The Effect of "Freebies" on Runs Allowed and Winning in NCAA Division I Baseball

Daniel J. Mundfrom, Michelle L. DePoy Smith
Eastern Kentucky University, 521 Lancaster Avenue, Richmond, KY 40475

Abstract
Morgan Ensberg, a former Major League Baseball player and current minor league coach in the Houston Astros organization, developed a statistic to measure the impact of "freebies" on the outcome of a baseball game. The Morgan Ensberg Index (MEI) is a composite of walks, errors, stolen bases allowed, wild pitches, and hit batsmen. In 2012, Mundfrom and Smith extended the MEI to include balks, passed balls, and catcher's interference and investigated the relationship between "freebies" and runs allowed and games won for NCAA Division I baseball teams using data from the 2011 and 2012 seasons. The analyses presented here are an extension of those results using data from the 2013, 2014, and 2015 seasons. In addition, the use of a logistic regression model is explored to analyze individual game data from the 2014 and 2015 seasons for selected teams in NCAA Division I baseball.

Key Words: Baseball, Freebies, NCAA, Runs Allowed, Winning

1. Introduction

Former Eastern Kentucky University (EKU) baseball coach, Jason Stein, posed the following question, "What is the impact of "Freebies" on winning in baseball? The motivation behind the question was to attempt to quantify this effect to make it easier for players to realize the consequences of their actions while in the field that allow the opposing team to gain some advantage that they did not earn. Players (and coaches) are well aware that miscues while in the field are not productive and can have undesirable consequences, but determining the actual effect of these actions may help his players to concentrate harder, be more vigilant in their approach and work harder to eliminate these actions.

For the purpose of this paper, freebies are defined to be "an action by the defensive team that allows at least one base runner and/or batter to advance at least one base." Coach Stein provided a list of eight such actions that he felt fit into this category: walks, hit batsmen, errors, passed balls, balks, wild pitches, stolen bases allowed, and catcher's interference.

2. Data

Season totals for each of these variables, plus number of games played, number of wins, number of losses, number of runs scored, number of runs allowed, and the number

intentional walks were recorded for each NCAA Division I baseball team during the 2011, 2012, 2013, 2014, and 2015 seasons, for a total sample size of n = 1497.

3. Previous Research

Mundfrom and Smith (2012) performed simple linear regression analyses predicting both runs allowed and wins using NCAA Division I baseball data for the 2011 and 2012 seasons. For the variable "runs allowed" their results showed that, on average, for every three freebies committed your opponent scored two more runs, and again on average, for every three freebies committed you will win 11 fewer games. These models, respectively, explained about 64% and 39% of the variation in the response variables. Schaffer, Mundfrom, and Smith (2013) examined the same question and performed similar analyses for Major League Baseball (MLB) using data from the 2003-2012 seasons. A comparison between MLB and NCAA Division I showed that freebies have a smaller effect in the major leagues than they do in college baseball.

4. Analysis I: Analysis of Team Data

The same analyses performed by Mundfrom and Smith (2012) were performed here using data from all NCAA Division I baseball teams for the 2011 – 2015 seasons, i.e., separate simple linear regressions were performed to predict runs allowed and number of wins using the number of freebies committed as the predictor variable.

5. Analysis II: Analysis of Individual Game Data

Logistic Regression Models were used to predict the number of wins using the number of freebies committed per game for individual game data. These analyses were performed for selected teams due to the tremendous amount of data that needed to be obtained and analyzed. The following analyses were performed using individual game data:

- EKU data for the 2014 season using n = 53 games (2 unusual games were removed from the analysis).
- The 16 teams (n = 277) that qualified for the 2014 College World Series (CWS).
- A sample of 12 teams from the 2015 season – we randomly selected 5 teams that played in the NCAA Tournament, five teams that did not qualify for the NCAA Tournament, and the CWS winner and runner-up.

6. Results

6.1 Results of Team Data

The team data were analyzed using simple linear regression analysis. Summary statistics for the number of freebies per game for the 2011 – 2015 seasons were as follows: $n = 1497$, mean = 8.16, standard deviation = 1.58, minimum = 4.03, and maximum = 15.08. The simple linear regression model $y = -0.103 + 0.670x$ was used to predict the runs allowed

per game using freebies per game. Using this model we can say on average, for every three freebies committed per game, your opponent will score 2 more runs per game. This model explains about 66% of the variation in the runs allowed per game. The 95% Prediction Interval to predict the number of runs allowed with 8 freebies per game is from 3.8 to 6.7 runs. The scatterplot and regression line for all five seasons is presented in Figure 1.

