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

Daniel J. Mundfrom **Michelle L. DePoy Smith**
Eastern Kentucky University

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. Simple linear regression models were computed to investigate the relationship between freebies and runs allowed and games won for National College Athletic Association (NCAA) Division I baseball teams using data from the 2011 - 2015 seasons. Using this model we can say on average, for every three freebies committed per game, your opponent will score 2 more runs per game. Binary logistic regression models were constructed to analyze individual game data from the 2015 season for selected NCAA Division I baseball teams. We estimate for each additional freebie in a game, the odds of winning decreases by between 16% and 23%, while holding game location and NCAA tournament participation at fixed values.

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 players to concentrate harder, be more vigilant in their approach and work harder to eliminate these actions. 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.

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 (Ensberg, 2011). The Morgan Ensberg Index (MEI) is a composite of walks, errors, stolen bases allowed, wild pitches, and hit batsmen.

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 per game" their results showed that, on average, for every three freebies committed per game your opponent scored two more runs, and again on average, for every three freebies committed per game you will win 11 fewer games. These models, respectively, explained about 64% and 39% of the variation in the response variables. The authors also found similar results using the MEI as a predictor for the variable "runs allowed per game." Their results showed that, on average, for every three MEI freebies committed per game your opponent scored two more runs, with the model explaining about 66% of the variation in runs allowed per game. They also determined that including intentional walks did not affect the analyses. 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.

Herein, we extend the analyses of Mundfrom and Smith (2012) for the team data using data from all NCAA Division I baseball teams for the 2011–2015 seasons. In addition, individual game data from the 2014 and 2015 seasons for selected NCAA Division I baseball teams will be used to construct logistic regression models to predict the number of wins using the number of freebies committed per game.

Data

Season totals were recorded for each NCAA Division I baseball team for each of the 2011- 2015 seasons for each of eight freebie variables, plus number of games played, number of wins, number of losses, number of runs scored, number of runs allowed, number of intentional walks, and NCAA tournament participation (no, yes). The resulting data totaled n = 1497 team results to be analyzed. Most teams will have five sets of observations, one for each season. However, yearly differences in team rosters and schedules should minimize the dependence among the observations. We computed the variable, freebies per game, as the season total for all eight freebie variables combined divided by the number of games played. In addition, we examined the variable, runs per game, as the season total of number of runs scored divided by the number of games played.

Due to the tremendous amount of individual game data we chose not to examine all individual game data for all NCAA Division I baseball teams. For the individual game data, the team, opposing team, game location (away, home, neutral), outcome (loss, win), tie (no, yes), total number of the eight freebies, number of runs scored, number of runs allowed, number of hits, number of hits allowed, number of intentional walks, NCAA tournament participation (no, yes), World Series appearance (no, yes) were recorded. We randomly selected five teams that played in the NCAA Tournament, randomly selected five teams that did not qualify for the NCAA Tournament, and the College World Series (CWS) winner and runner-up which yielded a sample size of n = 713 individual games.

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. 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 $\hat{y} = -0.103 + 0.670x$ 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 appears in 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. We also performed the team data analysis by conference, with results presented in Table 2. Eastern Kentucky University is a member of the Ohio Valley Conference (OVC). For the OVC, we obtained the regression equation: $\hat{y} = 0.75 + 0.65x$. 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.

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 which did not qualify for the tournament ($p = 0.057$) which was not statistically significant at the 5% significance level. The scatterplot of the data with separate regression equations is displayed in Figure 2.

