Regular Season vs Playoffs: Shooting Efficiency Analysis

STAT 107 Team 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 below.

STEP 1: Load datasets

```
regular <- read_csv("Regular Season.csv")
playoffs <- read_csv("Playoffs.csv")

# Check structure of both datasets
glimpse(regular)
```

```
## Rows: 28
## Columns: 14
## $ ...1   <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, ~
## $ Player <chr> "Shai Gilgeous-Alexander", "Jalen Williams", "Chet Holmgren", "~
## $ Team   <chr> "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", ~
## $ GP     <dbl> 76, 69, 32, 54, 57, 71, 76, 68, 74, 36, 47, 69, 54, 37, 78, 73,~
## $ W      <dbl> 63, 55, 26, 44, 49, 57, 62, 55, 63, 31, 40, 58, 45, 30, 47, 46,~
```

```
## $ L      <dbl> 13, 14, 6, 10, 8, 14, 14, 13, 11, 5, 7, 11, 9, 7, 31, 27, 28, 1~
## $ Min    <dbl> 34.2, 32.4, 27.4, 19.3, 27.9, 29.2, 22.9, 27.6, 21.7, 16.6, 16.~
## $ PTS    <dbl> 32.7, 21.6, 15.0, 7.1, 11.2, 10.1, 12.0, 8.4, 10.2, 6.5, 5.9, 6~
## $ FGM    <dbl> 11.3, 8.2, 5.2, 2.6, 4.9, 3.6, 4.7, 3.4, 3.5, 2.5, 2.1, 2.5, 0.~
## $ FGA    <dbl> 21.8, 16.9, 10.7, 5.8, 8.4, 8.4, 9.6, 7.2, 7.9, 5.1, 4.7, 5.1, ~
## $ `FG%`  <dbl> 51.9, 48.4, 49.0, 44.6, 58.1, 43.5, 48.8, 47.4, 44.0, 49.5, 43.~
## $ `3PM/` <dbl> 2.1, 1.8, 1.4, 1.1, 0.0, 2.4, 1.7, 1.1, 2.6, 0.6, 1.3, 1.0, 0.3~
## $ `3PA`  <dbl> 5.7, 4.9, 3.6, 3.1, 0.3, 5.8, 4.5, 3.1, 6.3, 1.7, 3.3, 2.5, 1.2~
## $ `3P%`  <dbl> 37.5, 36.5, 37.9, 35.3, 0.0, 41.2, 38.3, 35.6, 41.2, 38.3, 39.9~
```

```r
glimpse(playoffs)
```

```
## Rows: 28
## Columns: 14
## $ ...1   <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, ~
## $ Player <chr> "Shai Gilgeous-Alexander", "Jalen Williams", "Chet Holmgren", "~
## $ Team   <chr> "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", ~
## $ GP     <dbl> 23, 23, 23, 23, 23, 23, 22, 23, 21, 12, 17, 16, 10, 9, 23, 23, ~
## $ W      <dbl> 16, 16, 16, 16, 16, 16, 15, 16, 14, 8, 11, 12, 7, 6, 15, 15, 15~
## $ L      <dbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 4, 6, 4, 3, 3, 8, 8, 8, 8, 8, 8, 8, ~
## $ Min    <dbl> 37.0, 34.6, 29.8, 24.4, 22.4, 28.9, 13.8, 22.4, 10.0, 7.0, 8.3,~
## $ PTS    <dbl> 29.9, 21.4, 15.2, 9.2, 8.1, 7.9, 6.0, 5.6, 5.1, 3.4, 2.6, 2.4, ~
## $ FGM    <dbl> 10.1, 7.7, 5.3, 3.1, 3.6, 2.6, 2.1, 2.2, 1.7, 1.3, 0.9, 1.0, 0.~
## $ FGA    <dbl> 21.9, 17.2, 11.6, 7.0, 5.8, 7.1, 5.4, 5.2, 3.5, 2.9, 2.1, 2.5, ~
## $ `FG%`  <dbl> 46.2, 44.9, 46.2, 45.0, 61.9, 36.6, 39.5, 42.9, 49.3, 45.7, 42.~
## $ `3PM/` <dbl> 1.4, 1.5, 1.2, 1.6, 0.0, 2.1, 1.1, 0.9, 1.1, 0.4, 0.5, 0.3, 0.3~
## $ `3PA`  <dbl> 4.9, 5.0, 4.0, 3.9, 0.0, 6.1, 3.1, 2.7, 2.7, 1.1, 1.5, 1.3, 0.4~
## $ `3P%`  <dbl> 28.3, 30.4, 29.7, 41.1, 0.0, 34.3, 36.2, 32.3, 41.1, 38.5, 36.0~
```

