MIT SLOAN SPORTS ANALYTICS CONFERENCE presented by ESPN

42 ANALYTICS

A Starting Point for Predicting Team Success in Volleyball

Brian Xu and Scott Morey / The Kinkaid School / brianxu16@gmail.com

Introduction

- What happens if two perfect volleyball teams face each other? Both teams would hit every serve in, and they would side out against the other team with 100% efficiency, causing the set to continue forever as no team could get a two point lead. By thinking about volleyball in this way, it becomes clear that a team must side out more efficiently than their opponent in order to win. The only way for one of the perfect teams to win is by earning points when the opposing team is in serve receive.
- While side out percentage is the most important metric in volleyball, each skill in volleyball impacts a team's side out efficiency. How good a team is at passing, serving, and hitting and how good the opposing team is at serving, blocking, and digging affects how likely it is for a team to side out. We seek to determine the importance of each skill by identifying the skills that contribute the most to winning.
- Currently, there are very few conclusive results on the relative importance of each fundamental skill. The biggest problem researchers face is disassociating skills from each other. When a hitter gets a kill, it is hard to determine exactly how much the pass, set, or hit contributed to getting the kill. Similarly, the opposing teams defense will affect a hitter's kill percentage. Volleymetric analysts Corley Bagley and Brett Ware created a zero-to-ten rating system for each skill that enables teams to see what skills they are weak in. This helps individual teams focus on their weak spots but does not shed insight on what skills contribute the most to winning. Unfortunately, Bagley and Ware used proprietary data so we could not use their skill ratings systems when creating our model.
- We take inspiration from Dean's Four Factor equation which accurately predicted the number of games an NBA team would win in a season and identified the relative importance of each variable that made up the model. By developing an equation for volleyball similar to Dean's and comparing the coefficients of the variables in our equation, we can determine the skills most relevant to winning.
- By ranking skills in importance, coaches can develop more focused training sessions that will improve the skills most relevant to winning and evaluate players better.

Objectives

- Create a mathematical model that accurately predicts set win percentage by conducting regression analysis.
- Determine the most relevant factors for winning sets in volleyball and determine how much each one impacts winning.

Methodology

- We used R to perform min-max normalization on the data and to conduct all other statistical analysis.
- Created multivariable regression models using dependent variables that represented each skill.
- Discarded the variables with low coefficient t values until only statistically significant predictors remained.
- Picked the model with the lowest RMSE value and highest R^2 value.
- Checked linear regression assumptions for chosen model and compared coefficient confidence intervals.

Results

- We successfully found a multiple regression model that accurately predicts winning. The three key factors for winning in volleyball are Serve Error % Differential, Serve Ace % Differential, and Attack Efficiency Differential.
- All factors have large coefficient t values, indicating a relationship between the factor and winning exists. Our model also has a MAE value of 0.0284 which means the model is wrong by 2.84% on average.

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.084080 -0.027636 -0.001424  0.026089  0.095527

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.15189    0.01528   9.941  < 2e-16 ***
AcePctDiff    0.21276    0.03082   6.904 3.31e-10 ***
SErrPctDiff  -0.11778    0.01773  -6.642 1.19e-09 ***
AttackEffDiff 0.60355    0.02532  23.833  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 0.03608 on 111 degrees of freedom
Multiple R-squared:  0.9466,    Adjusted R-squared:  0.9451
F-statistic: 655.5 on 3 and 111 DF,  p-value: < 2.2e-16
```

Table 1: Summary of Model

- All linear regression assumptions are met. The residual plot shows little heteroscedasticity and normal probability plot shows the residuals are normally distributed. All factors appear linear and no significant collinearities exist.

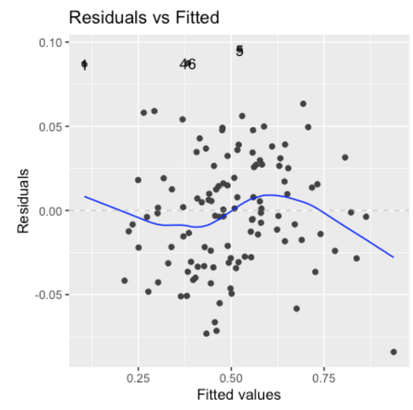

Figure 1: Residual Plot

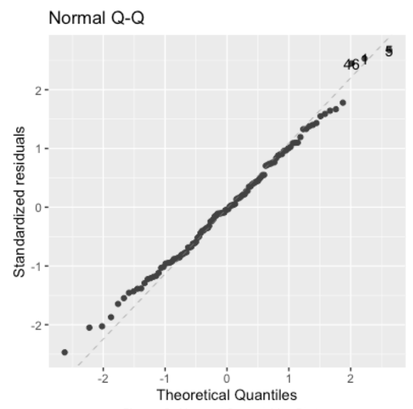

Figure 2: Normal Probability Plot

- Based on the coefficients, the weighting of SE% Diff, SA% Diff, and AttEff Diff is 22.7, 12.6, and 64.7. While we can't say this weighting is definitive because there may be interactions between the variables, the coefficient confidence intervals for each variable are distinct ranges, suggesting that AttEff Diff influences winning more than SE% Diff which influences winning more than SA% Diff. Our model accurately predicts set win percentage and explains 95% of the variation in wins.

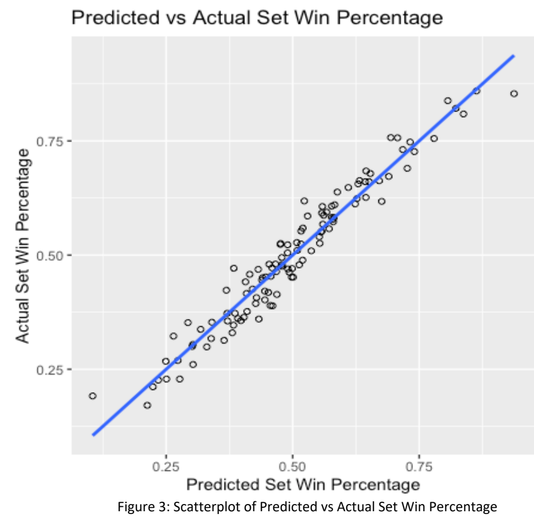

Figure 3: Scatterplot of Predicted vs Actual Set Win Percentage

Conclusion

- Serving the ball into the court is more important than getting an ace. Going back to the perfect teams scenario, this makes sense as teams tend to miss more serves than serve aces and missing your serve denies your team the opportunity to win points while the opponent team is in serve receive. Getting these points is important for lowering the opponent team's side out efficiency. Furthermore, missing your serve gives your opponents a free point.
- While we shed some insight on what skills contribute the most to winning, we were unable to definitively rank the skills in importance. For one, not all skills were included in the model as there was collinearity between Attack Efficiency Differential and statistics that measured receive and blocking. Likewise, too many factors (pass, set, block) influence Attack Efficiency Differential to draw any conclusions on how much more important hitting is than serving.