This is a story. The story of one person obsessed with numbers and sports. He wants to use those numbers to explain what happens in sports. In this episode, the person, (me) wants to examine professional sports in North America. I want to understand throughout sports history, what teams are considered the best regular season team over the history of the league by use of statistical analysis. I also think it is important to indicate the worst team in the history of the sport. To close off the circle, I also want to determine the most average of teams in the history of a particular sport. I will go into more detail after the introduction.
So, there have been many, many teams from each league as well as other leagues from one sport. Some lasted one season and others are still in existence. So, what is the criterion to determine what teams are eligible to be considered?

In Statistics there are three measures of central tendency, the ***mean***, the ***median*** and the ***mode***. The mode is the most frequently occurring number in a dataset. This measure will not tell which team is eligible or not and I consider that the data is Unimodal, so we do not use this measure for these sports data.
When it's unique, the mode is the value that appears the most often in a data set and it can be used as a measure of central tendency, like the median and mean. But sometimes, there is no mode or there is more than one mode.

There is no mode when all observed values appear the same number of times in a data set. There is more than one mode when the highest frequency was observed for more than one value in a data set. In both of these cases, the mode can't be used to locate the centre of the distribution.

The mode can be used to summarize categorical variables, while the mean and median can be calculated only for numeric variables. This is the main advantage of the mode as a measure of central tendency. It's also useful for discrete variables and for continuous variables when they are expressed as intervals.
Example: During a hockey tournament, Audrey scored 7, 5, 0, 7, 8, 5, 5, 4, 1 and 5 points in 10 games. After summarizing the data in a frequency table, you can easily see that the mode is 5 because this value appears the most often in the data set (4 times). The mode can be considered a measure of central tendency for this data set because it's unique.[1]
(https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch11/mode/5214873-eng.htm)

Now, the next measure to consider is the median. This is where a data set is divided into two where there are 50% of the data below a middle number and 50% are above this number. This measure is used in census data when looking at income. Income can vary from people who are technically defined as poor, an income below the poverty level and people who are billionaires. By using the median, the data is divided into two, by use of a middle number and this middle number would be considered the middle and any number above this would be considered eligible.

Example: Imagine that a top athlete in a typical 200-metre training session runs in the following times: 26.1 seconds, 25.6 seconds, 25.7 seconds, 25.2 seconds, 25.0 seconds, 27.8 seconds and 24.1 seconds. How would you calculate his median time?

Rank associated with each value of 200-meter running times
Table summary
This table displays the results of Rank associated with each value of 200-meter running times. The information is grouped by Rank (appearing as row headers), Times (in seconds) (appearing as column headers).

Rank	Times (in seconds)
1	24.1
2	25.0
3	25.2
4	25.6
5	25.7
6	26.1
7	27.8

There are n = 7 data points, which is an uneven number. The median will be the value of the data points of rank
(n + 1) ÷ 2 = (7 + 1) ÷ 2 = 4.
The median time is 25.6 seconds.[2]

The third measure is the mean, (average). This is the one that most people are aware of by calculating their academic average in school. The grades of courses in one semester are added up and then divided by the total number of courses. The resulting number is the mathematical average of the grades. However, there is one problem with the mathematical average. If there are many low numbers in the dataset, this will result in an average being lowered as it is pulled down because of the low numbers. The reverse is also true; if there are many large numbers this will pull the average higher and produce an erroneous result.

Example: Mount Royal hosts a soccer tournament each year. This season, in 10 games, the lead scorer for the home team scored 7, 5, 0, 7, 8, 5, 5, 4, 1 and 5 goals. What is the mean score of this player?
The sum of all values is 47 and there are 10 values. Therefore, the mean is 47 ÷ 10 = 4.7 goals per game.[2] (https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch11/mean-moyenne/5214871-eng.htm)

In terms of the sports teams, imagine teams with a 5 year existence analyzed with teams who have been around for 100 hundred years.

As for how many teams to be considered for the criteria of the statistical analysis, I will use the criteria of 30 seasons. A sample size of 30 is considered to be representative of the entire population, so any team that has played a minimum of 30 seasons will be considered representative of the entire league.[3] The ***central limit theorem*** states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed. This will hold true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large (usually $n \geq 30$). In using the theorem there will be at least a minimum of 30 seasons that represents all seasons.

