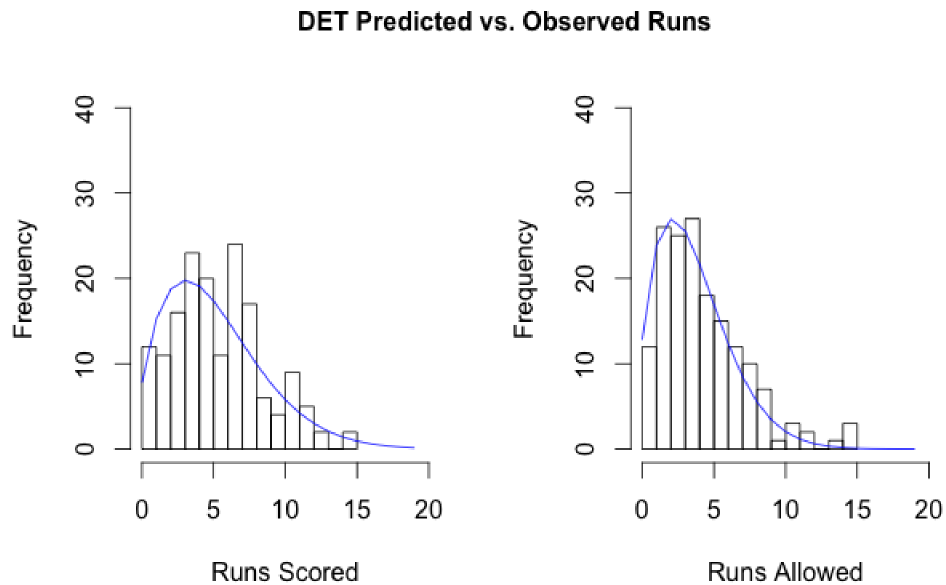

FIGURE A.9: Fitted Weibull Distributions for the Detroit Tigers

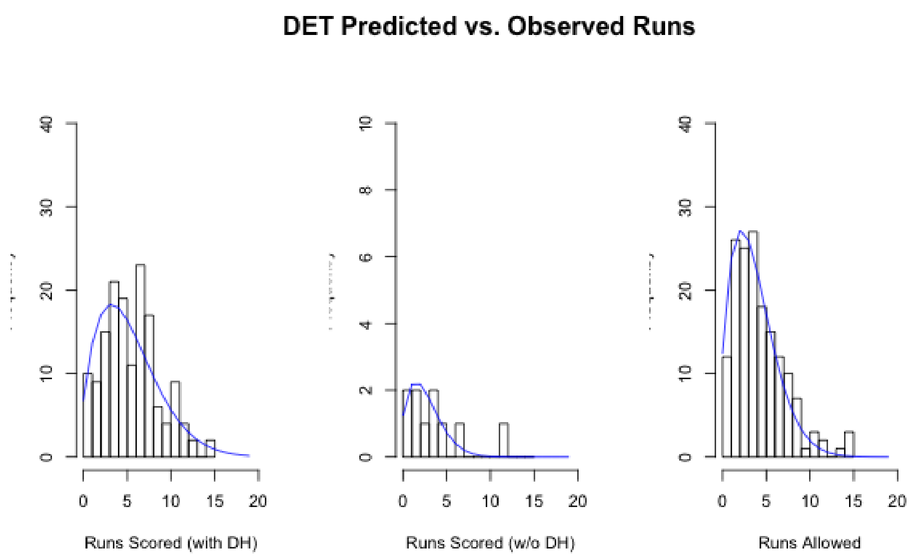

FIGURE A.10: Fitted Weibull Distributions for the Detroit Tigers, Incorporating Designated Hitters

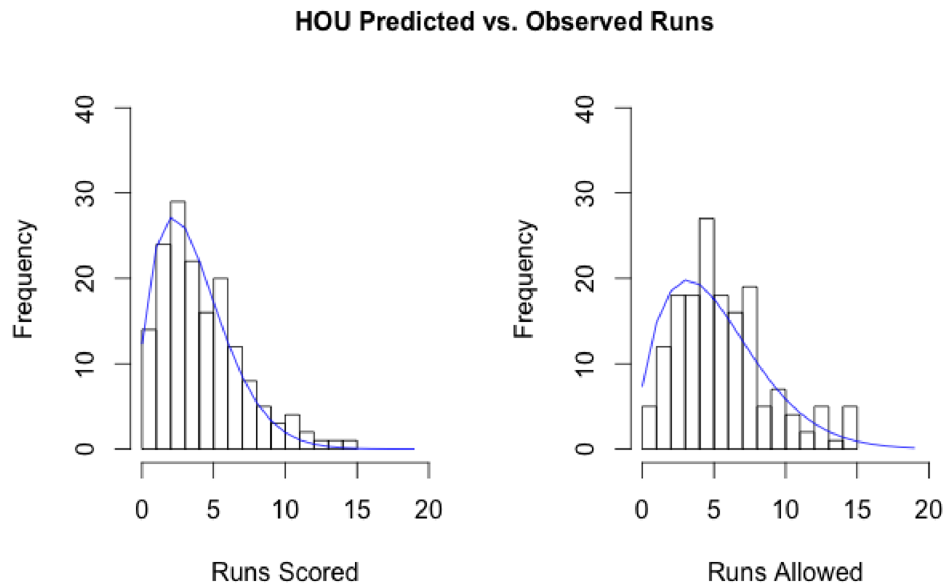

Figure A.11: Fitted Weibull Distributions for the Houston Astros

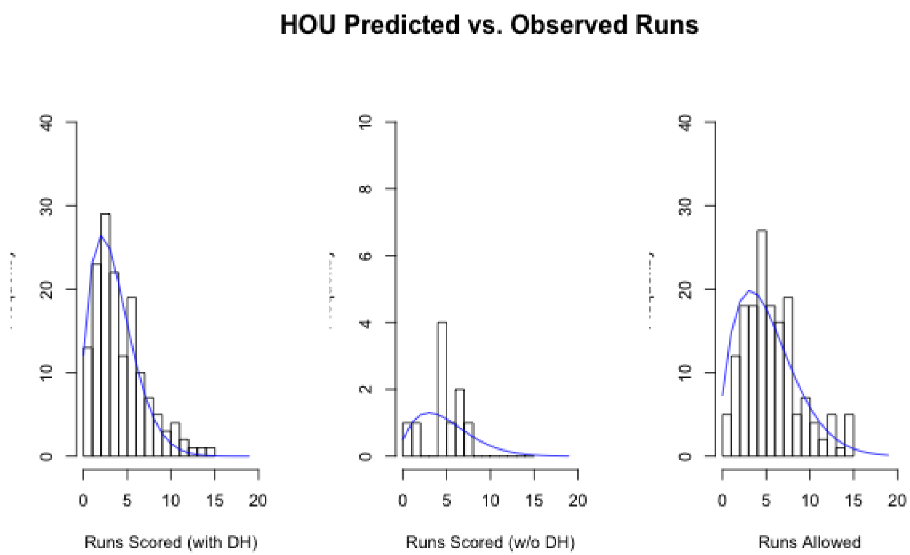

Figure A.12: Fitted Weibull Distributions for the Houston Astros, Incorporating Designated Hitters

