

When do shooting percentages become representative of a player's true talent? An introductory analysis.

Chris Gauthier

2024-09-10

Introduction

In this document I will perform an analysis where I identify at what number of shot attempts shooting percentages start to become indicative of a player's true shooting talent.

This will include an analysis of RMSE values comparing the collective players' n-attempt samples with their season-long percentages.

I will analyze

- FG% vs 2PFG% vs 3PFG%
- <8 ft. vs 8-16 ft. vs. >24 ft.

TL;DR: ~75 shots is the rough cut off I have decided where we can start to rely on a shooting sample.

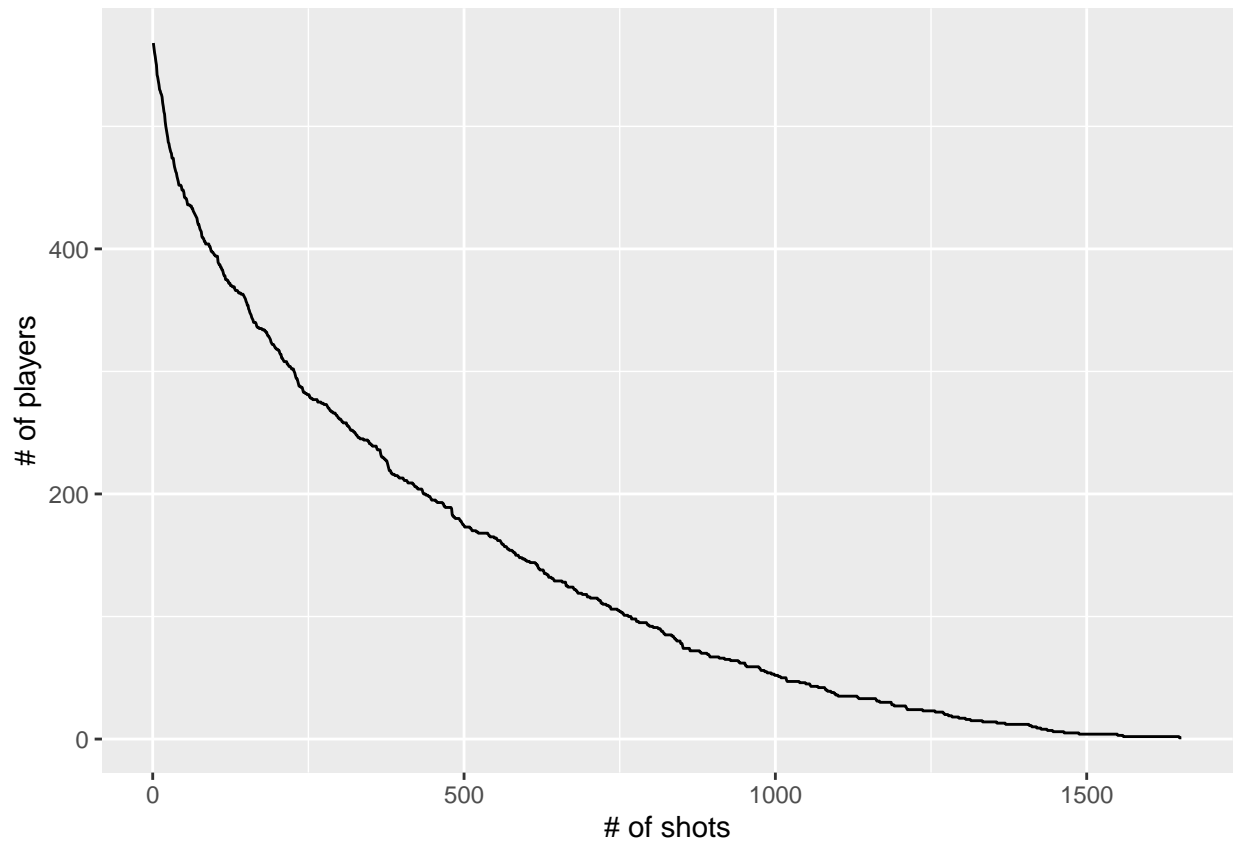
Methodology

The main function at play here is a random sampling function I created. After providing a minimum number of shot attempts I want to consider for a player pool, I take a random sample of every player's season shot total in increments of 10. I then took the Root Mean Squared Error (RMSE) for each random sample, regressing to the season shot total. Using RMSE here is just a way of quantifying how far/close we are to the player's true FG% for the particular subset we are looking at.

