Analyzing Home Court Advantage in the National Basketball Association Through the Lens of Free Throw Statistics

By Len Huang

Abstract.

This project aims to study the effect of home court advantage in the National Basketball Association through a descriptive statistical analysis of free throw percentage, as well as free throws made per 100 possession. Being the home court team can be assumed to be advantageous, but in this study, we see how closely this assumption matches up with quantitative measurements of free throws through the use of effect size, more specifically, Cohen's d. Although some instances of data biased towards showing home court advantage, the magnitude of these instances were mostly negligible, showing the lack of influence home court advantage has on NBA teams.

I. Introduction

In the world of sports, home court advantage (HCA) is the concept that the team who is competing on their territory has an upper hand over the visiting team. Statisticians have long been skeptic of these sport ideas; this study will further explore that skepticism. [1] In recent years, HCA has been declining, claiming three main reasons: decreased referee dependency, improved travel conditions, and the increasing use of television as an alternative to seeing events live. With this, there are many reasons such as the support of local fans, familiarity with the competition area, and the lack of miscellaneous fatigue factors all contribute to this belief of HCA.

VI. Acknowledgements

Free throws are an example of a measurement for quantifying HCA. Often through distraction, or the lack thereof, fans can play a pivotal role in influencing the success of free throws. As such, we will analyze free throw statistics in this study. [2] Data taken from TeamRankings.com that has "a vast repository of sports stats and data, publishes it all on the site, and uses it to power proprietary power ratings systems and algorithmic models." From this website, I used their available data on free throws made per 100 possessions (FTMper100) and free throw percentage (FT%) . I will draw similarities between overall statistics as the control group, and the home and away statistics as treatments. When referring to the averages of either FTMper100 or FT%, I will be referring to the average of all 30 NBA teams for each year. All numbers are taken from the regular season in order to maximize the data pool and minimize bias seen in the postseason.

II. Data Collection

To collect the data, I needed, I had to find an efficient way to get statistics for home, away, and overall, without copying the other necessary information. In order to do so, I created a Python web-scraping program utilizing the BeautifulSoup library to extract the data needed. This data was put into CSV (comma separated value) files which I organized alphabetically by team, then converted into Google Spreadsheets so I could conduct analysis on them.

```
import bs4 as bs
import urllib
import csv

url = input("Enter the url: ")
filename = input("Enter the file name. Don't forget to add .csv at the end: ")

sauce = urllib.request.urlopen(url).read()
soup = bs.BeautifulSoup(sauce, 'lxml')

csv_file = open(filename, 'w')
csv_writer = csv.writer(csv_file, lineterminator = '\n')
csv_writer.writerow(['Team','Overall','Home','Away'])

x = []
y = []
for info in soup.find_all('td'):
    x.append(info.text)

for i in range(30):
    tup = (x[8*i+1], x[8*i+2], x[8*i+5], x[8*i+6])
    csv_writer.writerow([tup[0], tup[1], tup[2], tup[3]])

csv_file.close()
```

(Figure 1) Python program made by Len Huang. Interactive program that utilizes BeautifulSoup to get data from certain columns of a table.

III. Analysis Method

When analyzing the FTMper100, we use something called effect size, a quantitative measure of the magnitude of a phenomenon. [2] In our case, this means the effect size tells us how much of an impact being at home or away has on the mean number of FTMper100. Specifically, we will use *Cohen's d* in this study, a measure created by statistician and psychologist Jacob Cohen. This measurement is defined as the difference between to means divided by a standard deviation for the data.

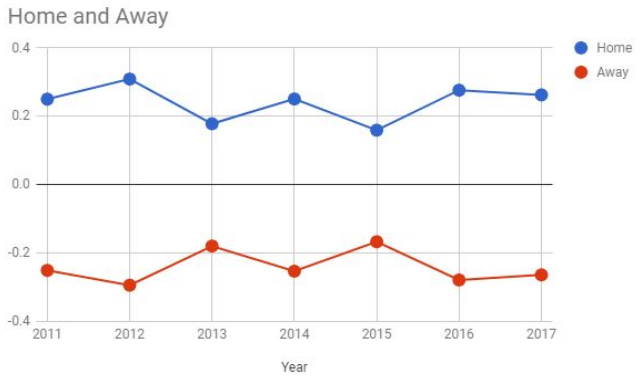
$$d = \frac{\bar{x}_1 - \bar{x}_2}{s} = \frac{\mu_1 - \mu_2}{s}.$$

(Figure 2) Cohen's d Formula

The statistic derived from this equation shows to what extent the averages of a treated group will influence a control group, and in what way: positively or negatively. For our purposes, A negative effect size would mean that the average FTMper100 is lowered by the "treatment" and a positive effect size would mean that the average is raised by the "treatment."

