2024 NBA Future Analytics Stars Coding Exercise

James Sheplock — December 4, 2023

Methodology

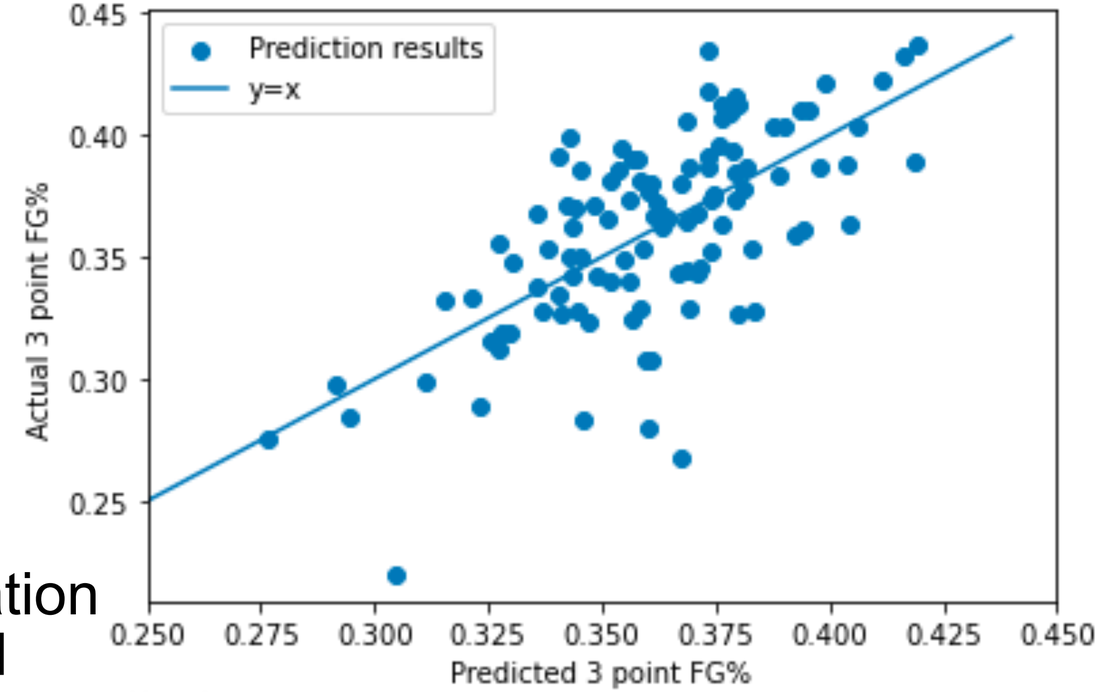
- Task: Predict each player's three-point percentage at the end of the 2022-23 season given their shooting statistics from October/November 2022
- Approach: Create a multiple linear regression model using the October/November shooting stats to predict the end-of-season three-point field goal percentage
- Result: Model using corner and non-corner three-point field goal percentage and free throw percentage

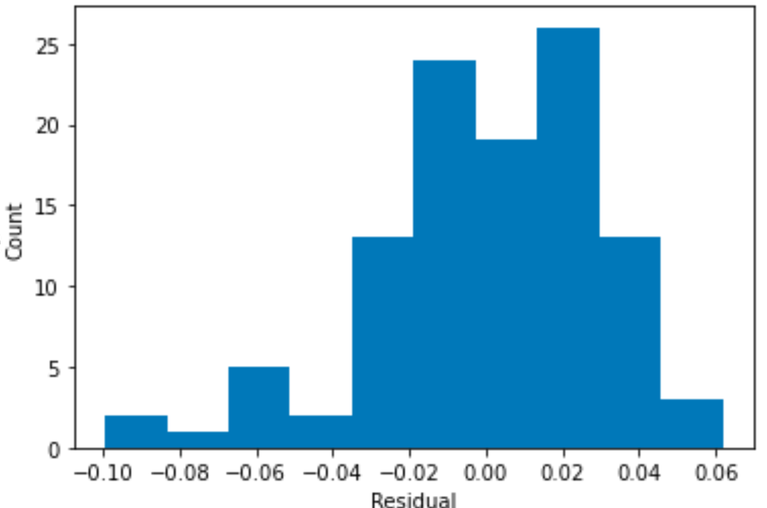
$$\text{ThreePctSeason} = 0.0882 + 0.134 \times \text{NonCornerThreePctOctNov} + 0.298 \times \text{CornerThreePctOctNov} + 0.147 \times \text{FreeThrowPctOctNov}$$

- Explanation: These variables had the most significant impacts on the model. Repeated F-tests did not favor rejecting this model by adding additional variables. The accuracy of the model did not significantly improve with adding any additional variables or using different combinations; this is the simplest and most accurate approach.
- Works Cited: Codecademy (model comparison - link) and FiveThirtyEight (skill score - link)

Prediction Accuracy

- Accuracy measured using root mean square errors: **2.9%**
- This is the average deviation between the predicted and actual end-of-season three point field goal percentage
- Plot to the right shows predicted vs. actual values, and histogram shows each residual
- This accuracy metric provides a direct interpretation for the results: given a predicted three point field goal percentage, how accurate can I expect that prediction to be?
  - The average difference between a prediction and the eventual end-of-season three point field goal percentage is less than 3 percentage points





Evaluating model skill

- In addition to RMS error, would like to evaluate model to verify it adds value over naive approaches
- Compare my model to 2 naive models using skill score based on mean square errors:
  - 1 - (model MSE) / (naive MSE)
  - Naive model 1: End of season stats = October/November stats
  - Naive model 2: End of season stats = Average player stats
- With positive scores of 0.52 and 0.45, I confirm my model improves on the naive approaches. The distance from 1 indicates room for improving this model.

Future Studies

- Given more time, I would like to study the possibility of predicting three point field goal percentage by separately predicting corner and non-corner percentages and number of corner and non-corner shots over the season. I did not have time to investigate this, but it strikes me as an obvious potential improvement worth looking at in the future.