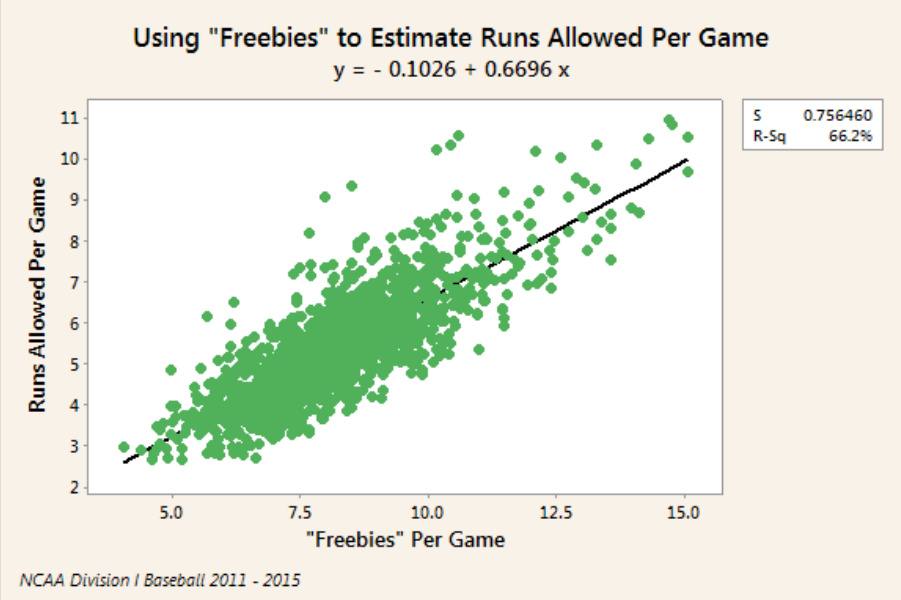

Figure 1.

We separately analyzed the team data for the five seasons, the results are presented in Table 1, which shows the model has been quite consistent over the past 5 seasons.

Table 1.

Year	**Intercept**	**Slope**	**R^2**	**P-value**	**95% PI for Predicted Runs Per Game 8 Freebies are Allowed.**		
					Lower Limit	**Point Estimate**	**Upper Limit**
2011	-0.16	0.70	0.64	<.0001	3.80	5.43	7.06
2012	-0.06	0.67	0.68	<.0001	3.82	5.28	6.74
2013	0.14	0.63	0.68	<.0001	3.84	5.16	6.49
2014	-0.28	0.67	0.65	0.001	3.57	5.07	6.57
2015	-0.06	0.67	.69	<.0001	3.89	5.33	6.76
All	-0.10	0.67	0.66	<.0001	3.77	5.25	6.74

We also examined the team data analysis by conference, which are presented in Table 2. Eastern Kentucky University is a member of the Ohio Valley Conference (OVC). For the

OVC, we obtained the regression equation: $y = 0.75 + 0.65x$. The 95% Prediction Interval is from 4.3 to 7.7 runs per game when 8 freebies are allowed. This model explains about 43% of the variation in the runs allowed per game.

Table 2.

