**Probability and Applied Stats**

**Analysis of NBA Scoring Patterns of Over the Years**

**Table of Contents:**

**Introduction…………………………………………………………………………………………..….1**

The role statistics play in sports……………………………………………………………..…1

Thesis……………………………………………………………………………………………..1

Overview of the NBA dataset used………………………………….…………………………1

**Averages………………………………………………………………………………………………….1**

Early scoring averages, sample variance and standard deviation…………………………1

Style and pacing……………………………………………………………………………….…2

More serious athletes……………………………………………………………………………2

**Conditional Probabilities………………………………………………………………………………2**

Chances of winning with a certain score throughout the decades ………………………...2

Softer officiating………………………………………………………………………………….3

Players taking advantage……………………………………………………………………….3

**Countable Additivity Axiom…………………………………………………………………………...3**

Chances of scoring in certain range of points throughout the decades…………...……….3

Addition of 24-second shot clock……………………………………………………...……….3

Modern era……………………………………………………………………………….……….3

**Expected Value…………………………………………………………………………………………..4**

Scaling factor of 1.5……………………………………………………………………..……….4

Comparison to points scored averages……………………………….……………...……….4

Introduction of three-point shot………………………………………………………...……….4

Golden State Warriors dynasty…………...…………………………………………...……….4

**Chebyshev analysis…………………………………………………………………………………….5**

Range of points typically scored per decade………………………………………………….5

Gamma distribution showing scoring consistency…………...……………………..………..5

**Combinations and permutations……………………………………………………………………..5**

Unique scores in each decade……………………...………………………………………….5

Different ways to select and arrange 2 scores……………………………………….……….5

Offensive boom in the 70s, 80s, and 90s……………………...……………………..……….6

**Binomial Distribution…………………………………………………………………………………...6**

League expansion……………...………………………………………………………………..6

Chances of scoring over 100 points in each decade………………………………..……….7

**Conclusion…………………………………………………...…………………………………………..7**

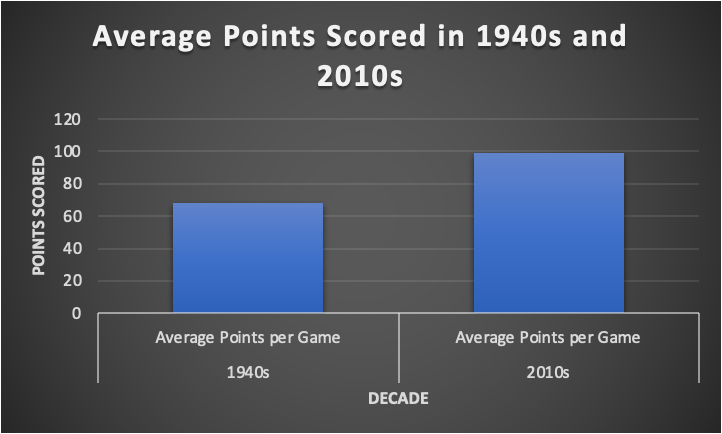
Restate calculations…………………………………………………………………….……….7

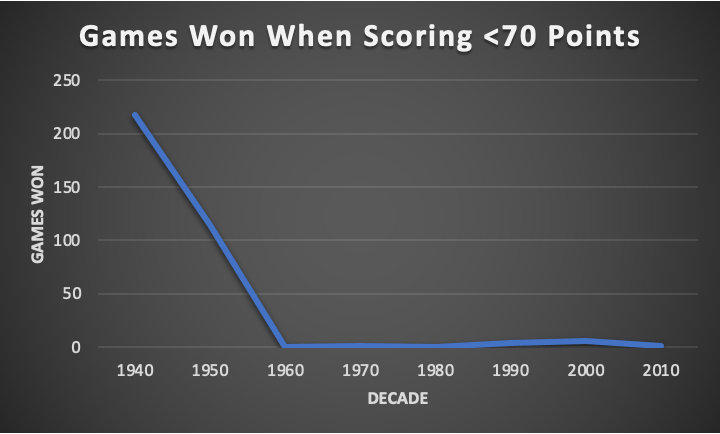
Restate NBA changes…………………………………………………………………..……….7