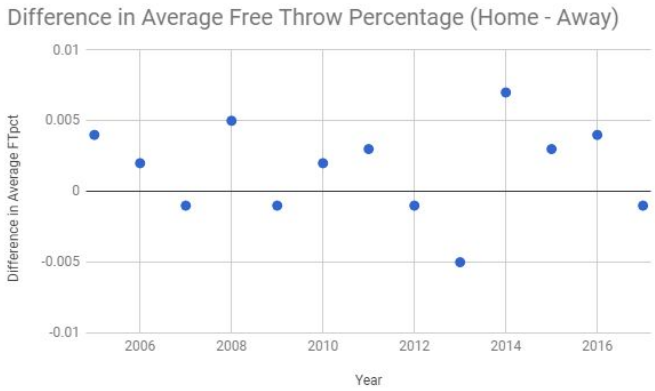
When analyzing the FT%, I took the average FT% of games played at home for a given year, and the the average of games played away for a given year. Then, I took the difference between these two averages (Home minus Away) to see whether or not FT% was higher at home or away, a factor contributing to HCA.

IV. Results



(Figure 3) Graph of Effect Size of Average FTMper100 from 2011-2017 Regular Seasons.

[3] This graph shows the progression of effect size of average free throws made per 100 possessions from the 2011 to 2017 regular season. From this graph, a trend arises: home stats have a positive effect size, while away stats have a negative effect size. However, the magnitude of the effect size stays relatively the same around the range of 0.1 to 0.3.



(Figure 4) Graph of Differences in Average FT% from 2005-2017 Regular Season

[4] This scatterplot shows the difference in average FT% for each season from the 2005 to 2017 regular season. There were eight years (2005, 2006, 2008, 2010, 2011, 2014, 2015, 2016) where the FT% at home was greater than that of away, but five years (2007, 2009, 2012, 2013, 2017) where the FT% at home was less than that of away. The magnitude of the difference in each year was less than 1%.

V. Discussion

As deduced from Figure 3, the relationship between positive and negative effect sizes with home and away treatments respectively shows that average FTMper100 is higher in home games and lower in away games when compared to all of the games as a whole. This relationship hints at HCA: for every 100 possessions, more free throws are being made by the home team, and less by the away team. However, because the effect size is generally low in magnitude (<0.3), we can assume that the treatments have a very small effect on FTMper100, and thus non-convincing evidence in favor of HCA.

The lack of evidence for HCA is further exhibited in Figure 4. Although there were more instances in which FT% at home was greater than that of away, there is noticeable pattern in these observations. Furthermore, the magnitude of these differences (less than 1%) are very small, which can be considered negligible. There is no relationship between FT% and its treatments, and there is not enough evidence to support that home games will always have greater FT% than away games. As such, there is non-convincing evidence in favor of HCA

In both Figure 3 and 4, there was a lack of convincing evidence in favor of HCA with regards to free throw statistics. As such, we see there is no distinct mathematical proof of HCA in the regular season of the NBA.

Given countless confounding and or lurking variables, there are many limitations to this study. However, there is still valuable insight gained from these numbers: while many people may think HCA is an influential factor in the NBA's regular season, the numbers show that it isn't.

VII. References

[1] Haberstroh, T. (2015, January 28). Home-court advantage? Not so much. Retrieved from http://www.espn.com/nba/story/_/id/12241619/home-court-advantage-decline

[2] Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hove: Lawrence Erlbaum Associates

[3] NBA Team FTM per 100 Possessions. (n.d.). Retrieved from https://www.teamrankings.com/nba/stat/ftm-per-100-possessions

[4] NBA Team Free Throw %. (n.d.). Retrieved from https://www.teamrankings.com/nba/stat/free-throw-pct

VII. Acknowledgements

The data was made available through teamrankings.com and the data collection was made possible through the online tutorials of Corey Schafer. Knowledge of the HTML needed to collect the data was supplemented by Mr. Arthur Davis of Holmdel High School. The whole paper was made possible through the assistance and guidance of Mr. John Ambrose of Holmdel High School.