Z-Score Explained

Unfortunately, most people are not familiar with statistics. So, to explain what was done to the data, an explanation of the statistics behind the answers needs to be presented.

First, there are terms that the reader should understand, don't worry there will not be a test at the end of the chapter.

Key Terms: ***Normal distribution***: a bell-shaped, symmetrical distribution in which the mean, median and mode are all equal
Z scores (also known as standard scores): the number of standard deviations that a given raw score falls above or below the mean
Standard normal distribution: a normal distribution represented in z scores. The standard normal distribution always has a mean of zero and a standard deviation of one.
Standard Deviation: A standard deviation (or σ) is a measure of how dispersed the data is in relation to the mean. Low standard deviation means data are clustered around the mean, and high standard deviation indicates data are more spread out. A standard deviation close to zero indicates that data points are close to the mean, whereas a high or low standard deviation indicates data points are respectively above or below the mean.[3]
https://www.nlm.nih.gov/nichsr/stats_tutorial/section2/mod8_sd.html

Overview
What is a distribution? A distribution is an arrangement of values of a variable displaying their observed or theoretical frequency of occurrence. A bell curve showing how the class did on the last exam would be an example of a distribution. The following four dimensions can characterize all distributions:

1. ***Central Tendency***—what are the mean, median and mode(s) of the distribution?

2. ***Variability***—all distributions have a variance and standard deviation (they also have a range and IQR, but those are less important in inferential statistics).

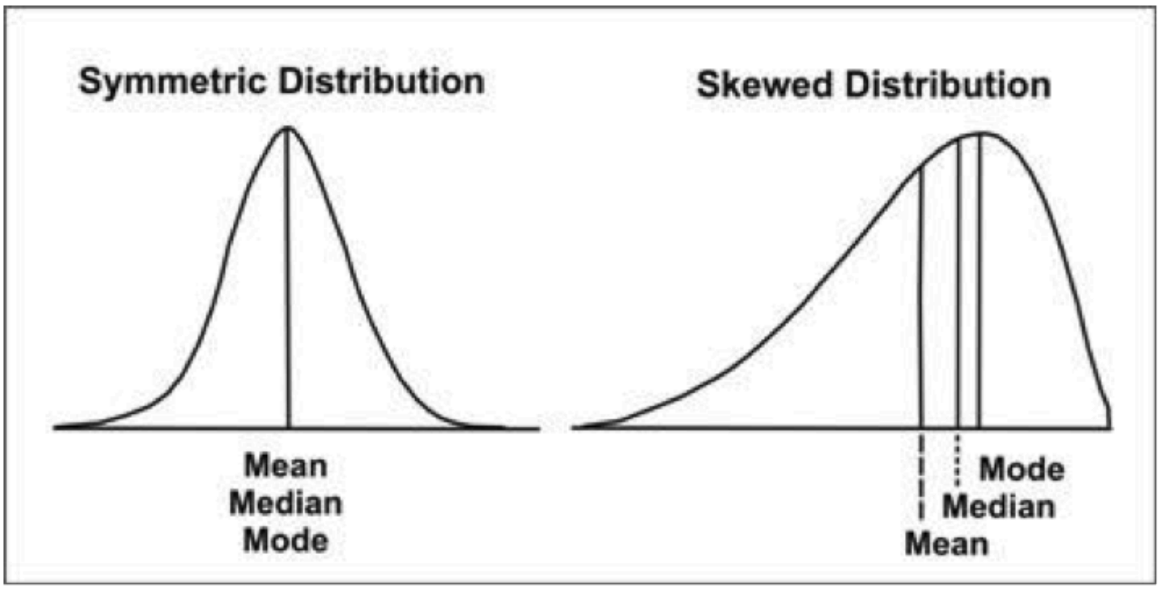
3. ***Range*** -- measure of variability is the range, given as the difference between the largest and the smallest results. It has no statistical significance, however, for small data sets.[4]
https://www.britannica.com/topic/range-statistics