Step 2: Keep only relevant columns

```r
regular_clean <- regular %>%
  select(Player, Team, GP, Min, PTS, FGM, FGA, `FG%`, `3PM/`, `3PA`, `3P%`)

playoffs_clean <- playoffs %>%
  select(Player, Team, GP, Min, PTS, FGM, FGA, `FG%`, `3PM/`, `3PA`, `3P%`)
```

STEP 3: Rename columns for consistency

```r
names(regular_clean) <- c("Player", "Team", "GP", "Minutes", "Points",
                    "FGM", "FGA", "FG_percent", "TPM", "TPA", "TP_percent")

names(playoffs_clean) <- c("Player", "Team", "GP", "Minutes", "Points",
                    "FGM", "FGA", "FG_percent", "TPM", "TPA", "TP_percent")
```

STEP 4: Combine datasets

```
combined <- dplyr::bind_rows(
  dplyr::mutate(regular_clean, type = "Regular"),
  dplyr::mutate(playoffs_clean, type = "Playoffs")
)
```

Visualization

- We include the visualization file that generates distribution histograms, paired box plots, and correlation scatter plots for FG percent. The functions used are defined in 02__funct__Plots.R and are invoked inside the visualization file.

Processing

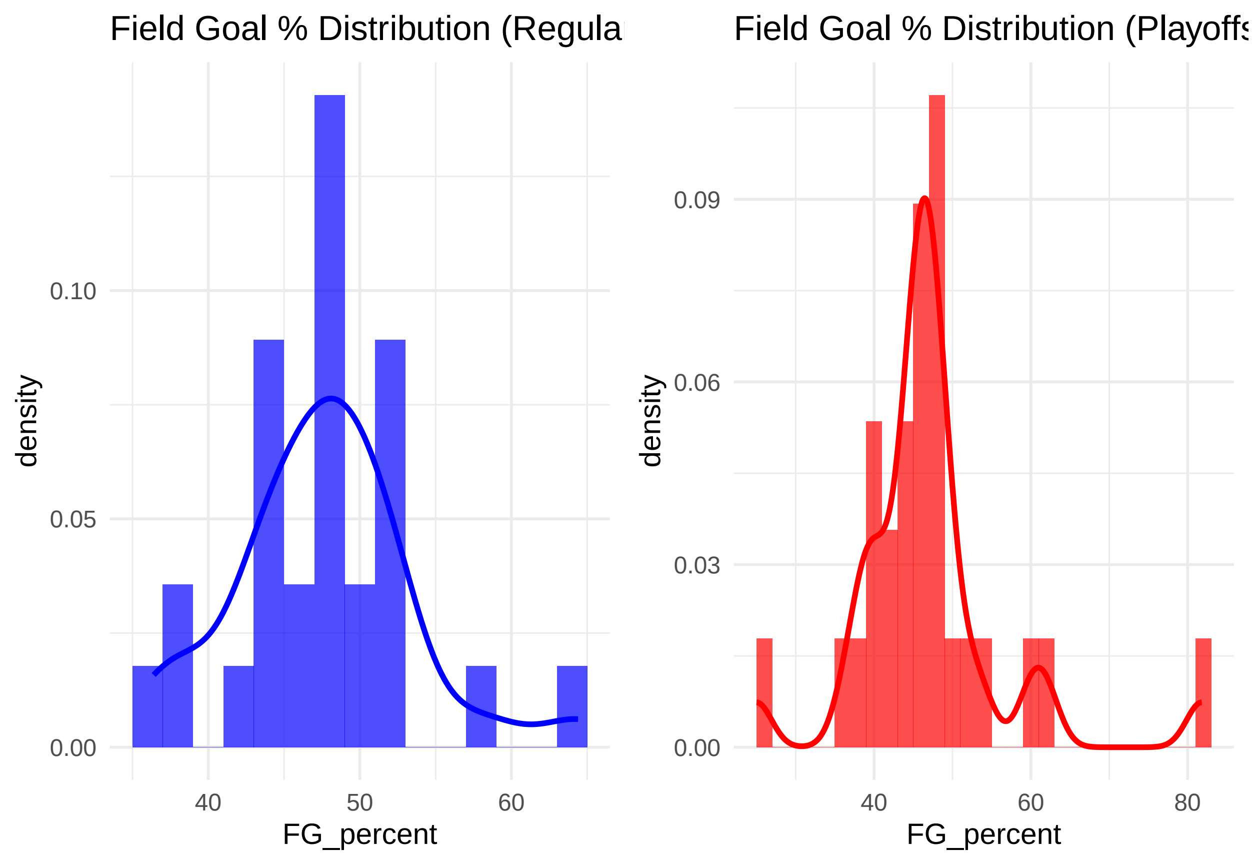
```
# 1. Load datasets
regular <- read_csv("Regular Season.csv")
playoffs <- read_csv("Playoffs.csv")

# 2. Keep only relevant columns
regular_clean <- regular %>%
  select(Player, Team, GP, Min, PTS, FGM, FGA, `FG%`, `3PM/`, `3PA`, `3P%`)

playoffs_clean <- playoffs %>%
  select(Player, Team, GP, Min, PTS, FGM, FGA, `FG%`, `3PM/`, `3PA`, `3P%`)

# 3. Rename columns for consistency
names(regular_clean) <- c("Player", "Team", "GP", "Minutes", "Points",
                          "FGM", "FGA", "FG_percent", "TPM", "TPA", "TP_percent")

names(playoffs_clean) <- c("Player", "Team", "GP", "Minutes", "Points",
                           "FGM", "FGA", "FG_percent", "TPM", "TPA", "TP_percent")

# 4. Process: Create a paired dataset
# We use an inner_join to keep only players who appear in BOTH datasets
paired_stats <- inner_join(
  regular_clean,