Conference	**# of Obs.**	**Intercept**	**Slope**	**R^2**	**Model**	**95% PI for Predicted Runs Per Game 8 Freebies are Allowed.**		
					P-value	**Lower Limit**	**Point Estimate**	**Upper Limit**
AMERICAN ATHLETIC	27	0.38	0.54	0.56	<.0001	3.2	4.72	6.25
ATLANTIC COAST	64	0.38	0.55	0.59	<.0001	3.68	4.82	5.95
AMERICA EAST	32	0.56	0.6	0.44	<.0001	2.93	5.39	7.85
ATLANTIC 10	65	0.6	0.56	0.61	<.0001	3.94	5.11	6.29
ATLANTIC SUN	50	0.97	0.58	0.47	<.0001	3.8	5.64	7.48
BIG 12	46	-0.73	0.73	0.57	<.0001	3.93	5.12	6.31
BIG EAST	38	0.55	0.56	0.61	<.0001	3.91	5.04	6.18
BIG SOUTH	57	0.5	0.6	0.54	<.0001	3.88	5.27	6.66
BIG TEN	56	-0.1	0.68	0.53	<.0001	3.64	5.33	7.01
BIG WEST	46	0.09	0.65	0.67	<.0001	4	5.29	6.58
COLONIAL	49	-0.03	0.71	0.66	<.0001	4.2	5.64	7.09
CONFERENCE USA	54	0.79	0.51	0.45	<.0001	3.6	4.87	6.13
GREAT WEST	27	0.2	0.7	0.62	<.0001	3.76	5.78	7.8
HORIZON	30	1.25	0.52	0.58	<.0001	3.67	5.43	7.2
INDEPENDENT	15	0.7	0.62	0.41	0.0097	2.32	5.65	8.99
IVY LEAGUE	40	0.66	0.6	0.62	<.0001	3.92	5.47	7.01
MID-EASTERN	43	-1.01	0.77	0.85	<.0001	3.78	5.13	6.48
METRO ATLANTIC	49	0.09	0.64	0.77	<.0001	4.13	5.2	6.28
MID-AMERICAN	60	1.01	0.54	0.46	<.0001	3.188	5.3	6.72
MISSOURI VALLEY	40	0.17	0.59	0.54	<.0001	3.72	4.93	6.14
MOUNTAIN WEST	32	1.89	0.46	0.18	0.0148	3.72	5.59	7.47
NORTHEAST	41	-0.3	0.7	0.68	<.0001	3.95	5.33	6.71
OHIO VALLEY	51	0.75	0.65	0.43	<.0001	4.28	5.98	7.68
PAC-12	54	0.6	0.54	0.45	<.0001	3.62	4.93	6.24
PATRIOT	30	-0.93	0.76	0.8	<.0001	4.13	5.11	6.1
SOUTH EASTERN	66	0.56	0.53	0.35	<.0001	3.44	4.84	6.23
SOUTHERN	53	0.93	0.56	0.54	<.0001	4.29	5.45	6.6
SOUTHLAND	61	-0.02	0.64	0.59	<.0001	3.96	5.06	6.17
SUN BELT	51	1.82	0.45	0.39	<.0001	4.17	5.4	6.63
WEST COAST	46	0.43	0.59	0.47	<.0001	3.87	5.11	6.36
SUMMIT	30	0.59	0.58	0.69	<.0001	3.78	5.22	6.65

SOUTH WESTERN	50	-0.69	0.72	0.76	<.0001	3.26	5.06	6.85
WESTERN ATHLETIC	44	-0.58	0.76	0.69	<.0001	4.03	5.49	6.95

We additionally analyzed the team data by comparing those teams that qualified for the NCAA Tournament with those that did not. The slope for teams that qualified for the NCAA tournament is 0.07 less than for teams that did not qualify for the tournament (P = 0.057). The scatterplot of the data with separate regression equations is displayed in Figure 2.

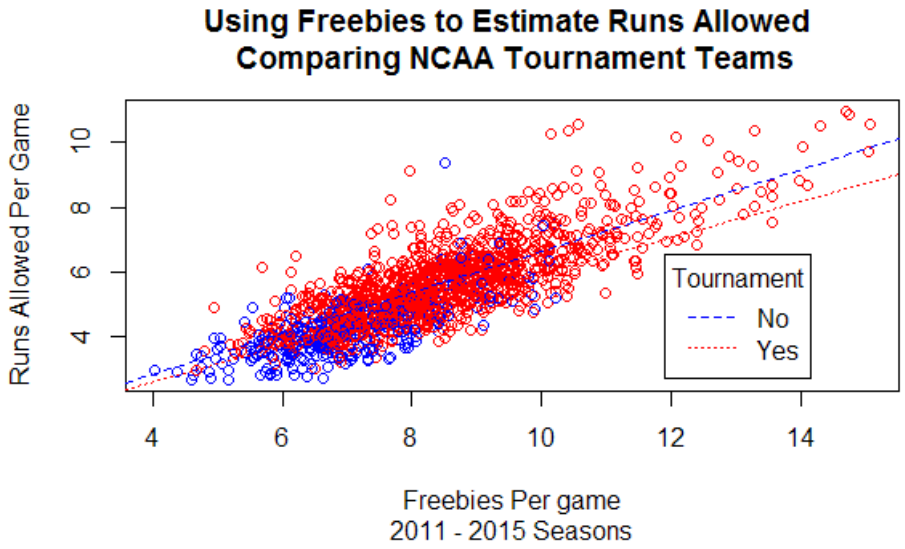

Figure 2.