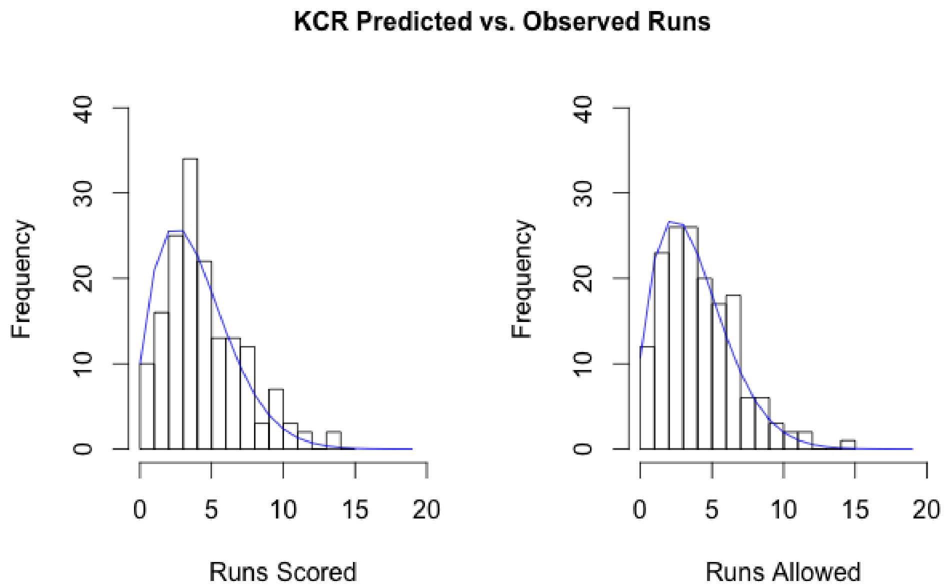

Figure A.13: Fitted Weibull Distributions for the Kansas City Royals

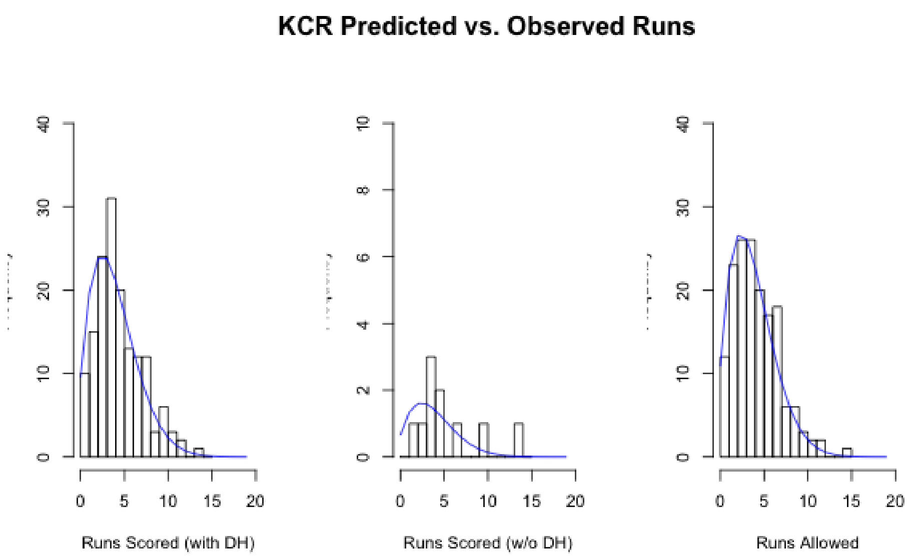

Figure A.14: Fitted Weibull Distributions for the Kansas City Royals, Incorporating Designated Hitters