Why Root Mean Squared Error? I went with RMSE because, in a way, I am comparing a "model prediction" to the actual value. I am comparing what we think a player's shot making talent is at 10 shots, 20 shots, etc. compared to the player's total shots, what serves in this analysis as the player's "shooting talent." The behavior we see later on is consistent with RMSE, R-Squared, and even just taking a simple difference between the sample % and season-long %.

Deciding on a qualifying # of shot attempts

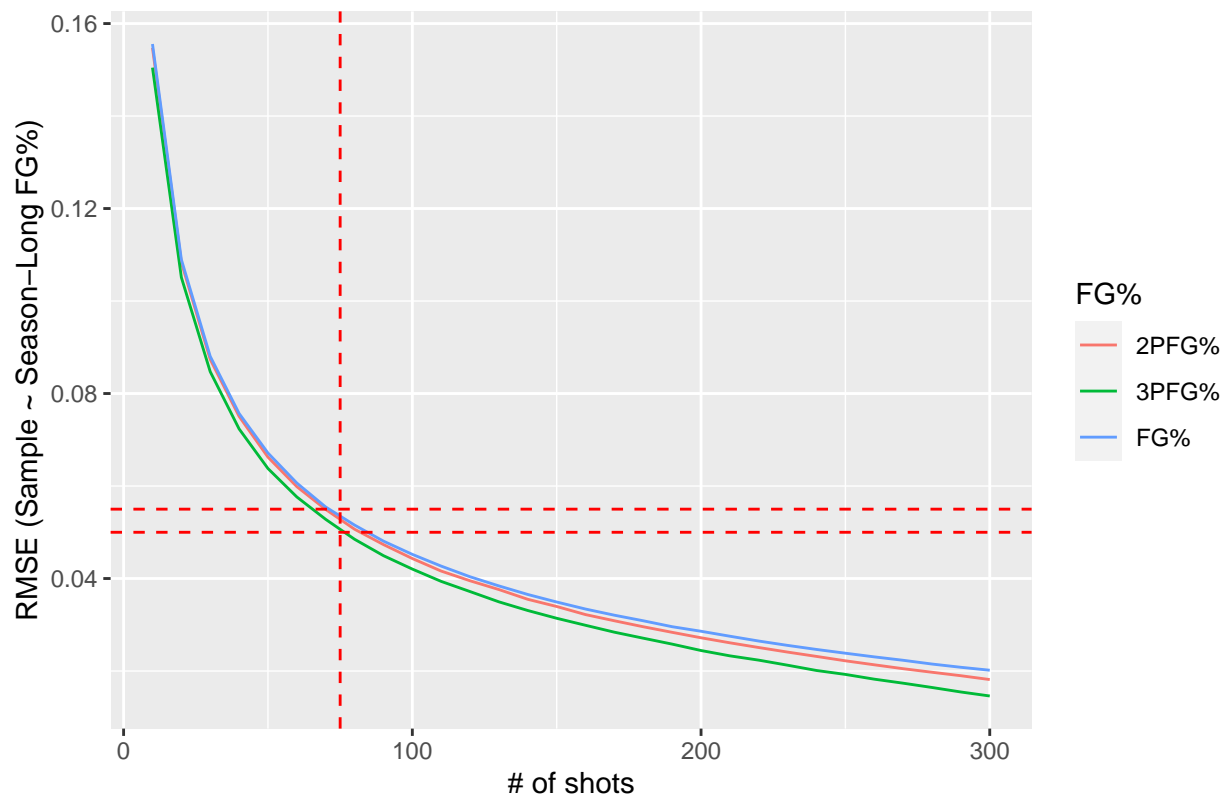
I chose the arbitrary value of 300 shot attempts as a cut off for the population of players I examined. Below is a chart showing the size of the player pool present at n-attempts. While I am only showing the chart for FG%, I reviewed this chart for all shooting percentage subsets I analyzed. While I could have gone with a shot requirement of 500 for the all-encompassing FG%, I opted for 300 as it was a more appropriate number for the smaller shot pools.