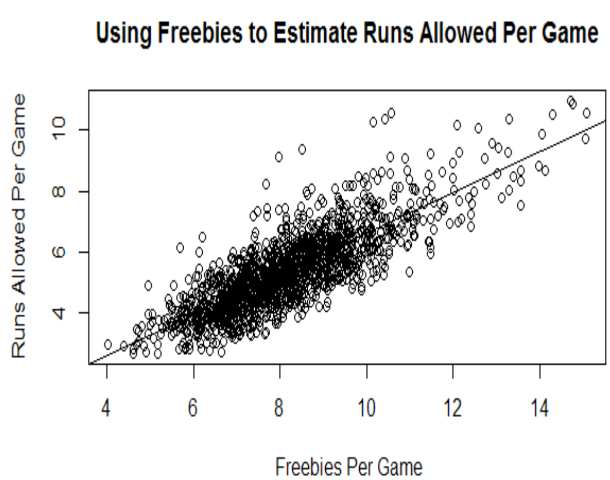

Figure 1. Plot of "Freebies" and Runs per Game

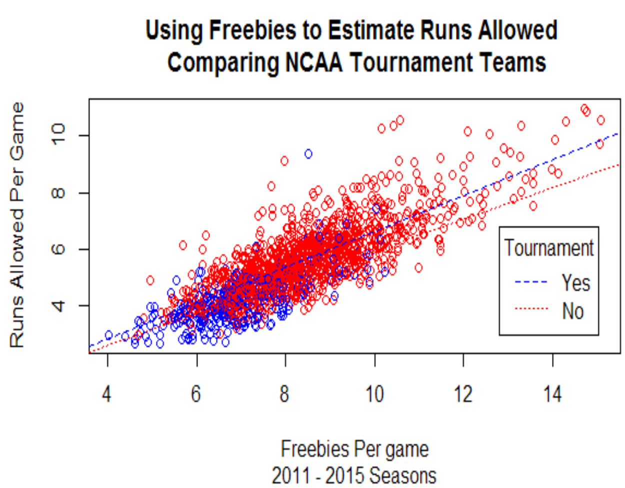

Figure 2. Plot of "Freebies" and Runs per Game

Table 1. Linear Regression Model Results by Seasons with Prediction Intervals for 8 Freebies per Game

					95% PI for Predicted Runs Per Game		
					Lower Limit	Point Estimate	Upper Limit
2011	-0.16	0.70	0.64	< 0.0001	3.80	5.43	7.06
2012	-0.06	0.67	0.68	< 0.0001	3.82	5.28	6.74
2013	0.14	0.63	0.68	< 0.0001	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	0.69	< 0.0001	3.89	5.33	6.76
All	-0.10	0.67	0.66	< 0.0001	3.77	5.25	6.74

Table 2. Simple Linear Regression Model Results by Conference

					95% PI for Predicted Runs Per Game		
					Lower Limit	Point Estimate	Upper Limit
American Athletic	0.38	0.54	0.56	< 0.0001	3.20	4.72	6.25
Atlantic Coast	0.38	0.55	0.59	< 0.0001	3.68	4.82	5.95
American East	0.56	0.60	0.44	< 0.0001	2.93	5.39	7.85
Atlantic 10	0.60	0.56	0.61	< 0.0001	3.94	5.11	6.29
Atlantic Sun	0.97	0.58	0.47	< 0.0001	3.80	5.64	7.48
Big 12	-0.73	0.73	0.57	< 0.0001	3.93	5.12	6.31
Big East	0.55	0.56	0.61	< 0.0001	3.91	5.04	6.18
Big South	0.50	0.60	0.54	< 0.0001	3.88	5.27	6.66
Big Ten	-0.10	0.68	0.53	< 0.0001	3.64	5.33	7.01
Big West	0.09	0.65	0.67	< 0.0001	4.00	5.29	6.58
Colonial	-0.03	0.71	0.66	< 0.0001	4.20	5.64	7.09
Conference USA	0.79	0.51	0.45	< 0.0001	3.60	4.87	6.13
Great West	0.20	0.70	0.62	< 0.0001	3.76	5.78	7.80
Horizon	1.25	0.52	0.58	< 0.0001	3.67	5.43	7.20
Independent	0.70	0.62	0.41	0.0097	2.32	5.65	8.99
Ivy League	0.66	0.60	0.62	< 0.0001	3.92	5.47	7.01
Mid-Eastern	-1.01	0.77	0.85	< 0.0001	3.78	5.13	6.48
Metro Atlantic	0.09	0.64	0.77	< 0.0001	4.13	5.20	6.28
Mid-American	1.01	0.54	0.46	< 0.0001	3.19	5.30	6.72
Missouri Valley	0.17	0.59	0.54	< 0.0001	3.72	4.93	6.14
Mountain West	1.89	0.46	0.18	0.0148	3.72	5.59	7.47
Northeast	-0.30	0.70	0.68	< 0.0001	3.95	5.33	6.71
Ohio Valley	0.75	0.65	0.43	< 0.0001	4.28	5.98	7.68
PAC-12	0.60	0.54	0.45	< 0.0001	3.62	4.93	6.24
Patriot	-0.93	0.76	0.8	< 0.0001	4.13	5.11	6.10
South Eastern	0.56	0.53	0.35	< 0.0001	3.44	4.84	6.23
Southern	0.93	0.56	0.54	< 0.0001	4.29	5.45	6.60
Southland	-0.02	0.64	0.59	< 0.0001	3.96	5.06	6.17
Sun Belt	1.82	0.45	0.39	< 0.0001	4.17	5.40	6.63
West Coast	0.43	0.59	0.47	< 0.0001	3.87	5.11	6.36
Summit	0.59	0.58	0.69	< 0.0001	3.78	5.22	6.65
South Western	-0.69	0.72	0.76	< 0.0001	3.26	5.06	6.85
Western Atlantic	-0.58	0.76	0.69	< 0.0001	4.03	5.49	6.95