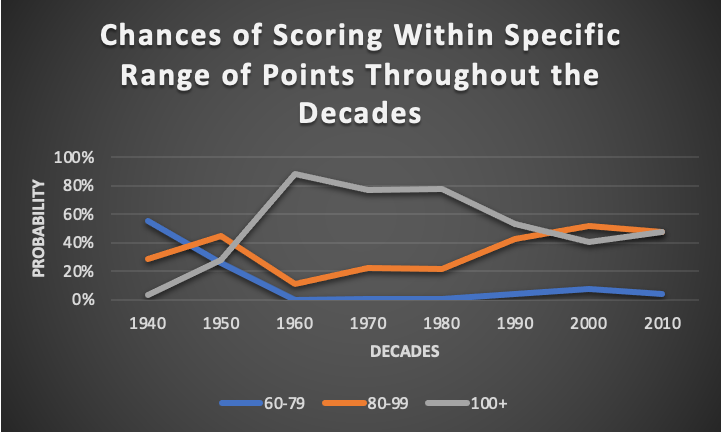
NBA today………………………………………………………………………………..……….8

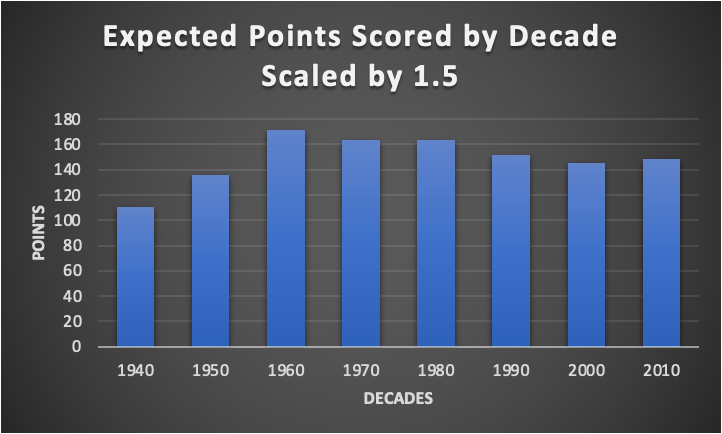
**Works Cited………………………………………………………………………………………………9**

Anyone who is a fan of sports can testify that occasionally, the announcers will pull some obscure statistic out of thin air that will make the fans wonder how they could even keep track of something like that. Heck even if they are not a typical New York Knicks fan, there are good odds that someone has heard some type of bizarre sports statistic that's made them stop and think for a minute. This is because when it comes down to it, all sports are broken down into many different statistical figures that have been tracked for many years. The National Basketball Association is no exception to this, and just like anything else the NBA and its players are constantly growing and improving. This is how we end up with these great statements by ex or current players claiming that “this record will never be broken” until a new young talent comes in the league and does the unthinkable. If you compare the number of points teams and players are scoring today against teams and players 20-30 years ago, anybody would easily be able to tell that the scoring numbers have shot way up over the years. Using a very large NBA dataset, this report will extensively compare and contrast various scoring patterns by teams from the years 1947-2015, as well as team performances and game outcomes that are able to provide insights into just how much the league has changed throughout the decades. The data used is from “nbaalleo.csv” created by FiveThirtyEight, a branch of ABC news that uses statistics to create stories about American society. In this case they have created a massive dataset that details every game's outcome, points scored by both teams, the date of the game, if it took place in the playoffs or not, along with a few other categories to give great detail on each NBA season.

