Regular Season vs Playoffs: Shooting Efficiency Analysis

STAT 107 Group 20

2025-11-08

Abstract

- This report examines whether NBA players' shooting efficiency changes between the regular season and the playoffs. Using player level data from the Oklahoma City Thunder, we focus on field goal percentage and three point percentage. We rely on a dedicated cleaning file to standardize variables and on a visualization file to produce distribution, paired comparison, and correlation plots. Preliminary visuals suggest modest declines in playoff efficiency. In the final phase we will run paired tests and report effect sizes with confidence intervals.

Introduction

- Playoffs often feature stronger defenses and higher pressure. These conditions may affect shooting outcomes. We ask whether field goal percentage and three point percentage differ between the two stages for the same players and whether regular season performance correlates with playoff performance.

Data and Cleaning

- The datasets are Regular Season.csv and Playoffs.csv. Cleaning and variable standardization are performed in the separate R Markdown file 11_DataCleaning.Rmd.
- Step 1: Load the datasets. First, both the regular-season and playoff datasets were imported into the workspace. After loading them, we checked the structure of each dataset to understand what variables they included, how the data were formatted, and whether there were any immediate issues. This gave us a baseline understanding of the raw data before making any changes.
- Step 2: Keep only the relevant columns. Next, we reduced each dataset to just the variables needed for the analysis. This step removed extra or unnecessary columns so that the remaining datasets contained only the player identity, team, playing time, scoring totals, and shooting statistics we planned to work with.
- Step 3: Rename columns for consistency. Because the column names in the two datasets were not fully consistent, all selected variables were renamed to follow the same naming style. This standardization ensured that both datasets used identical labels, which is important for combining them later and for avoiding confusion during analysis.
- Step 4: Combine the datasets. Finally, the cleaned regular-season and playoff datasets were merged into a single dataset. A new variable was added to label each observation as either "Regular" or "Playoffs," making it possible to compare player performance across the two contexts within one unified dataset.

Visualization

- We include the visualization plots that visualize the distribution of shooting efficiency, paired boxplots for player-level performance comparison, and correlation scatter plots between regular and playoff FG%. The -
- Distribution of Shooting Efficiency: The first set of plots shows histograms of field-goal percentage for the regular season and the playoffs. These visualizations allow us to compare how player shooting efficiency is spread out in each setting. By looking at the shapes of the two distributions, we can see whether players tend to shoot better, worse, or similarly in the postseason compared to the regular season.
- Paired Comparison of Performance: Next, we created a paired boxplot using only players who appear in both datasets. This visualization places regular-season and playoff field-goal percentages side by side for the same individuals. Comparing the two boxplots helps us understand whether there is a common shift in performance—such as a general decrease or increase in shooting efficiency—from the regular season to the playoffs.
- Correlation of Performance. Finally, we produced a scatterplot that compares each player's regular-season field-goal percentage to their playoff field-goal percentage. Each point represents one player, showing both values together. This plot helps reveal whether players who shoot well in the regular season also tend to shoot well in the playoffs, and how consistent shooting performance is across the two contexts.

Statistical Analysis

- Step 1: Identify playoff teams. We first loaded the datasets, then extracted a list of all teams that appear in the playoff dataset. This allows us to distinguish which teams made the playoffs and which did not, which is necessary for comparing team shooting performance between the two groups.
- Step 2: Aggregate regular-season data to the team level. Player-level statistics were summed to produce team-level totals for field-goals made and attempted, as well as three-pointers made and attempted. From these totals, team field-goal percentage and three-point percentage were calculated. This aggregation converts individual player performance into overall team shooting efficiency.
- Step 3: Classify teams as Playoff or Non-Playoff. Each team was labeled based on whether it appeared in the playoffs. This classification created two groups: "Playoff" and "Non-Playoff", for subsequent statistical comparisons. Only the relevant columns (team name, playoff status, and shooting percentages) were kept for clarity.
- Step 4: Organize data for comparison. We separated the shooting percentages into two groups for each metric: field goal percentage (FG%) and three point percentage (3P%). Each metric was split into values for playoff teams and non-playoff teams. This organization allows us to compare the average performance of playoff and non-playoff teams.
- Step 5: Compute Cohen's d. Cohen's d was calculated for FG% and 3P% to measure the effect size of the difference between playoff and non-playoff teams. Cohen's d quantifies how large the difference is in standard deviation units, giving context beyond just statistical significance.
- Step 6: Conduct t-tests. Unpaired t-tests (Welch's t-test, assuming unequal variances) were performed to determine whether differences in FG% and 3P% between playoff and non-playoff teams are statistically significant. This step provides formal evidence of whether playoff teams shoot more efficiently than non-playoff teams.
- Step 7: Summarize and save results. The results were summarized by printing the mean shooting percentages, t-test results, and Cohen's d for each metric. Finally, the team-level data with shooting percentages and playoff status were saved to a CSV file for reporting or further analysis.

Summary Statistics

- Group players by stage: We first organized the combined dataset by stage, separating players' statistics into two groups: regular season and playoffs. This grouping allows us to compute summary statistics for each stage separately, rather than mixing all players together.
- Calculate summary statistics: For each group, we calculated the following: The total number of players included (players_n), the average field-goal percentage (FG_mean) and its standard deviation (FG_sd), the average three-point percentage (TP_mean) and its standard deviation (TP_sd), these measures provide a concise overview of shooting performance and variability for each stage.
- Display results in a table: Finally, the summary statistics were presented in a formatted table within the report. This table allows us to quickly compare overall shooting efficiency and variability between the regular season and playoffs, making it easy to see whether shooting tends to improve, decline, or remain consistent across stages.