Individual Game Data

For the 2015 season we randomly selected five teams from those which qualified for the NCAA tournament and five teams which did not qualify for the tournament, and the winner and runner-up of the CWS which resulted in a total of 713 games. The binary logistic regression model using the predictors freebies per game, game location (away (0), home (1), or neutral (2)), and NCAA tournament participation (no (0), yes (1)) to predict the probability of winning was significant ($P < 0.0001$, Cox & Snell Generalized $R^2 = 19\%$). In addition, the Wald Chi-Square test statistics for all three predictors were significant at the 5% significance level, the results are given in Table 3. The maximum likelihood parameter estimates and standard errors are given in Table 4.

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 presented in Table 5.

Using the average number of 8 freebies per game, Table 6 contains the probability of winning with 95% confidence. Using the binary logistic regression model, we estimate with 95% confidence that the probability that an NCAA tournament selected team will win a game on a neutral site, when 8 freebies per game are committed is between 0.55 and 0.76.

We also computed logistic regression models to predict wins (Y) using the number of freebies committed per game (X), for each team individually. We can say that for the University of Virginia, 2015 CWS Champion, we are 95% confident that for each additional freebie in a game, the odds of winning decreases by between 11% and 38%. The odds ratio estimate and confidence intervals for freebies are presented in Table 7.

Table 3. Wald Chi-Square Tests for Predictors

Effect	df	Wald Chi-Square	P-value
Freebies	1	90.2295	< 0.0001
Home	2	12.2351	0.0022
NCAA	1	10.2547	0.0014

Table 4. Maximum Likelihood Parameter Estimates and Standard Errors (SE) for Binary Logistic Regression Model

Parameter	df	Estimate	SE
Intercept	1	2.4080	0.3005
Freebies	1	-0.2193	0.0231
Home 0	1	-0.5793	0.2733
Home 1	1	0.0207	0.2712
NCAA 0	1	-0.5530	0.1727

Table 5. Odds Ratio and Confidence Limits for Binary Logistic Regression Model Parameters

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)

Table 6. Estimated Probability of Winning for the Binary Logistic Regression Model When 8 Freebies per Game are Committed

Location	NCAA	Winning Probability	95 % Confidence Limits	
Home	No	0.530	0.452	0.608
Away	No	0.383	0.311	0.460
Neutral	No	0.525	0.393	0.654
Home	Yes	0.663	0.597	0.722
Away	Yes	0.519	0.446	0.591
Neutral	Yes	0.658	0.545	0.755

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.

Table 7. Odds Ratios and Confidence Limits for Each School Resulting from Logistic Regression Models Using Freebies Committed per Game to Predict Wins

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 (CWS Runner-Up)	Yes	0.957	0.839	1.093
Virginia (CWS Champion)	Yes	0.743	0.621	0.889

References

Ensberg, M. (2011). *Morgan Ensberg's baseball IQ, The MEI* [Web log post]. Retrieved from https://morganensberg.wordpress.com/2011/06/11/the-mei/

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

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

Send correspondence to: Daniel Mundfrom
Eastern Kentucky University
Email: daniel.mundfrom@eku.edu