6.2 Results of Individual Game Data

The individual game data were analyzed using logistic regression models. We first analyzed the 53 games for Eastern Kentucky University during the 2014 season. The two games versus Youngstown were omitted from the analysis due to unusually high scoring by both teams. The fitted logistic regression model $\hat{y} = \frac{e^{(1.1277-0.1296x)}}{1+e^{(1.1277-0.1296x)}}$ was significant (P = 0.0118, R^2 = 11.3) For each additional freebie (X) in a game, we estimate the odds of winning to decrease by 12%. 95% CI: (0.777, 0.973). In addition, the probability of winning when 8 freebies are committed in a game is 0.52. 95% CI: (.363, .678). The data and model are presented in Figure 3.

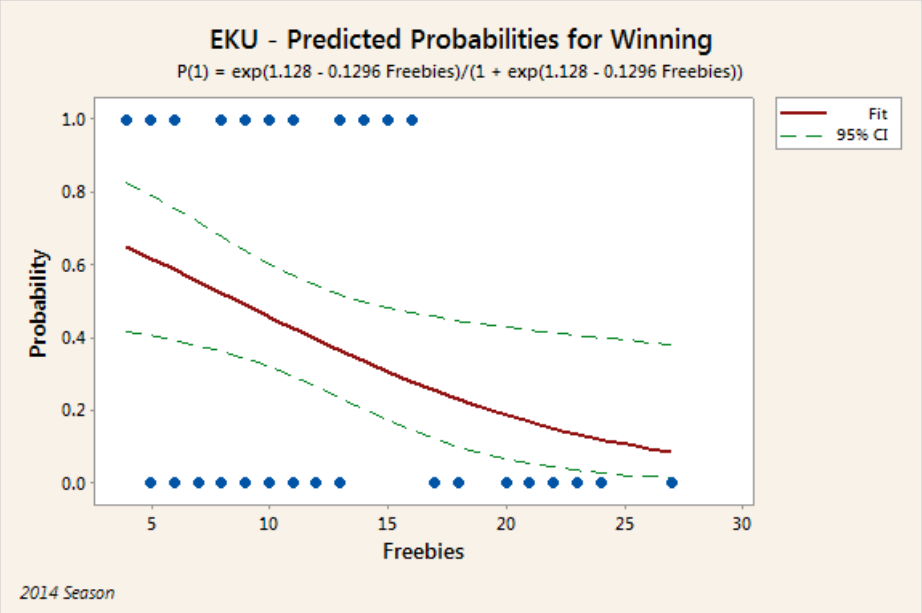

Figure 3.

Our next analysis concerned the eight teams that played in the 2014 College World Series. A total of 277 games were analyzed for the eight teams which resulted in the fitted logistic regression model $\hat{y} = \frac{e^{(2.5228-0.2208x)}}{1+e^{(2.5228-0.2208x)}}$ which was significant ($P < 0.0001$, $R^2 = 10.8\%$). For each additional freebie (X) in a game, we estimate the odds of winning to decrease by 20%; 95% CI: (0.738 0.871). The probability of winning when 8 freebies are committed in a game is 0.68; 95% CI: (.615, .740). The data and model are presented in Figure 4.