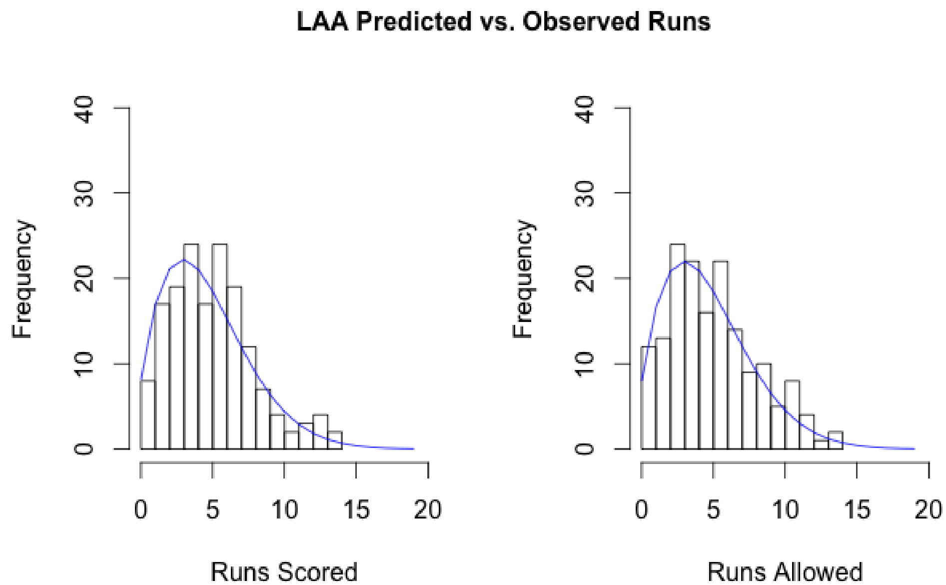

FIGURE A.15: Fitted Weibull Distributions for the Los Angeles Angels

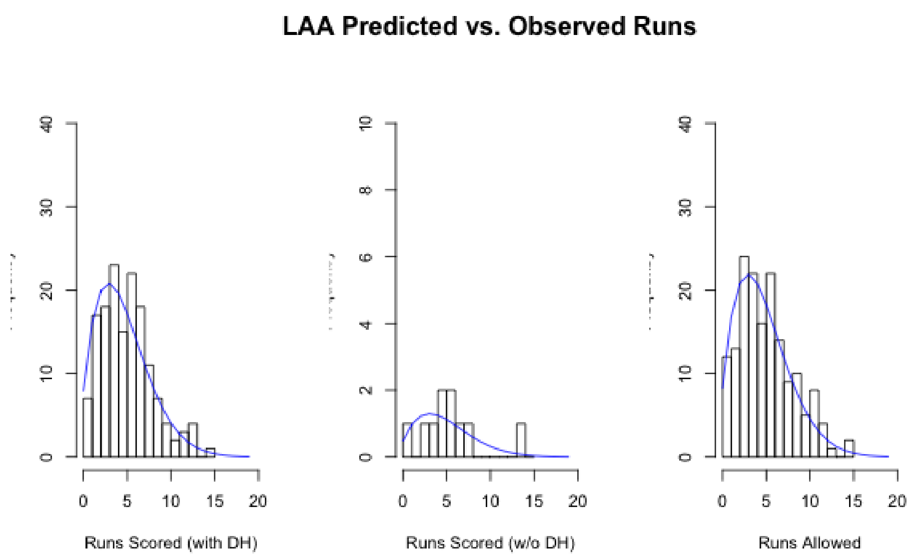

FIGURE A.16: Fitted Weibull Distributions for the Los Angeles Angels, Incorporating Designated Hitters

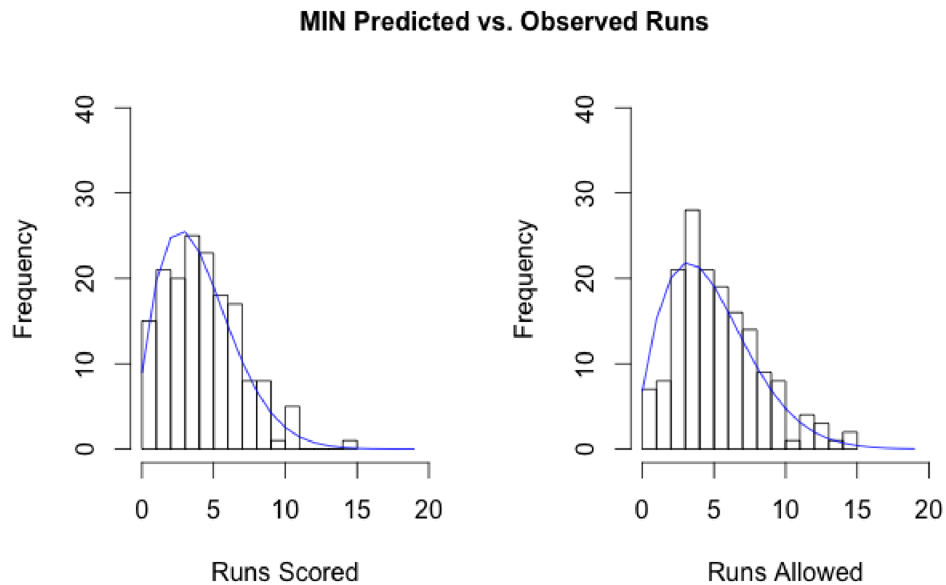

Figure A.17: Fitted Weibull Distributions for the Minnesota Twins

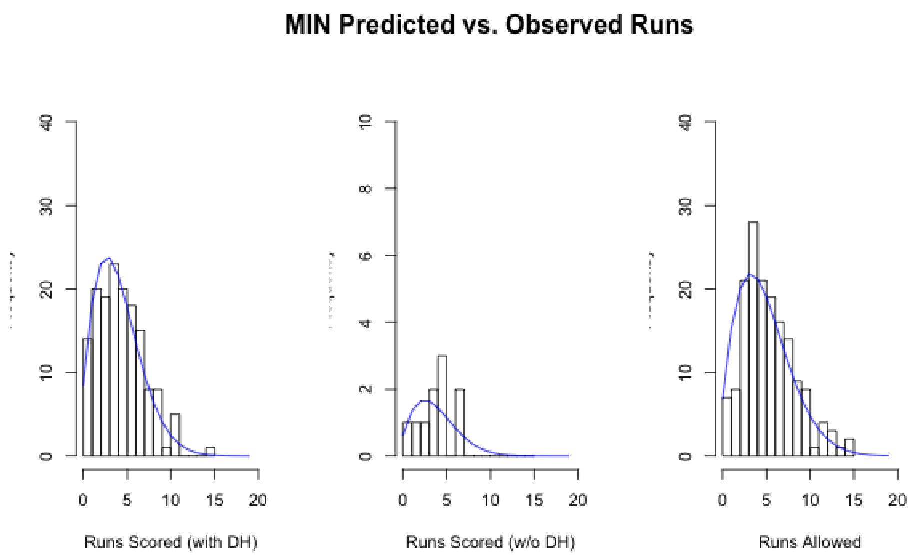

Figure A.18: Fitted Weibull Distributions for the Minnesota Twins, Incorporating Designated Hitters

