41_Statistical_Analysis

2025-12-06

```r
# Data loading
df_reg <- readr::read_csv("Regular Season.csv")
```

```
## New names:
## Rows: 570 Columns: 15
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (2): Player, Team dbl (13): ...1, Age, GP, W, L, Min, PTS, FGM, FGA, FG%, 3PM/,
## 3PA, 3P%
## i Use ‘spec()‘ to retrieve the full column specification for this data. i
## Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## * ‘‘ -> ‘...1‘
```

```r
df_playoff <- readr::read_csv("Playoffs.csv")
```

```
## New names:
## Rows: 50 Columns: 15
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (2): Player, Team dbl (13): ...1, Age, GP, W, L, Min, PTS, FGM, FGA, FG%, 3PM/,
## 3PA, 3P%
## i Use ‘spec()‘ to retrieve the full column specification for this data. i
## Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## * ‘‘ -> ‘...1‘
```

```r
# Identify playoff teams
# Get a vector of unique team codes that appear in the playoff data
playoff_teams <- unique(df_playoff$Team)

# Aggregate regular season data to team level
df_reg_team <- df_reg %>%
  group_by(Team) %>%
  summarise(
    FGM_total = sum(FGM, na.rm = TRUE),
    FGA_total = sum(FGA, na.rm = TRUE),
    TPM_total = sum(`3PM/`, na.rm = TRUE),
    TPA_total = sum(`3PA`, na.rm = TRUE),
    .groups = 'drop'
  ) %>%
  mutate(
    # Calculate team FG%
    FG_percent_team = (FGM_total / FGA_total) * 100,
    # Calculate team 3P%
    TP_percent_team = ifelse(
      TPA_total > 0,
      (TPM_total / TPA_total) * 100,
      0
    )
  )

# Classify teams
df_comparison <- df_reg_team %>%
  mutate(
    Team_Status = ifelse(
      Team %in% playoff_teams,
      "Playoff",
      "Non-Playoff"
    )
  ) %>%
  select(Team, Team_Status, FG_percent_team, TP_percent_team)

assign("df_comparison", df_comparison, envir = .GlobalEnv)

# Prepare data for testing

# Separate the data into two groups for FG%
fg_playoff <- df_comparison %>% filter(Team_Status == "Playoff") %>% pull(FG_percent_team)
fg_non_playoff <- df_comparison %>% filter(Team_Status == "Non-Playoff") %>% pull(FG_percent_team)

# Separate the data into two groups for 3P%
tp_playoff <- df_comparison %>% filter(Team_Status == "Playoff") %>% pull(TP_percent_team)
tp_non_playoff <- df_comparison %>% filter(Team_Status == "Non-Playoff") %>% pull(TP_percent_team)

# Execute Statistical Comparisons (Unpaired T-tests)

# T-test for Field Goal Percentage (FG%)
# We use var.equal = FALSE for Welch's t-test, which assumes unequal variances.
t_test_fg <- t.test(fg_playoff, fg_non_playoff, var.equal = FALSE)

# T-test for Three Point Percentage (3P%)
t_test_tp <- t.test(tp_playoff, tp_non_playoff, var.equal = FALSE)

assign("t_test_fg_team", t_test_fg, envir = .GlobalEnv)
assign("t_test_tp_team", t_test_tp, envir = .GlobalEnv)

# Print Results

cat("--- Field Goal Percentage (FG%) Comparison ---\n")
```

```
## --- Field Goal Percentage (FG%) Comparison ---
```

```
cat("Playoff Mean (N=", length(fg_playoff), "): ", round(mean(fg_playoff), 2), "%\n", sep="")
```

```
## Playoff Mean (N=4): 44.73%
```

```r
cat("Non-Playoff Mean (N=", length(fg_non_playoff), "): ", round(mean(fg_non_playoff), 2), "%\n", sep="
```

```
## Non-Playoff Mean (N=26): 45.05%
```

```r
print(t_test_fg)
```

```
##
##  Welch Two Sample t-test
##
## data:  fg_playoff and fg_non_playoff
## t = -0.42853, df = 12.358, p-value = 0.6756
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.911506  1.281491
## sample estimates:
## mean of x mean of y
##  44.73128  45.04629
```

```r
cat("\n")
```

```r
cat("--- Three Point Percentage (3P%) Comparison ---\n")
```

```
## --- Three Point Percentage (3P%) Comparison ---
```

```r
cat("Playoff Mean (N=", length(tp_playoff), "): ", round(mean(tp_playoff), 2), "%\n", sep="")
```

```
## Playoff Mean (N=4): 33.29%
```

```r
cat("Non-Playoff Mean (N=", length(tp_non_playoff), "): ", round(mean(tp_non_playoff), 2), "%\n", sep="
```

```
## Non-Playoff Mean (N=26): 33.27%
```

```r
print(t_test_tp)
```

```
##
##  Welch Two Sample t-test
##
## data:  tp_playoff and tp_non_playoff
## t = 0.0089791, df = 3.6357, p-value = 0.9933
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.350305  5.383658
## sample estimates:
## mean of x mean of y
##  33.28588  33.26920
```