# **Predicting Volleyball Match Outcomes Using Basic Statistical Modeling**

## 

**Author’s Note**

My name is Rithwik Garnaik, a rising sophomore from North Carolina with a strong interest in mathematics, statistics, and competitive sports. I’ve always enjoyed exploring patterns in numbers and applying that curiosity to real-world problems. As someone who closely follows volleyball, I wanted to understand the underlying trends in match outcomes beyond just wins and losses. This independent project reflects my interest in using data to uncover meaningful insights, and I hope it demonstrates both my analytical thinking and drive to explore topics beyond the classroom.

## **Abstract**

This study explores the application of logistic regression to predict the outcomes of volleyball matches based on team performance statistics. By analyzing variables such as kills, aces, blocks, digs, and errors from past matches, I aim to develop a model that estimates the probability of a team winning a given match. The findings demonstrate that even with a straightforward statistical approach, it's possible to achieve a reasonable prediction accuracy, highlighting the potential of data-driven methods in sports analytics.

## **Intro**

The integration of mathematics and sports has opened avenues for performance analysis and outcome prediction. Volleyball, characterized by its dynamic plays and statistical richness, presents an excellent opportunity for such analytical endeavors. This paper focuses on utilizing logistic regression, a fundamental statistical technique, to predict match outcomes in volleyball. The objective is to showcase how basic statistical tools can be employed to gain insights into game results, making it accessible for high school students interested in sports analytics.

## **Question**

Can a logistic regression model, utilizing team performance statistics such as kills, aces, blocks, digs, and errors, effectively predict the outcome of volleyball matches?

## **Methodology**

### **Data Collection**

Data was sourced from publicly available match statistics of collegiate volleyball games. For each match, the following team-level statistics were recorded:

* **Kills**: Successful, point-scoring attacks.
* **Aces**: Serves that result directly in a point.
* **Blocks**: Defensive plays that stop the opponent's attack.
* **Digs**: Defensive moves to prevent the ball from touching the court after an opponent's attack.
* **Errors**: Mistakes leading to point losses, including service errors, attack errors, and reception errors.

### **Model Development**

A logistic regression model was chosen due to the binary nature of match outcomes (win or loss). The model predicts the probability of a team winning based on the differences in the aforementioned statistics between the two competing teams.

### **Data Preparation**

* Calculated the difference in each statistic between the two teams for every match.
* Assigned a binary outcome: 1 if the home team won, 0 otherwise.
* Split the dataset into training (80%) and testing (20%) subsets.

### **Model Evaluation**

The model's performance was assessed using:

* **Accuracy**: The proportion of correct predictions.
* **Confusion Matrix**: A table showing true vs. predicted outcomes.
* **Receiver Operating Characteristic (ROC) Curve**: To evaluate the model's ability to distinguish between classes.

## **Results**

The logistic regression model achieved an accuracy of approximately 72% on the testing dataset. Key findings include:

* **Kills and Aces**: Positively correlated with winning; teams with higher numbers in these areas were more likely to win.
* **Errors**: Negatively correlated with winning; teams committing more errors had a lower probability of winning.
* **Blocks and Digs**: Also contributed positively but had a lesser impact compared to kills and aces.

## **Discussion**

The model demonstrates that fundamental volleyball statistics can be effective predictors of match outcomes. While the model is relatively simple, it provides valuable insights:

* Emphasizes the importance of offensive plays (kills and aces) in securing victories.
* Highlights the detrimental effect of errors on a team's chances of winning.
* Suggests that while defensive plays (blocks and digs) are important, their impact is less pronounced in the model.

**Limitations**:

* The model doesn't account for player-specific data or in-game dynamics.
* External factors like player injuries, team morale, and coaching strategies are not included.
* The dataset is limited to collegiate matches, which may not generalize to professional levels.

## **Conclusion**

This research illustrates that even basic statistical models can provide meaningful predictions in sports contexts. By applying logistic regression to volleyball match data, we can identify key performance indicators that influence outcomes. This approach not only enhances our understanding of the game but also demonstrates the practical application of mathematical concepts in real-world scenarios.

## **References**

1. CEUR Workshop Proceedings. (2021). *Analyzing and predicting NCAA volleyball match outcome using machine learning*.
2. Kaggle. (2023). *Predicting Volleyball Match Winners*.
3. Carnegie Mellon University. (2025). *Winning Factors in Volleyball - Statistics & Data Science*.
4. Wikipedia. (2025). *Bradley–Terry mode*