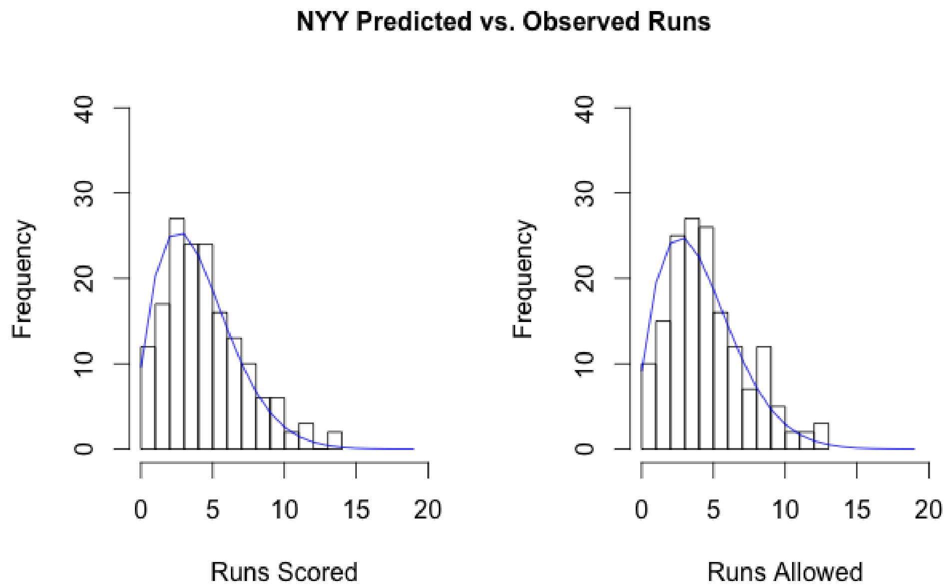

FIGURE A.19: Fitted Weibull Distributions for the New York Yankees

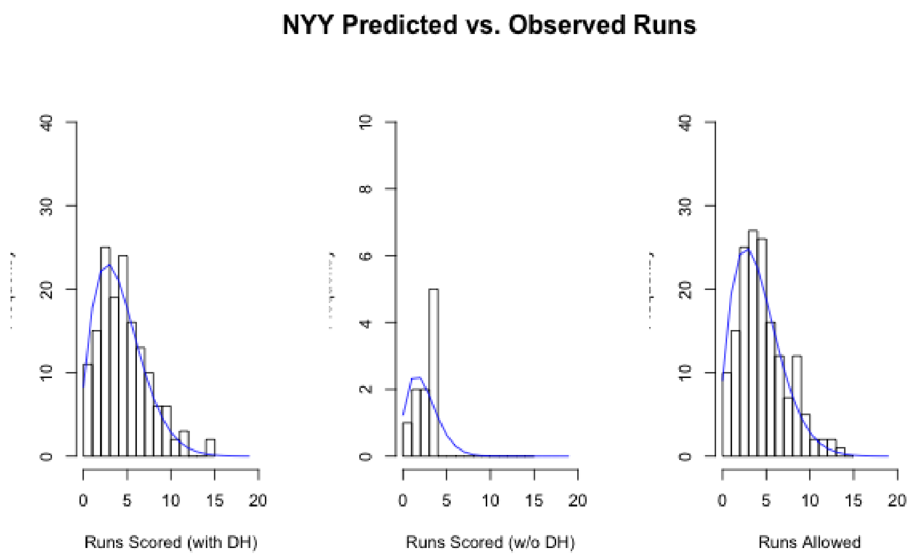

FIGURE A.20: Fitted Weibull Distributions for the New York Yankees, Incorporating Designated Hitters

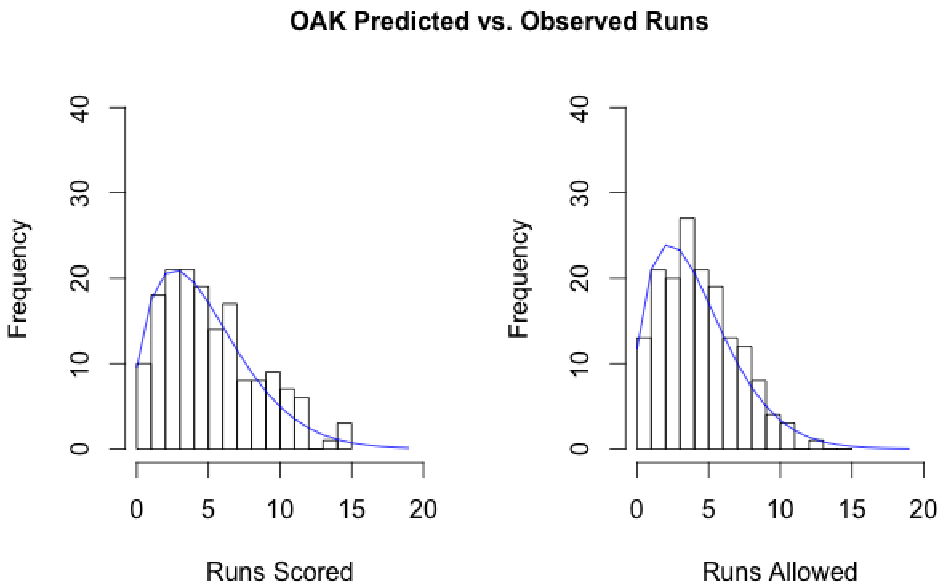

Figure A.21: Fitted Weibull Distributions for the New York Yankees

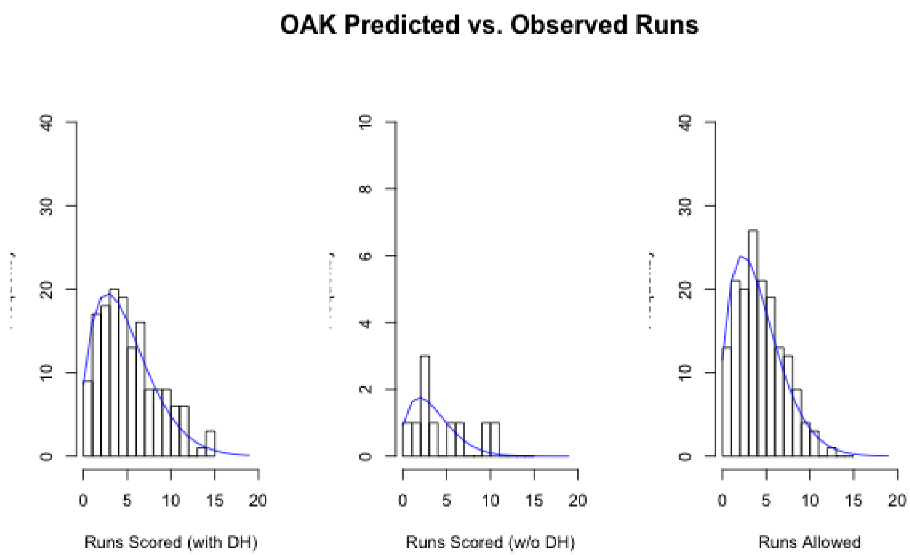

Figure A.22: Fitted Weibull Distributions for the New York Yankees, Incorporating Designated Hitters

