**Volleyball Statistics as Predictors of Win-Loss Percentage**

Heidi Rogers and Olivia Bojesen

**Introduction**

Within this study, we investigated how women’s volleyball team statistics are related to how often their team wins. Specifically, we posed the research question: which volleyball statistic best predicts a team’s win-loss percentage?

We gathered data on NCAA division I women's teams during the 2022 season as our sample for analysis (*NCAA College Women’s Volleyball DI Stats | NCAA.com*, 2018). This data was taken mid-season and the mean number of sets played by each team is 42 with a standard deviation of 6 sets. There were originally 344 teams in our dataset but due to unavailable data, we excluded nine of these teams from the analysis and the sample size we used was 335. The mean winning proportion for all 335 teams is 0.5108 with a standard deviation of .238.

**Methods**

The volleyball team data was analyzed in R 4.2.1. To begin answering our research question, we fit simple linear regression models for each team statistic with win-loss percentage to identify any relationships. Inspecting diagnostic plots for every predictor variable assisted us in distinguishing which team statistics fit our model best. We also looked at a forest plot and investigated which team statistic resulted in the greatest R-squared values to determine which has the most predictive power of win-loss percentage. These simple linear regression models were fit to the form:

Y=β0 + β1 X + ε

where Y is the response variable (win-loss percentage), X is the predictor variable (team statistic), β0 and β1 are the coefficients with β0 being the intercept of the regression line and β1 being the slope of the regression line, and ε is the error term. The team statistics that we used were aces, assists, solo blocks, block assists, total blocks, digs, kills, errors, total attacks, opponent kills, opponent errors, and opponent attacks. We used a significance level of < 0.05 throughout the analysis.

Additionally, we fit a multiple regression model to consider a more complex relationship between variables. For this model, we used four team statistics with the greatest R-squared values from our simple linear regression model together as the predictor variables for the winning proportion. This model was fit to the form:

Y=β0 + β1 + β2 + β3 + β4 + ε

**Results**

Overall, team statistics individually were not especially predictive of win-loss percentage. The team statistics that fit our simple linear model best were errors, assisted blocks, total blocks, and kills. Errors had an R-squared value of .158, assisted blocks had .1545, total blocks had .1472, and kills had .1307. However, the low R-squared values indicate that these variables have a weak relationship with win-loss percentage, and most of the variance in win-loss percentage is not explained by a team statistic on its own.

**Confidence intervals for each team statistic’s simple linear regression model**

Chart, scatter chart

Description automatically generated

The plot above shows the 95% confidence interval for each team statistic in simple linear regression where points on the graph represent the specific regression coefficient. The relationship between a variable and the win-loss percentage is not significant if the interval crosses the vertical line at 0. We can see that each of our top four variables – errors, block assists, total blocks, and kills – have confidence intervals that do not contain zero, and therefore they each have a relationship with win-loss percentage.

**Results of simple linear regression analysis of several NCAA Division I Women's Volleyball team statistics as predictors of winning percentage (n=335)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient of Determination (R2)** | **p-value** | **Regression Coefficient** |
| Errors\* | 0.158 | <0.001 | -0.002 |
| Kills\* | 0.131 | <0.001 | 0.001 |
| Total Blocks\* | 0.147 | <0.001 | 0.004 |
| Block Assists\* | 0.155 | <0.001 | 0.002 |
| Aces | 0.076 | <0.001 | 0.004 |
| Assists | 0.131 | <0.001 | 0.001 |
| Solo Blocks | 0.007 | 0.135 | 0.002 |
| Digs | 0.008 | 0.109 | 0.000 |
| Total Attacks | 0.002 | 0.384 | 0.000 |
| Opponent Kills | 0.100 | <0.001 | -0.001 |
| Opponent Errors | 0.052 | <0.001 | 0.001 |
| Opponent Attacks | 0.001 | 0.562 | 0.000 |

\*Best predictor variables

Team’s statistics and relationship to their corresponding win-loss percentage are shown in the plots below for our top four predictor variables. The regression line is overlayed with shading around the line showing the 95% confidence interval for winning percentage predictions based on our linear model:

![Chart, scatter chart

Description automatically generated]()

The relationships of winning percentage vs. block assists, total blocks, and kills are all linear and positive and there is significant evidence that these variables are predictive of winning percentage (p < 0.001 for each variable) and as they increase, winning percentage does as well. On the other hand, there is a negative linear relationship between winning percentage and errors. There is significant evidence that the number of errors is predictive of winning percentage (p < 0.001) and as the number of errors a team makes increases, their winning percentage decreases.

As previously stated, the R-squared values of these individual team statistics were low, and each statistic alone explained less than 16% of variance in winning percentage. Due to this, we fit a multiple regression model. We used the four variables with the highest R-squared values as the predictor variable and winning percentage still as the response variable.

Chart, scatter chart

Description automatically generated

There is significant evidence that these four variables together – block assists, total blocks, errors, and kills – are predictive of winning percentage (F4, 330 = 149.6, p < 0.0001). Additionally, this multiple regression model explains a larger proportion of the variance in winning percentage than simple linear regression (Adjusted R2 = 0.636), indicating a stronger relationship and is therefore a better model for predicting a team's win-loss percentage.

**References**

*NCAA College Women’s Volleyball DI Stats | NCAA.com*. (2018). NCAA.com. Web: 21 Sept. 2022. <https://www.ncaa.com/stats/volleyball-women/d1>