4. ***IQR*** -- When a data set has outliers or extreme values, we summarize a typical value using the median as opposed to the mean. When a data set has outliers, variability is often summarized by a statistic called the interquartile range, which is the difference between the first and third quartiles. The first quartile, denoted Q1, is the value in the data set that holds 25% of the values below it. The third quartile, denoted Q3, is the value in the data set that holds 25% of the values

above it. The quartiles can be determined following the same approach that we used to determine the median, but we now consider each half of the data set separately. The interquartile range is defined as follows:

Interquartile Range = Q3-Q1[5]
https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_summarizingdata/bs704_summarizingdata7.html

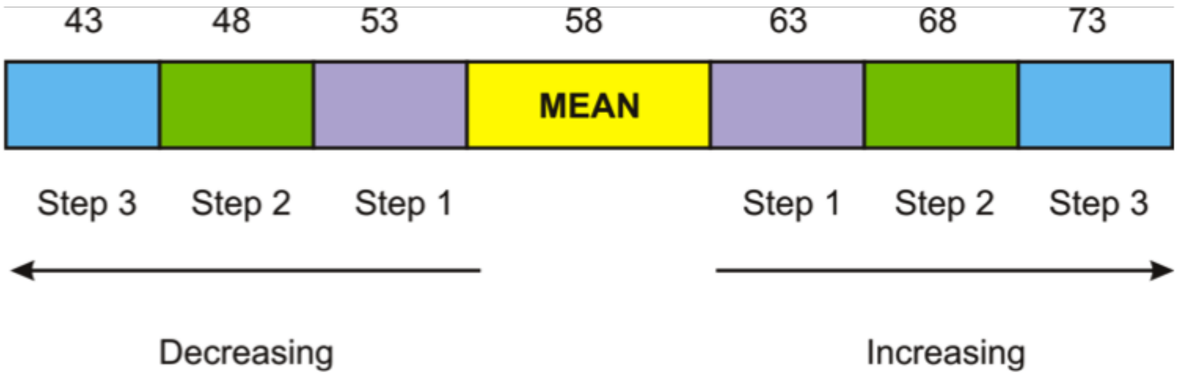
The Normal Distribution
The normal distribution is a bell-shaped, symmetrical distribution in which the mean, median and mode are all equal. If the mean, median and mode are unequal, the distribution will be either positively or negatively skewed. Consider the illustration below:



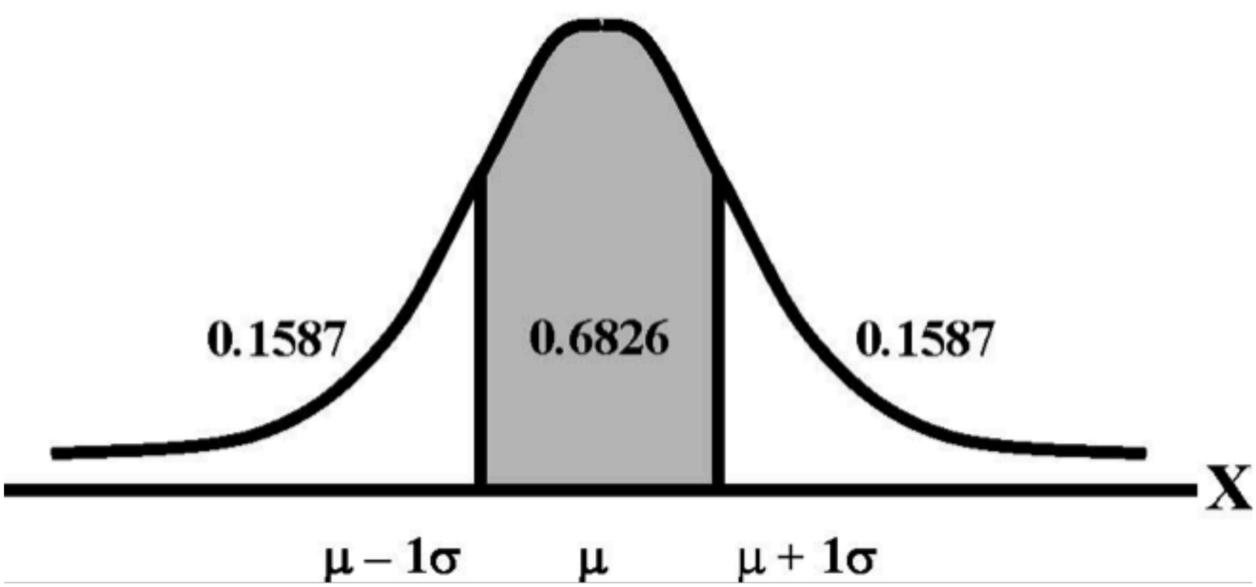
The Normal Distribution and the *Standard Deviation*