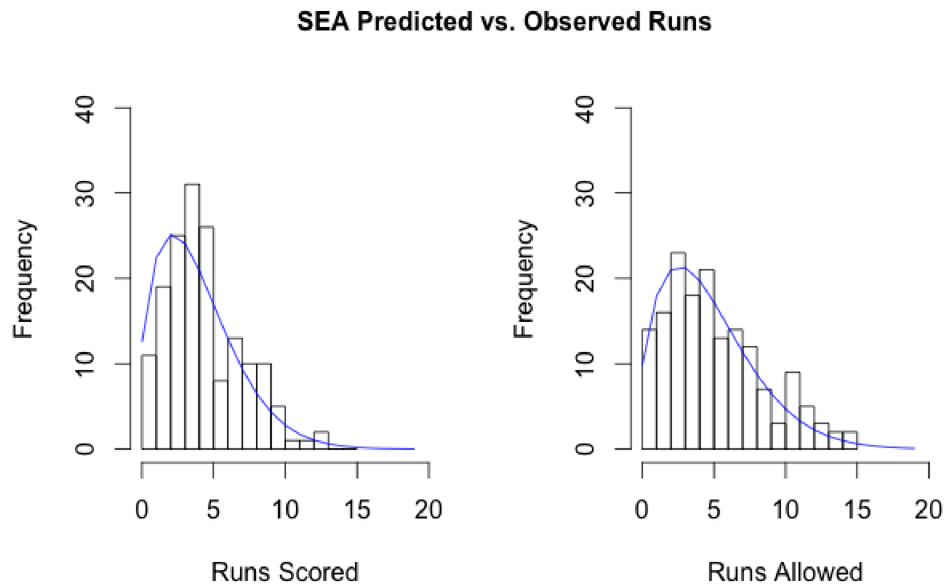

FIGURE A.23: Fitted Weibull Distributions for the Seattle Mariners

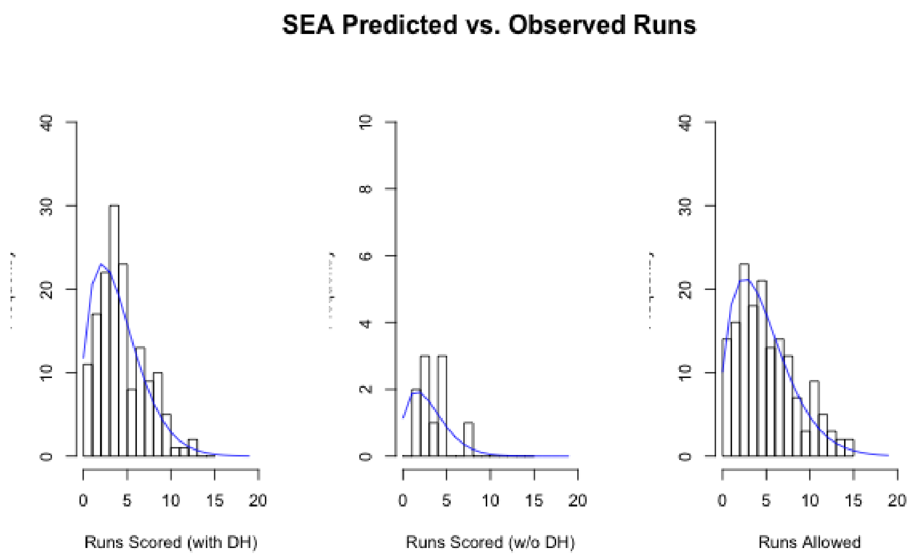

FIGURE A.24: Fitted Weibull Distributions for the Seattle Mariners, Incorporating Designated Hitters

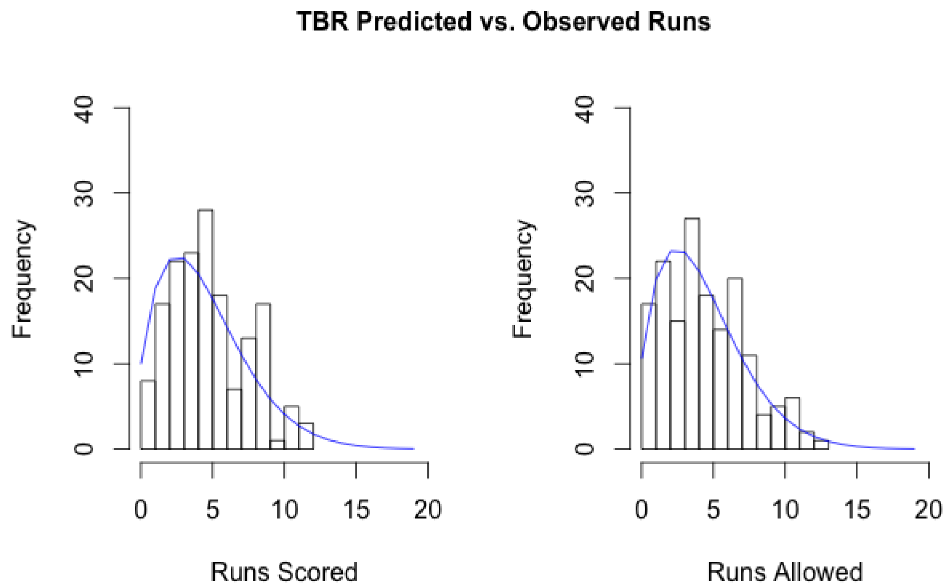

FIGURE A.25: Fitted Weibull Distributions for the Tampa Bay Rays

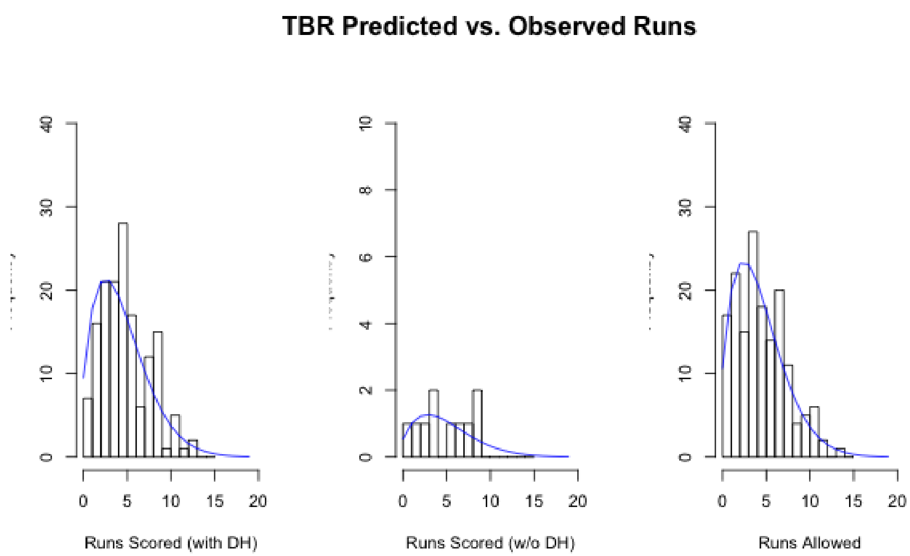

FIGURE A.26: Fitted Weibull Distributions for the Tampa Bay Rays, Incorporating Designated Hitters

