Predicting Factors That Influence Kill Probability in Volleyball

Maura Collins

Advisor: Peter Hovey, Ph.D.

Abstract

This project aims to predict the probability of a kill on left-side and right-side attacks in volleyball. A logistic regression model was used to look at the relationship of these variables to the probability of a kill on a single attack.

Data Collection

Data was collected through UD's Women's Volleyball Team and their opponents in 2019 Only data points from teams ranked in the top 50 for RPI that year were used.

Data Characteristics

Attack: accounts for left side vs. right side and for slow vs. faster tempo
- Go: left side, slower
- Black: left side, faster
- 9: right side, slower
- Fire: right side, faster

Pass grade: tells the quality of the first touch or pass for that specific play
- Scale 0-3, 3 being highest quality

Type: illustrates what kind of play this attack occurred during
- First ball: a play off a serve
- Transition: a play during a rally

Logistic Regression

A logistic regression was used to model the probability of a kill for a given play. This used because the dependent variable in is binary, being either a kill or not a kill. This analysis was run for 736 observations with 3 independent variables.

Effect Summary

Source	LogWorth		PValue
Type	2.612		0.00244
Attack	1.383		0.04141

P(Outcome 2) Formula vs. Attack
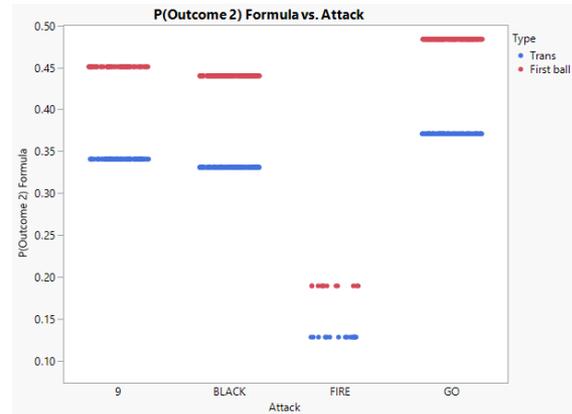
Parameter	Estimate	Std Error	ChiSq	Prob>ChiSq
Intercept	-0.72173	0.14931	32.06	0.0001
Attack[9]	0.29256	0.19111	2.47	0.1159
Attack[Black]	0.24854	0.17385	2.20	0.1382
Attack[Fire]	-0.96561	0.41359	6.99	0.0082
Type[Trans]	-.23175	0.07685	9.18	0.0024

Results

- Data points for the possible combinations of variables were created to account for the probability and occurrence of those specific events.
- The independent variables that are statistically significant at the 5% level are type and attack with p-values 0.00244 and 0.04141, respectively. Pass grade had a p-value that was deemed not significant.
- Coefficients indicate that a play being a transition attack decreases the probability of a kill by 0.23175, compared to it being a first ball attack.

Limitations and Future Work

- Due to accessibility of this type of data, it was mainly for teams in the Mid-West region. Gathering a wider spread of data from across the country could give a better model.
- Creating other variables that could impact the probability of a kill from a different aspect.
- The individual plays may not be independent of one another in this sample.