  playoffs_clean,
  by = "Player",
  suffix = c("_reg", "_playoff")
)

# Glimpse the final paired data
glimpse(paired_stats)
```

```
## Rows: 28
## Columns: 21
## $ Player             <chr> "Shai Gilgeous-Alexander", "Jalen Williams", "Chet ~
## $ Team_reg           <chr> "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "O~
```

```
## $ GP_reg             <dbl> 76, 69, 32, 54, 57, 71, 76, 68, 74, 36, 47, 69, 54,~
## $ Minutes_reg        <dbl> 34.2, 32.4, 27.4, 19.3, 27.9, 29.2, 22.9, 27.6, 21.~
## $ Points_reg         <dbl> 32.7, 21.6, 15.0, 7.1, 11.2, 10.1, 12.0, 8.4, 10.2,~
## $ FGM_reg            <dbl> 11.3, 8.2, 5.2, 2.6, 4.9, 3.6, 4.7, 3.4, 3.5, 2.5, ~
## $ FGA_reg            <dbl> 21.8, 16.9, 10.7, 5.8, 8.4, 8.4, 9.6, 7.2, 7.9, 5.1~
## $ FG_percent_reg     <dbl> 51.9, 48.4, 49.0, 44.6, 58.1, 43.5, 48.8, 47.4, 44.~
## $ TPM_reg            <dbl> 2.1, 1.8, 1.4, 1.1, 0.0, 2.4, 1.7, 1.1, 2.6, 0.6, 1~
## $ TPA_reg            <dbl> 5.7, 4.9, 3.6, 3.1, 0.3, 5.8, 4.5, 3.1, 6.3, 1.7, 3~
## $ TP_percent_reg     <dbl> 37.5, 36.5, 37.9, 35.3, 0.0, 41.2, 38.3, 35.6, 41.2~
## $ Team_playoff       <chr> "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "O~
## $ GP_playoff         <dbl> 23, 23, 23, 23, 23, 23, 22, 23, 21, 12, 17, 16, 10,~
## $ Minutes_playoff    <dbl> 37.0, 34.6, 29.8, 24.4, 22.4, 28.9, 13.8, 22.4, 10.~
## $ Points_playoff     <dbl> 29.9, 21.4, 15.2, 9.2, 8.1, 7.9, 6.0, 5.6, 5.1, 3.4~
## $ FGM_playoff        <dbl> 10.1, 7.7, 5.3, 3.1, 3.6, 2.6, 2.1, 2.2, 1.7, 1.3, ~
## $ FGA_playoff        <dbl> 21.9, 17.2, 11.6, 7.0, 5.8, 7.1, 5.4, 5.2, 3.5, 2.9~
## $ FG_percent_playoff <dbl> 46.2, 44.9, 46.2, 45.0, 61.9, 36.6, 39.5, 42.9, 49.~
## $ TPM_playoff        <dbl> 1.4, 1.5, 1.2, 1.6, 0.0, 2.1, 1.1, 0.9, 1.1, 0.4, 0~
## $ TPA_playoff        <dbl> 4.9, 5.0, 4.0, 3.9, 0.0, 6.1, 3.1, 2.7, 2.7, 1.1, 1~
## $ TP_percent_playoff <dbl> 28.3, 30.4, 29.7, 41.1, 0.0, 34.3, 36.2, 32.3, 41.1~
```