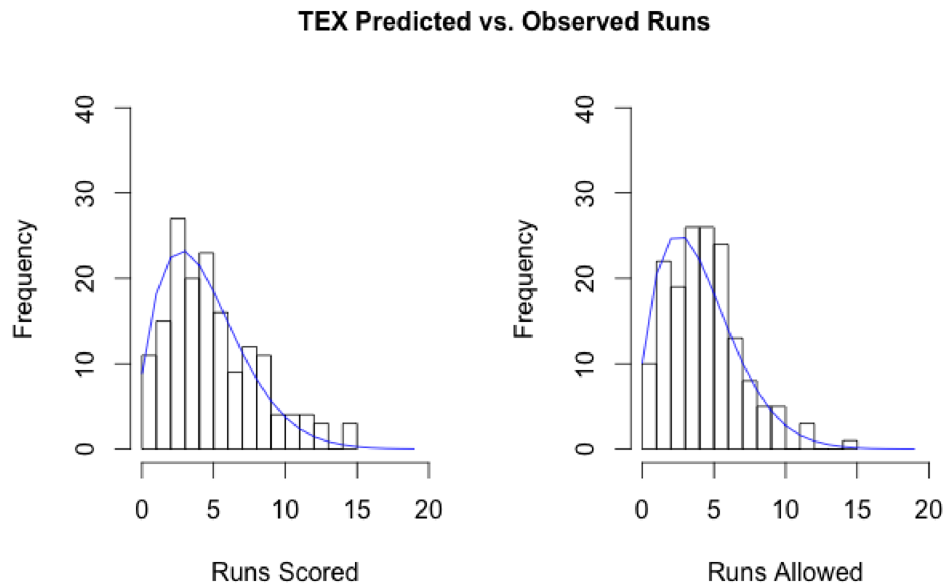

Figure A.27: Fitted Weibull Distributions for the Texas Rangers

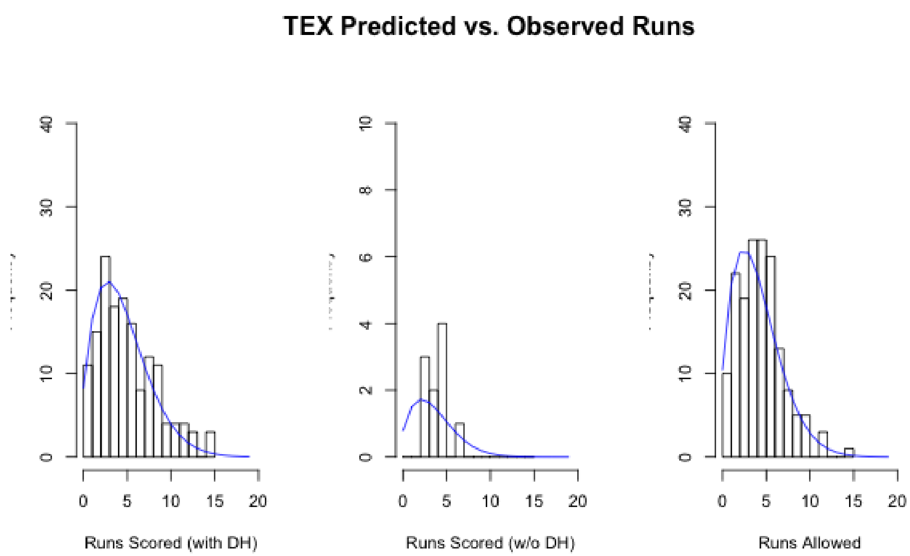

Figure A.28: Fitted Weibull Distributions for the Texas Rangers, Incorporating Designated Hitters

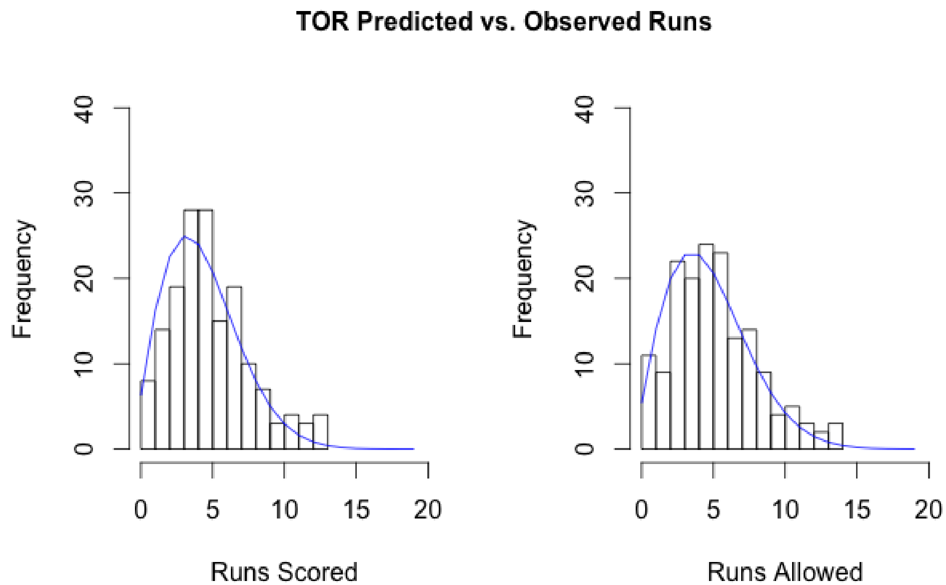

Figure A.29: Fitted Weibull Distributions for the Toronto Blue Jays

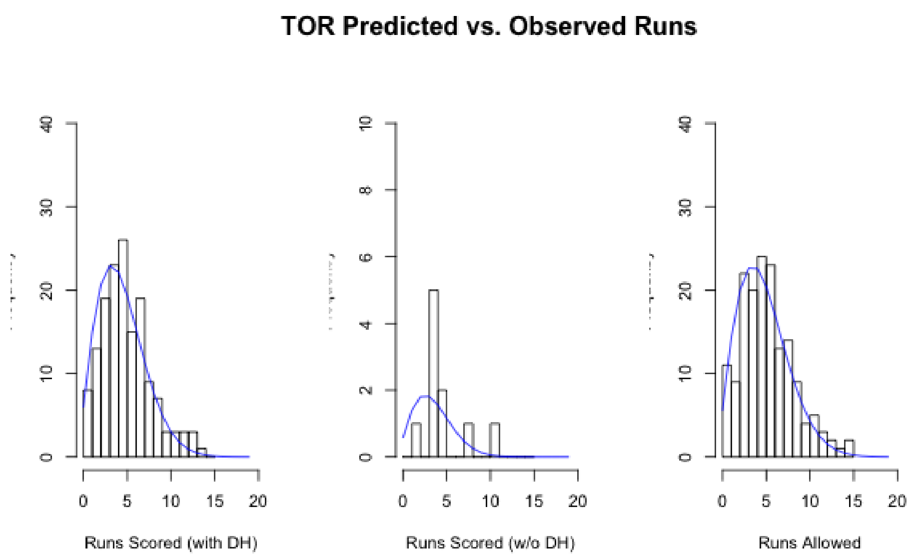

Figure A.30: Fitted Weibull Distributions for the Toronto Blue Jays, Incorporating Designated Hitters