When talking about the normal distribution, it's useful to think of the standard deviation as being steps away from the mean.
One step to the right or one step to the left is considered one standard deviation away from the mean. Two steps to the left or two steps to the right are considered two standard deviations away from the mean. Likewise, three steps to the left or three steps to the right are considered three standard deviations from the mean.
The standard deviation of a dataset is simply the number (or distance) that constitutes a complete step away from the mean. Adding or subtracting the standard deviation from the mean tells us the scores that constitute a complete step.
Below is a distribution with a mean of 58 and a standard deviation of 5. For example, if the standard deviation were added to the mean, you would get a score of 63 (58 + 5 = 63).
In stats terminology, we would say that a score of 63 falls exactly "one standard deviation above the mean." Similarly, we could subtract the standard deviation from the mean (58 – 5 = 53) to find the score that falls one standard deviation below the mean.

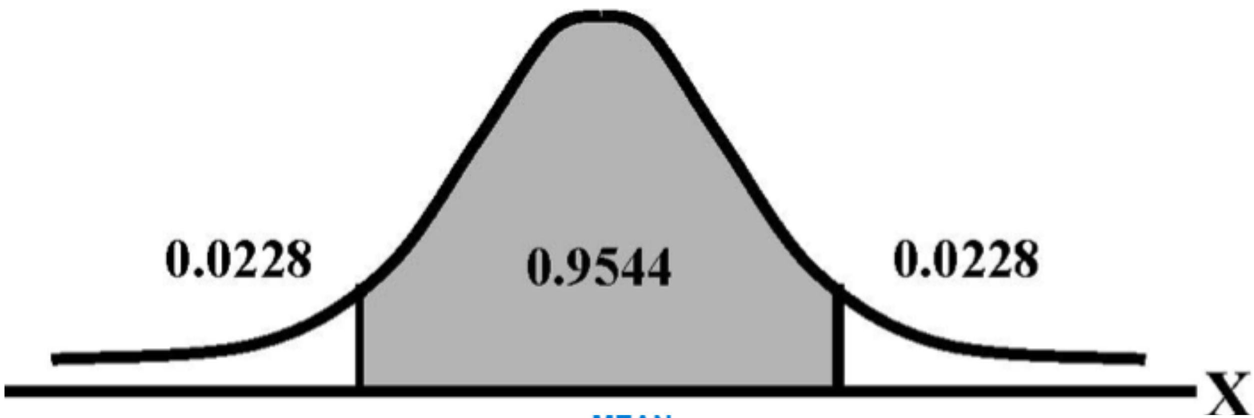
Normal distributions are important due to ***Chebyshev's Theorem***, which states that for a normal distribution a given standard deviation above and/or below the mean will always account for the same amount of area under the curve.