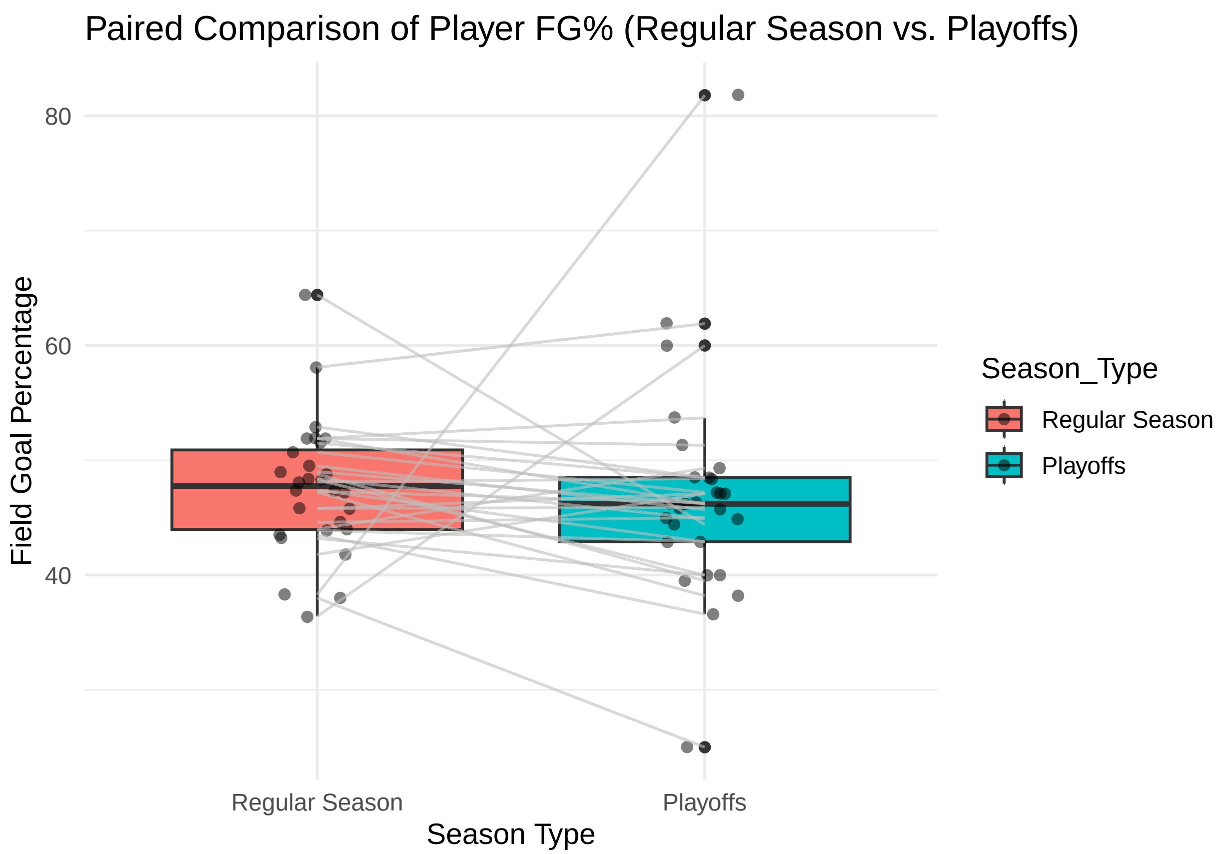
Distribution of Shooting Efficiency

```r
plot_distribution_histograms(
  regular_df = regular_clean,
  playoff_df = playoffs_clean,
  metric_col = "FG_percent",
  plot_title_prefix = "Field Goal %",
  bin_width = 2
)
```