To begin, let's take a look at the scoring trends in the first ever season in the NBA. On November 1, 1947, the first ever game was played in the National Basketball Association. The New York Knicks took down the Toronto Huskies in a 68 - 66 barnburner which would be considered either an offensive disaster or defensive game of the decade by today's standard. In 1947 the average points scored by a team was 67.92 points per game, with a sample variance of 146.40 and standard deviation of 12.10. These values show that this first season was on the lower scoring side of things. The sample variance is a larger value however, which shows that there was a wide range of scoring by teams whether it was higher or lower than 67.92 points. One possible reason for this large variance could be that in the earlier seasons in the NBA there were less teams, as well as games in a single season. This could cause the variance to be amplified because the sample size is smaller. When compared to the values taken from the 2010 to 2015 seasons it is easy to see the immediate difference in offensive production. Teams during this period averaged 99.16 points per game, with a sample variance of 142.18 and standard deviation of 11.92. As you can see the points scored by teams per game in this era has risen significantly by about 31 points. This is a very large jump, if a team were to lose a game by 31 points it would be considered a blowout. The sample variance and standard deviation only dropped a little bit indicating there remains a similar level of scoring dispersion even with the higher points averages. When observing the chart below it becomes clearer just how big of a gap there is between decades.This large jump in points is explainable through a plethora of different reasons. Games long ago were played completely differently in terms of style and pacing. Compared to today the game of basketball was moving at a snail's pace. Teams put emphasis on ball movement, set plays, and getting the ball to their slower “big men” who were much more dominant back then. Players 60 years ago compared to today were also not nearly as athletic, this is most likely due to the way that athletes train today compared to back then when the arenas were filled with cigarette smoke and the players were “partying” before, after, and even during the games. Athletes of today put millions of dollars a year into their bodies to ensure that they stay at their physical peaks for as long as possible. This dedication to their craft is what has allowed players such as Lebron James, Stephen Curry, Kevin Durant, and more and more players who seem to be able to play at a higher level for much longer than was previously thought possible.

Next, another statistic that shows a large difference in scoring throughout the decades is looking at the conditional probabilities of a team winning if they scored less than 70 points. In the 1940s there were 218 games won by teams that scored under 70 points that game. In all of the 2000s up until 2015 that had only happened 7 times. The chart below details how the chances of winning a game when scoring below 70 points drops to near zero after the 1950s. Something to also take note of is that 1947 was the first season, so the 40s “decade” only accounts for three years of data and there were almost 210 more occurrences of this back then. Surely such a large discrepancy must be caused by more than just the players getting better over the years, right? Correct, perhaps an even more prevalent reason of why such a large transformation has occurred over the years is due to the fact that the game is officiated completely differently than it was back then. In simple terms, the game back then was streetball compared to the way that athletes are protected on the court today. Fouls were called a lot less frequently allowing the defense to play much more physically in turn making it more difficult for the offense to score the basketball. These newer softer rules have been manipulated by NBA players with it really getting out of hand in more recent years by players such as James Harden. Many of today's points are scored at the foul line. The data above shows that if a team was to score below 70 points in the 1940s, they still had a 30% chance of winning that game. In the 2010s the chance of winning when scoring that little is about 1%.

To take it further, the Countable Additivity Axiom can be used to calculate the chances of teams scoring in a certain range of points to see how it has changed throughout the decades. The chances of a team scoring from 60-79 points in the 1940s was 55%, 80-99 points dropped to 29%, and their probability of scoring over 100 points is merely 3%. There is a massive change in these numbers as soon as the 1960s decade, where the chances of a team scoring from 60-79 points was 0.2%, 80-99 points increased to 11%, and their probability of scoring over 100 points is an impressive 88%. The chart above paints a picture of the discrepancy in scoring in these ranges over the years. There was a big boom in the 60s, 70s, and 80s until it began to stabilize near the 90s and 2000s. The reason for this can be pointed towards the massive rule change that was implemented on April 22, 1954. This date marked the day that the 24-second shot clock was added to the National Basketball Association. This is commonly referred to as the beginning of the modern era or shot clock era of the NBA. Most stats that are kept today only refer to years played after the shot clock was introduced, and for good reason. Now technically speaking, having less time to score the ball should make it harder to score and easier on the defense. However, with the new time limit in place for offensive possessions, this drastically increased the pace of the game which in turn led to more overall possessions for each team. This means more opportunities to score the ball leading to higher scoring games.

To continue, the expected value can be calculated using the dataset. With a scaling factor of 1.5 the expected value for a team's points in the 40s is about 110 points per game. Since it is already known that the average number of points scored in the 2010’s is about 99 points per game, this shows that it takes a scaling factor of 1.5 to bring the expected value of the teams in the 40s close to the average points in the 2010s. The chart below shows how these values differ throughout the decades, with a large rise coming again after the 1950s.There is one more concrete reason that can be looked to when trying to find the reason for the discrepancy. There is still one more huge change that the sport of basketball underwent in the 1979-80 season. Starting out as just a one-year trial, the National Basketball Association shook things up again by adding the three-point line. That is right, the perimeter arc that has made Stephen Curry one of the most famous NBA stars of all time has not always been a part of the NBA. This may be one of the more obvious reasons for change in point scoring, however if the dataset were to go even further into the future and made it to the 2020s it would show how the offensive scoring has jumped once again. Teams are scoring at all-time highs right now and it is all because of the three-point shot. When it was first introduced the three-point shot was not a large part of most players' game at that time. It took until the more recent years with teams like the Golden State Warriors who created a dynasty from 2015-2020 by becoming the most dangerous shooting team of all time. The three-point shot was such a threat for that team that they are often regarded as the greatest team of all time as in 2016 they broke the record for most wins in a season previously set by Michael Jordans Chicago Bulls in 1996. The Warriors had a record of 73-9 in the 2015-16 season and it was primarily led by the three-point basket.

