Executive Summary: Just How Insane Was Linsanity?

Background:

Linsanity was the time period in February 2012 when then perennial bench player Jeremy Lin had a historic rise into mainstream basketball. He took a struggling Knicks squad into a 7 game winning streak including wins against the contending Lakers and an iconic game winner against the Raptors. He single-handedly became a hero to many people, some who barely cared about basketball at the time. Even then President Barack Obama commented on him. Many, including the authors of this summary, remember this period of basketball fondly, thus we decided to look back with an objective lens using statistical analysis. We decided on using logistic regression, z-scores, confidence, and prediction intervals to see how Lin stacked up to other point guards during the 2011-2012 season as well as focusing on a 9-game streak where Linsanity reached its peak. We then used machine learning models for methods of bagging and random forests to create further predictive models of Jeremy Lin's performance during this time period, as well as the performance of Lin's peers at the point guard position over a similar stretch of games.

Analysis:

For the analysis we focused on three major statistics: PER, Box Plus/Minus (as well as normal Plus/Minus), and Game Score. PER is a metric made by John Hollinger, an ESPN columnist, in his own words he describes it as a way to "sum up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance." Box plus minus (BPM) looks at a players impact on score and finally game score looks at productivity of the player,

Some of the main takeaways from doing analysis on the 2011 and 2012 year alone was that Lin was an above average player when it came to PER and BPM. We did a basic linear regression analysis with the two variables which included a confidence interval and prediction interval. When it came to the year long dataset Lin was better than most point guards but still remained within the confidence interval and prediction interval. Honing in on the 9 game stretch between February 2nd and February 19th 2012, we see Lin's true prowess. Lin led point guards (all for except one) when it came to game score. When doing initial graphs of this stretch Lin seems to be the outlier and was showing his dominance. When doing a regression analysis with Game Score and Plus Minus we see that though Lin crosses the confidence interval threshold, he barely misses out on the prediction interval threshold. This remains the same when looking at the z score of the two variables as well. The z variable shows how far away the player was from the average of the variables. With this data we see that though Lin was incredibly efficient during this time he was not an "anomaly". One player was, though, someone who played more efficiently than Lin during this time. That was Hall of Famer Tony Parker.

Upon running a bootstrap aggregating (bagging) model, using b=100 bagged decision trees, and metrics in per game averages to predict box plus/minus, we collected a respectable RMSE of 5.3376, however, our model did a very poor job of predicting box plus/minus.When then attempting to use random forests to predict box plus/minus, we collected a better RMSE of 5.202, but yet again failed to get any real predictive power in our model, especially in the case of Lin, in which both models were drastically wrong. The bagging and random forest models predicted -9.0 and -0.029, respectively. Lin actually averaged a box plus/minus of 9.6 over the period of Linsanity. Even when accounting for better hyperparameters in our random forest model, we still could not fully attain a successfully predictive model, however, we did manage an improved RMSE in every model, getting as low as 5.07.