Appendix B

Fitted Weibull Distributions: Divisional Analysis

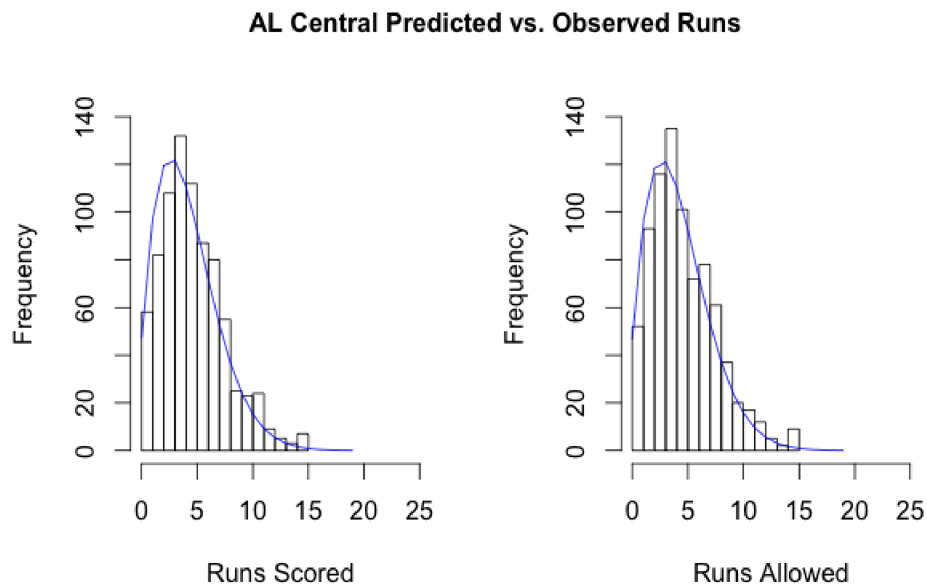

FIGURE B.1: Fitted Weibull Distributions for the AL Central

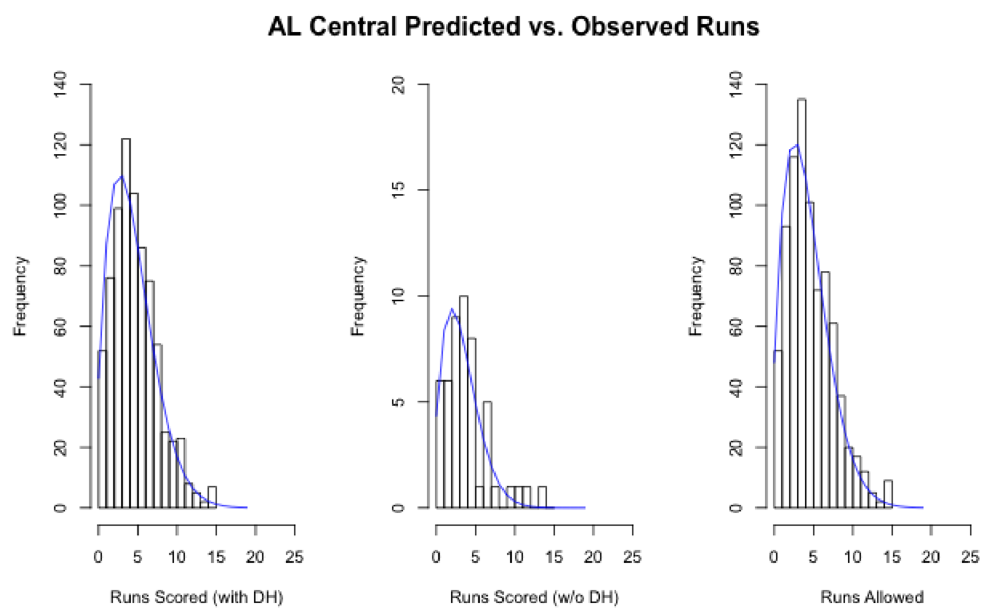

FIGURE B.2: Fitted Weibull Distributions for the AL Central, Incorporating Designated Hitters

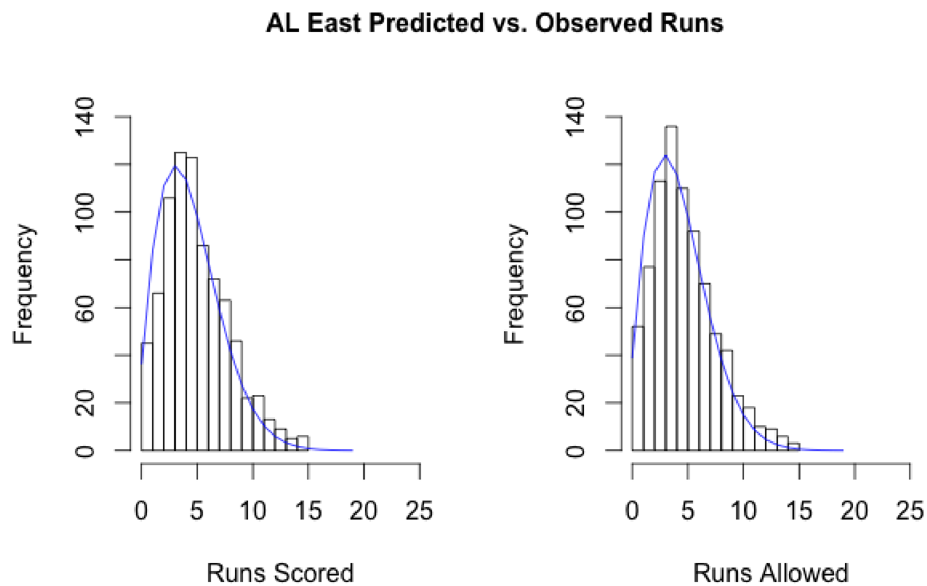

Figure B.3: Fitted Weibull Distributions for the AL East

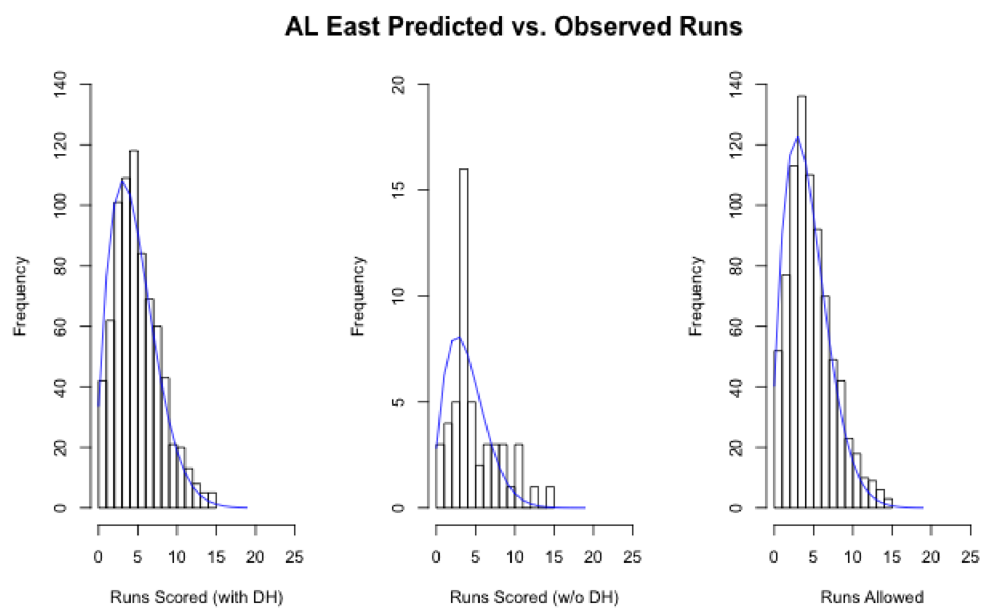

Figure B.4: Fitted Weibull Distributions for the AL East, Incorporating Designated Hitters