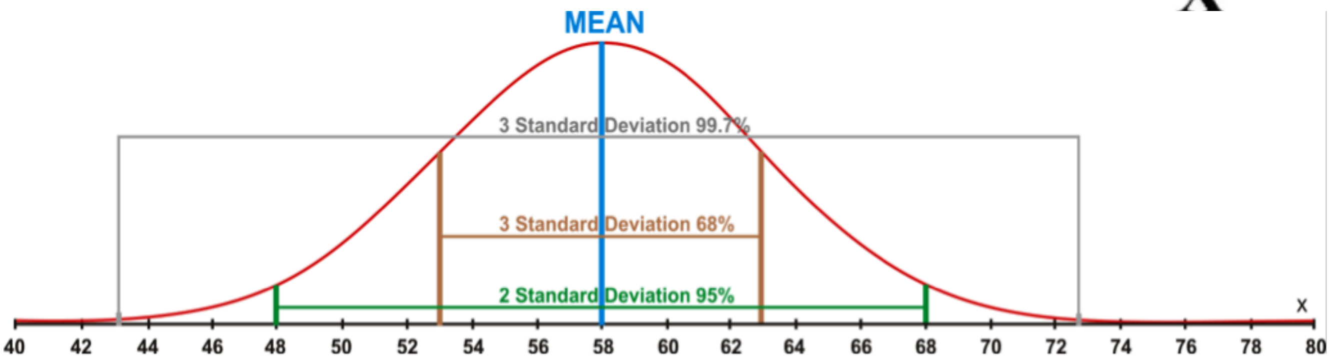


Look at the picture below. The shaded area represents the total area that falls between one standard deviation above and one standard deviation below the mean. Those Greek letters are just statistical notation for the mean and the standard deviation of a population. Regardless of what a normal distribution looks like or how big or small the standard deviation is, approximately 68 percent of the observations (or 68 percent of the area under the curve) will always fall within two standard deviations (one above and one below) of the mean.
Can you guess what proportion falls between the mean and just one standard deviation above it? If you guessed 34, you must be familiar with division (.68/2 = .34).



Now look at the next picture. It's basically the same as the first instance, only this time we're looking at two standard deviations above and below the mean. For any normal distribution, approximately 95 percent of the observations will fall within this area.





Z Scores
Z scores, which are sometimes called standard scores, represent the number of standard deviations a given raw score is above or below the mean. Sometimes it's helpful to think of z scores as just another unit of measurement.
If, for example, we were measuring time, we could express time in terms of seconds, minutes, hours or days.
Similarly we could measure distance in terms of inches, feet, yards or miles. We might have to do a little math to convert our data from one unit of measurement to another, but the thing we are measuring remains unchanged.

When we work with z scores, we're basically converting our existing data into a new unit of measurement: standard deviation units. All interval/ratio data can be expressed as z scores. We can convert any raw score into z scores by using the following formula:

$$Z = \frac{Y - \overline{Y}}{S_y}$$

In other words, we just need to subtract the mean from the raw score and divide by the standard deviation. Let's go back to our distribution with a mean of 58 and a standard deviation of 5. We can convert 63 (a raw score) into standard deviation units (z scores) fairly easily:

63-58/5 = 5/5 = 1

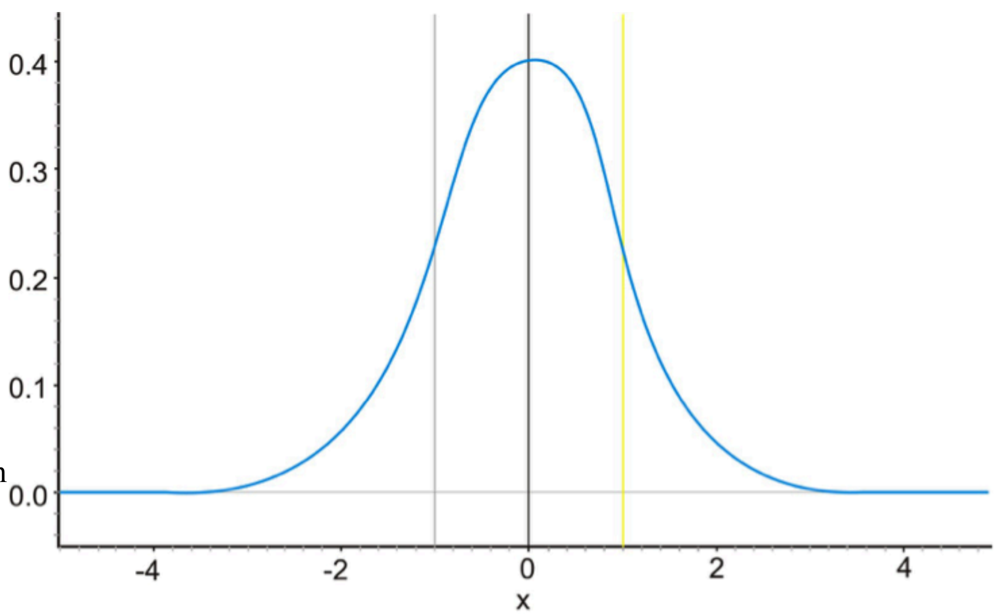
Just as one hour is equal to 60 minutes, a raw score of 63 in this distribution is equal to one standard deviation. The same holds true for observations below the mean:

53-58/5 = -5/5 = -1

In this case, because our answer is negative, we know that 53 falls exactly one standard deviation below the mean. Now suppose we wanted to convert our mean (58) into a z score:

58-58/5 = 0/5 = 0

When we convert our data into z scores, the mean will always end up being zero (it is, after all, zero steps away from itself) and the standard deviation will always be one. Data expressed in terms of z scores are known as the standard normal distribution, shown below in all of its glory.