Additionally, using a bound of 75%, the standard deviation, and mean from the decades the dataset can assist in calculating Chebyshev analysis of the decades. These calculations show that in the 1940s in 75% of games played the teams scored between 47.88 points and 99.75 points per game. The values drastically increase when looking at the 1970s where in 75% of games teams scored between 83.40 and 134.88 points. The 2000s were pretty similar numbers to the 70s, in 75% of games teams scored between 72.23 and 121.34 points per game.

| **Decade** | **Mean Score** | **Standard Deviation** | **Bounds (75% of games)** |
| --- | --- | --- | --- |
| 1940s | 73.81 | 12.97 | 47.88 to 99.75 |
| 1950s | 90.29 | 15.94 | 58.40 to 122.17 |
| 1960s | 114.25 | 12.56 | 89.13 to 139.37 |
| 1970s | 109.14 | 12.87 | 83.40 to 134.88 |
| 1980s | 109.17 | 12.5 | 84.16 to 134.18 |
| 1990s | 101.04 | 13.18 | 74.68 to 127.40 |
| 2000s | 96.78 | 12.28 | 72.23 to 121.34 |
| 2010s | 99.16 | 11.92 | 75.31 to 123.01 |

These numbers above again reflect the significant rise in offensive production that began in the 1960s and stabilized since then. Another statistic that backs this up is the gamma distribution through the decades. When looking at the shape of the points for the 1940s and 50s they are both about 32. On the other hand, when looking at the gamma distribution of any decade after the 50s they are all higher values ranging from 58.63 - 76.45. When it comes to gamma distribution higher values mean a more symmetric distribution, while lower values mean that the spread is larger and more skewed. Thus, backing up that the seasons after the 1960s held a more consistent offensive production.

Expanding upon this, taking an insight into the NBA’s scoring over the decades by means of combinations and permutations is a great way of researching the scoring patterns of the NBA decades. For example, in the 1940s and 50s scoring was less being done at a lower mark. This is shown in 1940 where there are only 74 unique scores, so there are 2701 different ways to select 2 scores, and 5402 ways those 2 scores could be arranged. However, in the 70s, 80s, and 90s, offenses were booming with power houses such as Kareem Abdul Jabbar, Larry Bird, Magic Johnson, and Michael Jordan playing in these decades. This is backed by the data taken from 1970 where there are 105 unique scores. This causes the number of ways to select and arrange 2 scores from this number to skyrocket. In 1970 there were 5460 different combinations of 2 scores, and 10920 different permutations of 2 scores. This shows the large scoring increase after the 1950 seasons. These numbers are shown on the table below for deeper clarification.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Decade** | **n (unique scores)** | **r** | **Combinations (C)** | **Permutations (P)** |
| 1940 | 74 | 2 | 2701 | 5402 |
| 1940 | 74 | 3 | 64824 | 388944 |
| 1940 | 74 | 5 | 16108764 | 1933051680 |
| 1950 | 102 | 2 | 5151 | 10302 |
| 1950 | 102 | 3 | 171700 | 1030200 |
| 1950 | 102 | 5 | 83291670 | 9995000400 |
| 1960 | 88 | 2 | 3828 | 7656 |
| 1960 | 88 | 3 | 109736 | 658416 |
| 1960 | 88 | 5 | 39175752 | 4701090240 |
| 1970 | 105 | 2 | 5460 | 10920 |
| 1970 | 105 | 3 | 187460 | 1124760 |
| 1970 | 105 | 5 | 96560646 | 1.16E+10 |
| 1980 | 98 | 2 | 4753 | 9506 |
| 1980 | 98 | 3 | 152096 | 912576 |
| 1980 | 98 | 5 | 67910864 | 8149303680 |
| 1990 | 108 | 2 | 5778 | 11556 |
| 1990 | 108 | 3 | 204156 | 1224936 |
| 1990 | 108 | 5 | 111469176 | 1.34E+10 |
| 2000 | 99 | 2 | 4851 | 9702 |
| 2000 | 99 | 3 | 156849 | 941094 |
| 2000 | 99 | 5 | 71523144 | 8582777280 |
| 2010 | 94 | 2 | 4371 | 8742 |
| 2010 | 94 | 3 | 134044 | 804264 |
| 2010 | 94 | 5 | 54891018 | 6586922160 |