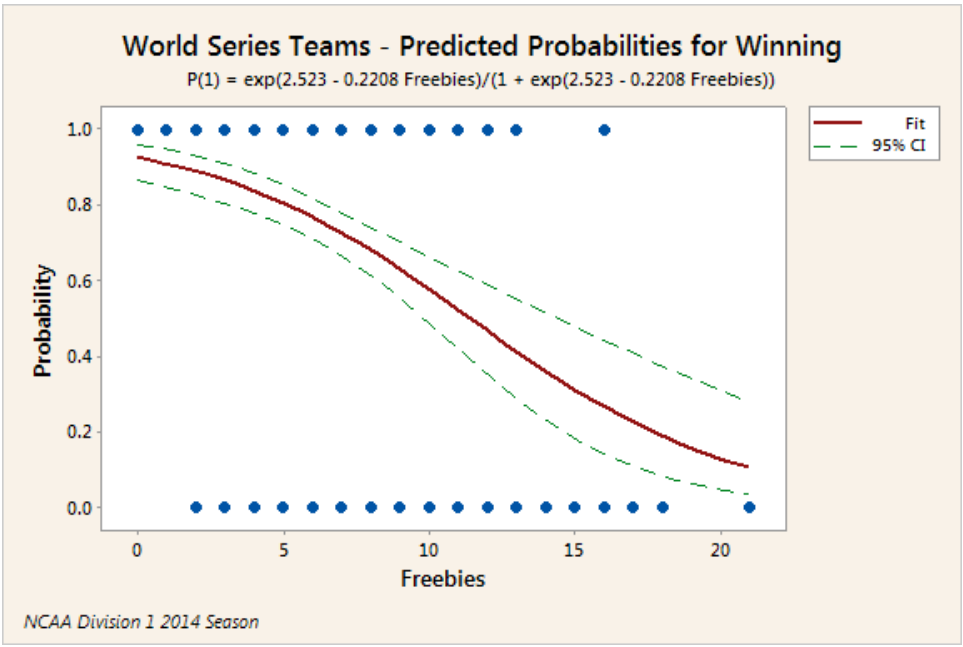

Figure 4.

For the 2015 season we randomly selected five teams from those who qualified for the NCAA tournament and five teams who did not qualify for the tournament, and the winner and runner-up of the college world series which resulted in a total of 713 games. The logistic model using the predictors freebies per game, game location (home, away, or neutral), and NCAA tournament (Yes, No) to predict the probability of winning was significant ($P < 0.0001$, $R^2 = 19\%$). In addition, each predictor was significant at the 5% significance level. We are 95% confident that for each additional freebie in a game, the odds of winning decreases by between 16% and 23%, while holding location and NCAA tournament play at fixed values. The additional odds ratio confidence intervals are contained in Table 3.

Table 3.

Predictor	**Odds Ratio**	**95% Wald Confidence Limits**
Freebies	0.803	(0.768, 0.840)
Location: Away Vs. Neutral	0.560	(0.328, 0.957)
Location: Home Vs. Neutral	1.021	(0.600, 1.737)
Location: Away Vs. Home	0.549	(0.386, 0.781)
NCAA: No Vs. Yes	0.575	(0.410, 0.807)

We also computed logistic regression models to predict wins (Y) using the number of freebies committed per game (X), which was computed for each team individually. The odds ratio estimate and confidence intervals for freebies are presented in Table 4.

Table 4.

School	**NCAA Tournament**	**Odds Ratio**	**95 % Confidence Limits**	
Chicago State	No	0.677	0.527	0.869
Lipscomb	No	0.791	0.669	0.935
UNC Asheville	No	0.891	0.792	1.001
Valparaiso	No	0.849	0.704	1.024
Wichita State	No	0.655	0.517	0.831
California	Yes	0.837	0.704	0.994
Mercer	Yes	0.725	0.594	0.885
Michigan	Yes	0.770	0.652	0.909
NC State	Yes	0.828	0.725	0.946
Texas State	Yes	0.757	0.634	0.905
Vanderbilt	Yes World Series Runner-Up	0.957	0.839	1.093
Virginia	Yes World Series Champion	0.743	0.621	0.889

7. Conclusions

Overall, the results were very consistent from year to year, as well as across all five years combined. No attempt was made to determine if any one freebie, or smaller subset thereof, was more detrimental than the others. For most teams the number of balks and catcher's interference occurrences were very small, so, individually, they are not as likely to be strong predictors. However, the overall focus was the effect of all freebies, not individual ones, so distinguishing one from the other did not appear useful and was not performed.

Obviously, all freebies are not "created equal." The effect of a walk in the first inning, an error in the fourth inning, and allowing a stolen base in the sixth inning, may not lead to any runs being scored by a team's opponent. On the other hand, a two-out error that prolongs an inning, or two hits, a walk, and a homerun, may result in several runs being scored as a result of that one single freebie.

Not surprisingly, the teams that qualified for the NCAA Tournament, and those that subsequently played in the CWS tend to have fewer runs allowed per freebie than those teams that did not qualify for Tournament play.

8. References

Mundfrom, D. J. & Smith, M. D. (2012). The effect of "freebies" on winning in baseball. A paper presented at the Joint Statistical Meetings, San Diego, CA, July, 2012.

Schaffer, J., Mundfrom, D. J., & Smith, M. D. (2013). The effect of "freebies" on run production in major league baseball. A paper presented at the Joint Statistical Meetings, Montreal, Quebec, Canada, August, 2013.