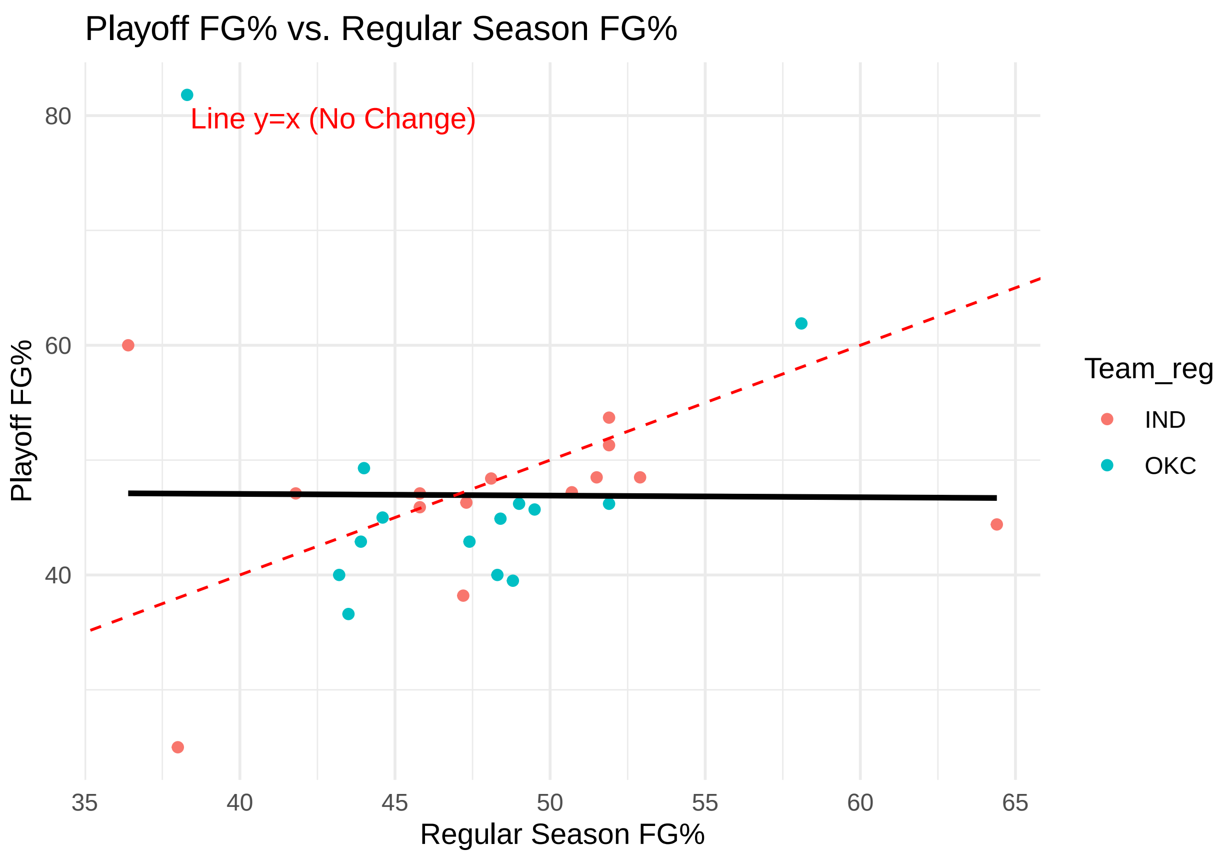
Paired Comparison of Performance

```
plot_paired_boxplot(
  paired_df = paired_stats,
  col_reg = "FG_percent_reg",
  col_playoff = "FG_percent_playoff",
  plot_title = "Paired Comparison of Player FG% (Regular Season vs. Playoffs)",
  y_label = "Field Goal Percentage"
)
```



Correlation of Performance

```
plot_correlation_scatter(
  paired_df = paired_stats,
  col_reg = "FG_percent_reg",
  col_playoff = "FG_percent_playoff",
  plot_title = "Playoff FG% vs. Regular Season FG%",
  x_label = "Regular Season FG%",
  y_label = "Playoff FG%"
)
```



Summary Statistics

- The table below summarizes group level statistics by stage to complement the visualizations.

```
desc <- combined %>%
  dplyr::group_by(type) %>%
  dplyr::summarise(
    players_n = dplyr::n(),
    FG_mean = mean(FG_percent, na.rm = TRUE),
    FG_sd = sd(FG_percent, na.rm = TRUE),
    TP_mean = mean(TP_percent, na.rm = TRUE),
    TP_sd = sd(TP_percent, na.rm = TRUE),
    .groups = "drop"
  )

knitr::kable(desc, caption = "Summary statistics for FG% and 3P% by stage")
```

Table 1: Summary statistics for FG% and 3P% by stage

type	players_n	FG_mean	FG_sd	TP_mean	TP_sd
Playoffs	28	46.94643	9.726767	33.58929	17.31765
Regular	28	47.59286	5.854971	33.52857	10.33211