Finally, another reason for the league being in such a better spot now than when it began is that the league continually expanded throughout the years, with more talks of expanding even today. This is prevalent when finding the binomial distribution for points scored by decade. In order to find this number, first the total number of games in the decade must be calculated. After doing this, the data immediately shows that the total number of games has continuously increased throughout the decades. This is due to the NBA constantly expanding its roster of teams and adding more games to the schedule. In these calculations, lies the probability of teams in each decade's chances of scoring more than 100 points.

|  |  |  |  |
| --- | --- | --- | --- |
| **Decade** | **n (Total Games)** | **p (Probability of >100 Points)** | **Total Probability** |
| 1940 | 1890 | 0.02910053 | 5.75E-25 |
| 1950 | 7190 | 0.25841446 | 2.35E-16 |
| 1960 | 10236 | 0.86694021 | 1.29E-24 |
| 1970 | 22426 | 0.74583073 | 2.09E-55 |
| 1980 | 20180 | 0.74821606 | 5.89E-55 |
| 1990 | 23300 | 0.50051502 | 0.00516285 |
| 2000 | 25810 | 0.37388609 | 3.32E-19 |
| 2010 | 15282 | 0.44529512 | 9.24E-31 |

In the 1950s the probability of a team scoring more than 100 points in a game was 25.84%. This is an increase from just 2.91% in the 40s, which is likely due to the introduction of the shot clock in 1954. However, beginning in the 60s teams jumped to an 87% chance of scoring over 100 points.

As you can see, the National Basketball Association has improved significantly in terms of offensive output, as shown by the analysis of scoring patterns throughout the decades since 1947. Statistics such as the average number of points teams scored in a single game in each decade and the variance of these values provide evidence of the consistent increase in scoring. Next, the conditional probability of teams winning a game if they scored less than 70 points showed that this number was good enough to secure a win back in the early days of the league. However, when looking at the data for the 2000s seasons, it became obvious 70 points would not be nearly enough to win in the modern NBA era with it only occurring 7 times in 15 years. Again, the offensive increase was supported by more statistics that derived from the Countable Additivity Axiom. Using this formula combined with the data supplied, the probabilities of a team scoring in certain ranges of points were calculated for each decade. These figures laid out that the probabilities of a team scoring over 100 points in a game rose significantly after the 1950s decade. It was deduced that this rise was more than likely a result of the addition of the 24-second shot clock in 1954. Then, using a scaling factor of 1.5, the expected value was found for the 1940s seasons to be only about 10 points higher than the average points scored by a team in the 2010s with no scaling factor. Additionally, Chebyshev analysis, combination and permutations, gamma and binomial distribution formulas were used to further the evidence of the NBA’s Offensive revolution since 1940. The direct cause of this change is hard to pinpoint as the league has changed immensely since it began in 1947. Nevertheless, there are certain rule changes such as the inclusion of the 24-second shot clock and the three-point line that can certainly help to provide explanation. Aside from just changing the rules, fans could also look to the expanded roster of the NBA with athletes taking their crafts more seriously nowadays compared to the time of the genesis of the league. All these reasons, combined with the fact that officiating in today's game makes it more difficult for teams to play hard and physical defense, contribute to the NBA becoming the offense heavy league it is today. These changes are seemingly not reverting anytime soon as teams today are scoring more points than ever, and it only seems to rise with every season as players figure out new exciting and creative ways to manipulate the games rules and put the ball in the basket.

Works Cited (APA)

Fivethirtyeight. “Data/NBA-ELO/Nbaallelo.Csv at Master · Fivethirtyeight/Data.” *GitHub*, 2015, github.com/fivethirtyeight/data/blob/master/nba-elo/nbaallelo.csv.