CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Elias B. Gilbert

Project Number

J1404

Project Title

Watch the Curve: Using Statistical Modeling to Predict the Next Baseball Pitch

Abstract

Objectives/Goals

The objective was to create computer software that would use a statistical model to accurately predict the next pitch to be thrown in a baseball game.

Methods/Materials

I used data from Tim Lincecum and the statistical programming language R to create 9 models that each took different approaches to guessing pitches. The first guessed randomly among the pitches. The second always guessed the most common pitch thrown. The third guessed based on the proportions that each pitch was thrown. The 4th and 5th were the same as the previous two but they guessed based on the pitch before instead of the whole game. The 6th and 7th were the same as the 4th and 5th but guessed based on the pitch count. The 8th and 9th models guessed based on both the pitch before and the pitch count. I used a train/test split using data from one game to predict pitches in a different game. Then, I used different numbers of training games to improve the predictions.

Results

The first, and most basic model got the pitch correct 20% of the time. The best performing model, model 8, got the pitch correct 32% of the time with a 1/1 train/test split. Then, when I used data from 3 games to guess the pitches in another game (a 3/1 train-test split), the model got the pitch correct 37% of the time.

Conclusions/Discussion

Several things were learned from this project. First, games have predictability. My model was able to get the pitch correct 37% of the time. Also, I found that both the pitch count and the pitch before affect the next pitch to be thrown. Third, using a larger training dataset gives you a more accurate prediction of the next pitch. This model could continue to be made better by adding more factors, and a bigger training dataset.

Summary Statement

I used statistical modeling to use previous pitches to predict the next kind of pitch that a pitcher will throw in a baseball game.

Help Received

My dad, Greg Gilbert, gave me helpful discussions and for teaching me coding when I did not know how to do something. My mom, Ingrid Parker, helped me lay out my poster and edited my text.