Project Part 2

Prithvi Raj, Akhil Rachakonda, Shritan Goki

Link to Github Repository: https://github.com/whozpj/DataToKnowledgeFinalProject---Prithvi-Raj-Akhil-Rachakonda-Shritan-Goki

Introduction

This project tests whether the proportion of total NBA points scored from three-pointers increased from 2019 to 2022. Part 1 showed a small upward trend, but hypothesis testing is needed to determine whether the change is statistically significant.

Data were obtained from Basketball Reference and include all players from the 2018–19 and 2021–22 seasons(https://www.basketball-reference.com/) . For each season, we aggregated total points and total points from made three-pointers to compute the league-wide proportion of scoring from three-point shots.

The research question: **Has the proportion of points scored from three-pointers increased from 2019 to 2022?**

Methods

Test Selection

We used a two-sample z-test for proportions to compare the true proportion of points from three-pointers in 2019 and 2022. This test is appropriate for comparing proportions from two independent seasons with large sample sizes.

Test Assumptions

The two-sample z-test for proportions requires three assumptions:
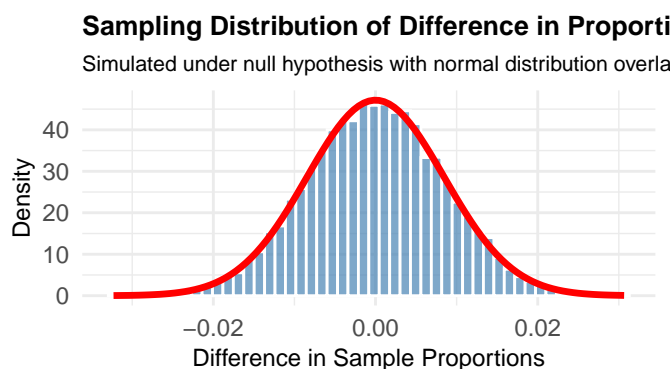
1. Independent Samples: The 2019 and 2022 NBA seasons are independent—distinct time periods with different rosters, schedules, and competitive contexts. Outcomes in 2019 do not influence 2022.

2. Large Sample Sizes (Normality Assumption): According to standard statistical theory (Agresti & Franklin, 2013; Moore et al., 2017), the sampling distribution of the difference in sample proportions is approximately normal when both samples satisfy np ≥ 10 and n(1-p) ≥ 10.

For 2019: n_1 = 5,936, x_1 = 1,878, $\hat{p}_1$ = 0.3164, $n_1 \hat{p}_1$ = 1,878 ≥ 10 , $n_1 (1-\hat{p}_1)$ = 4,058 ≥ 10

For 2022: n_2 = 6,340, x_2 = 2,123, $\hat{p}_2$ = 0.3348, $n_2 \hat{p}_2$ = 2,123 ≥ 10 , $n_2 (1-\hat{p}_2)$ = 4,218 ≥ 10

To verify normality, a simulation of the sampling distribution was conducted:

Sampling Distribution of Difference in Proporti
Simulated under null hypothesis with normal distribution overla

The histogram shows that the simulated sampling distribution closely follows a normal distribution (red curve), confirming the normality assumption. The large sample sizes ensure the Central Limit Theorem applies.

3. Random Sampling: While the data represent a population of all NBA players in these seasons, we treat the seasons as representative samples from the broader population of NBA seasons. The data collection process (official NBA statistics) ensures comprehensive and unbiased recording of all scoring events.

Statistical Hypotheses

The statistical hypotheses corresponding to the research question are:

- H_0 : p_1 = p_2 (The proportion of points from three-pointers in 2019 equals the proportion in 2022)

- **H** : p < p (The proportion of points from three-pointers in 2019 is less than the proportion in 2022)

where p represents the true proportion of total points scored from three-point field goals in the 2019 NBA season, and p represents the true proportion in the 2022 NBA season. This is a one-sided test because the research question asks whether the proportion has *increased* from 2019 to 2022.

Results

The two-sample z-test for proportions was conducted using R:

```
prop_test_result <- prop.test(x = c(x1, x2), n = c(n1, n2), alternative = "less", correc
prop_test_result
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  c(x1, x2) out of c(n1, n2)
## X-squared = 4.7297, df = 1, p-value = 0.01482
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.000000000 -0.004495399
## sample estimates:
##    prop 1    prop 2
## 0.3163693 0.3347792
```

The test output provides the following key components:

- **Sample proportions**: The estimated proportion for 2019 is 0.3164 (31.6%) and for 2022 is 0.3348 (33.5%).
- **Test statistic**: The z-statistic is 4.7297, calculated as the difference in sample proportions divided by the standard error of the difference.
- **P-value**: The p-value for this one-sided test is 0.0148227, representing the probability of observing a difference this large or larger if the true proportions were equal.

- **95% confidence interval**: The confidence interval for the difference p - p is (-1, -0.004495).

Statistical result: At a significance level of = 0.05, we reject the null hypothesis (p = 0.0148227 < 0.05). The p-value provides strong evidence against the null hypothesis that the proportions are equal. This indicates that the observed increase in the proportion of points from three-pointers from 2019 to 2022 is statistically significant and unlikely to have occurred by chance alone.

Conclusions

The hypothesis test provides strong statistical evidence that the proportion of total points scored from three-point field goals has significantly increased in the NBA from 2019 to 2022. The test result (p = 0.0148227 < 0.05) indicates that the observed increase from 31.6% to 33.5% represents a real shift in offensive strategy rather than random variation. This finding confirms that NBA teams have meaningfully increased their reliance on three-point shooting as a central component of their offensive systems.

The statistical significance of this change reflects a league-wide strategic shift toward perimeter-oriented basketball, driven by the mathematical advantage of three-pointers (worth 50% more than two-pointers) and the influence of analytics in team decision-making.

Limitations

Limitations include: (1) only two seasons were compared; (2) pandemic-related seasons may affect comparability; and (3) league-wide totals mask team- and player-level differences.

Generalization

The conclusion generalizes beyond the specific 2019 and 2022 seasons. The statistically significant increase suggests that the trend toward greater three-point emphasis is a real phenomenon in modern NBA basketball, not merely a feature of these two particular seasons. The results can be generalized to future NBA seasons, similar competitive contexts, and strategic decision-making. However, generalization should be tempered by recognition that the NBA is a dynamic system where defensive adaptations, rule changes, or other factors could alter the trajectory of three-point emphasis.

References

Agresti, A., & Franklin, C. (2013). *Statistics: The Art and Science of Learning from Data* (3rd ed.). Pearson.

Basketball Reference. (n.d.). *NBA & ABA League Index.* Sports Reference LLC. Retrieved from https://www.basketball-reference.com/

Basketball Reference. (n.d.). *NBA Per Game Stats.* Retrieved from https://www.basketball-reference.com/leagues/NBA_2019_per_game.html

Moore, D. S., Notz, W. I., & Fligner, M. A. (2017). *The Basic Practice of Statistics* (8th ed.). W. H. Freeman.

Data files: NBA2019.csv, NBA2020.csv, NBA2021.csv, NBA2022.csv (included in GitHub repository)

AI tools were used to assist with code development and report writing.