Results

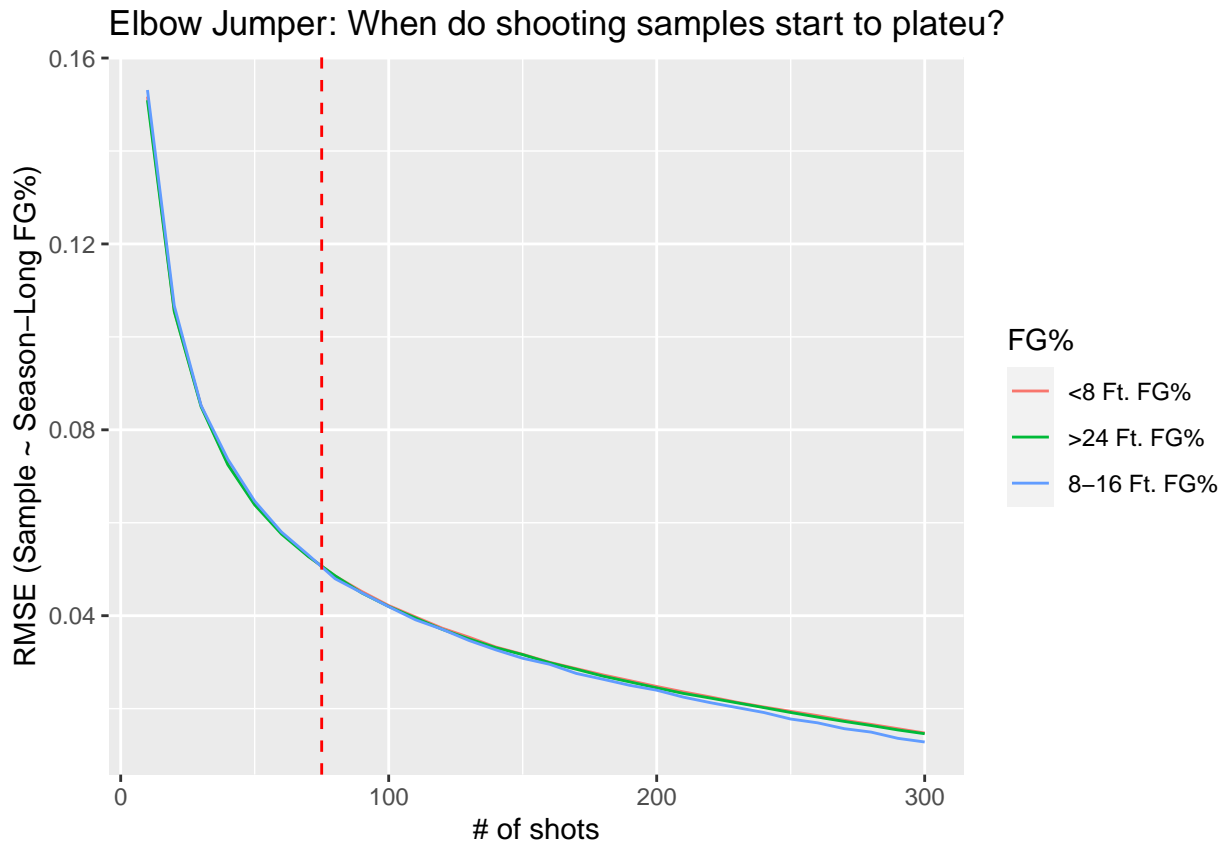
Below I have charted the “reliability curve” for FG%, 2PFG%, and 3PFG%:

Elbow Jumper: When do shooting samples start to plateau?



At the 75 shot cut off is around the “elbow point” where the decrease in error starts to slow down as the sample increases. The vertical line represents this 75-shot point, and the two horizontal lines represent the area in which all three of these curves intersect the vertical line, between 0.05 and 0.055 RMSE.

The other way I subset the shots of the 2023-24 NBA season was by distance in feet:



In this graph, all three FG%s have a similar threshold of 75 shots where they reach the elbow of their respective curve, at a RMSE of 0.05.

In this graph I have removed the 16-24 ft. subset as that group only had one player shoot ≥ 300 shots in this mid-range distance (DeMar DeRozan, 309).

Next Steps

I performed this analysis with very limited bandwidth. If I had more time, there are a number of ways I think I could have taken this analysis.

Ranges/confidence intervals I think this work could be layered on top of 2024-25 percentages that will appear early in the upcoming season. If Cole Anthony has shot ~200 shots, roughly split at ~150 2s and ~50 3s, could I provide a level of confidence for each of these three samples that we have arrived at Cole Anthony's true shooting talent?

Analyze other shooting metrics With more time I would have liked to take subsets by zone, look at the reliability of rim%, eFG%, etc.

Clean code practices With more time I would have been better about modularity within my code. While I have employed functions in my analysis, I had a lot of repeated code when I used these functions on different samples. There is probably a more efficient way about writing these functions where I could pass multiple subsets.

Conclusion

75 shots is a servicable arbitrary mark to reach for a shooting sample size where we can start to take a shooting percentage seriously. While there are probably more granular real cut offs for different percentages, zones, etc., there is a practicality of having a 75-shot catch-all threshold for quick assessments. Of course, this is not the be-all end-all, and not every shot is created equally. This 75-shot threshold is not meant to serve as law, but instead as a singular piece of information that serves as a soft guideline for decision making.