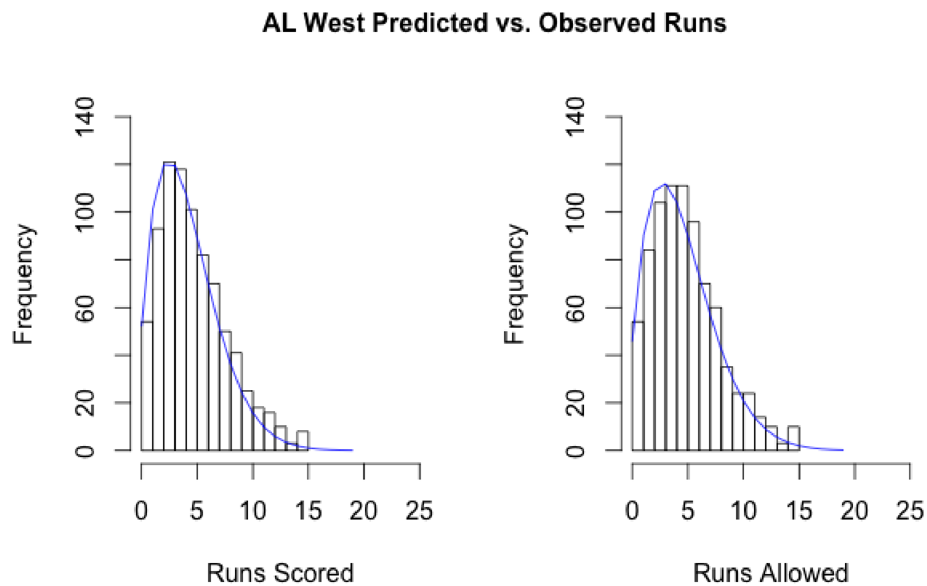

FIGURE B.5: Fitted Weibull Distributions for the AL West

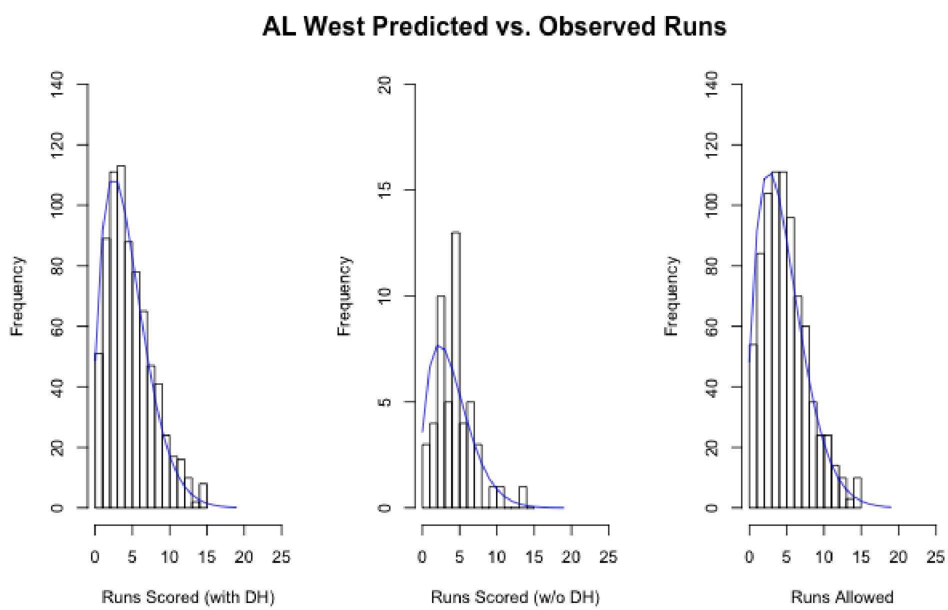

FIGURE B.6: Fitted Weibull Distributions for the AL West, Incorporating Designated Hitters

Bibliography

Baseball-Reference.com (2015), "Pythagorean Theorem of Baseball," Online, http://www.baseball-reference.com/bullpen/Pythagorean_Theorem_of_Baseball.

Ciccolella, R. (2006), "Are Runs Scored and Runs Allowed Independent?" *By the Numbers*, 16, 11–15.

Forbes, C., Evans, M., Hastings, N., and Peacock, B. (2011), "Weibull Distribution," in *Statistical Distributions*, chap. 46, pp. 196–198, Wiley.

Holm, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6, 65–70, http://www.jstor.org/stable/4615733.

Luo, V. (2014), "Relieving and Readjusting Pythagoras: Improving the Pythagorean Formula," http://web.williams.edu/Mathematics/sjmiller/public_html/math/papers/st/VictorLuo.pdf.

Miller, S. J. (2006), "A Derivation of the Pythagorean Won-Loss Formula in Baseball," http://arxiv.org/abs/math/0509698.

Perrett, J. J. and Mundfrom, D. J. (2010), "Bonferroni Procedure," in *Encylcopedia of Research Design*, ed. N. J. Salkind, pp. 98–102, Sage Publications, Inc., http://dx.doi.org/10.4135/9781412961288.

Rosenfeld, J. W., Fisher, J. I., Adler, D., and Morris, C. (2010), "Predicting Overtime with the Pythagorean Formula," *Journal of Quantitative Analysis in Sports*, 6.

Tuerlinckx, F. (2010), "Weibull Distribution," in *Encylcopedia of Research Design*, ed. N. J. Salkind, pp. 1616–1618, Sage Publications, Inc., http://dx.doi.org/10.4135/9781412961288.

Tymins, A. (2014), "Bringing Pythagorean Expectation to College Lacrosse," Online, http://harvardsportsanalysis.org/2014/05/bringing-pythagorean-expectation-to-college-lacrosse/.