The Standard Normal Distribution

We can

$$Y = \overline{Y} + Z(S_y)$$

We simply multiply the Z-score by the standard deviation and add that to the mean. So if we plug the numbers from our example into the formula we get:

Raw score = 58 + 1(5) = 63

Once we've got our heads around the normal distribution, Kuibyshev's theorem and z scores, we can use them to determine the percentage of our data that falls in a given area of our distribution. In order to do that, we need the cumulative z table, which I have posted on Canvas. The cumulative z table tells us what percentage of the distribution falls to the left of a given z score. I know that the table looks pretty intimidating, so we'll spend a significant amount of time going over this in class.

Main Points

The normal distribution is a symmetrical, bell-shaped distribution in which the mean, median and mode are all equal. It is a central component of inferential statistics.

The standard normal distribution is a normal distribution represented in z scores. It always has a mean of zero and a standard deviation of one.

We can use the standard normal table to calculate the area under the curve between any two points. [6](https://soc.utah.edu/sociology3112/normal-distribution.php)

Statistical Analysis:

As part of the analysis, I will examine each league chronologically. I will begin with Major League Baseball; move on to the National Hockey League, then the NFL and finally the NBA. So let's begin!

Major League Baseball: Major League Baseball began in 1871 with the National Association as the first professional baseball league. Below is the list of all baseball leagues recognized as major leagues.

The following leagues are now recognized as major leagues:
1871-1876-National Association
1876-Current – National League
1884 – Union Association
1890 – Players League
1900-Current – American League
1914-15 – Federal League
1920-31 – Negro National League
1937-62 – Negro American League
As can been there have been 8 major leagues since 1871, and as of 2022 there league are still in existence although now incorporated as on league; Major League Baseball. Based the earlier discussion about sample size, only teams that have played a minimum of 30 seasons from the National League and the American League will be evaluated.

Methodology

Here is the process that I used to come up with best, worst and whatever. Every year of a league, the league's teams played a schedule against the other teams. This was not the case in the American League and National League until incorporation of interleague play in 1997.
For each year each team in a particular league would have a record of wins, losses and possibly ties. The winning record of a team is calculated in the following manner, (Total Wins + Total Ties/Total Games Played) This establishes a winning percentage. This percentage would then be divided by the standard deviation. The average or the mean is calculated for the league teams. The mean is subtracted from the calculated winning percentage. This number is then squared and the square root is taken from this number and that is the standard deviation.

Example: National Association 1871 Philadelphia Athletics

	Games	Won	Lost	Tied	Total Points	Max Points	Winning Percentage
Philadelphia Athletics	28	21	7	0	42	56	0.75

Average Winning Percentage of the 1871 National Association (0.490386235)
Standard Deviation of the 1871 National Association (0.186407024)

Z-Score (Winning Percentage – Average Winning Percentage)/Standard Deviation
(0.75-0.490386235)/ (0.186407024) = 1.392725226.

This number represents how far above or below the particular team is from the average. This number, (1.392725226) indicates that the Philadelphia Athletics are 1.392725226 standard deviations above the average of the league in 1871.
Using Z-score conversion to a percentile, this Z-score converts to a percentile of 0.918148551. This number is close to the 92nd percentile. This is a high number relative to the rest of the league.
The z-scores are accumulated for the specific team across the league, whether they remain in the league and move to another league or the team drops from the league. An average Z-score is found for the specific team.
The Philadelphia Athletics played a total of 6 seasons, all of the seasons in the National Association. Their average Z-score was 0.417567596. This number is within 1 standard deviation or a Z-score of 1 or the average that in terms of a Z-score is zero.

Table of Z-scores: Below is a table of values associated with the z-scores and adjective used to describe the performance of the team across all the seasons the team played in the major league.

Z-Score Value	**Z-Value Characteristics**
-2 to -2.99	**Atrocious**
-1 to -1.99	**Abysmal**
-0 .01 to -99	**Negative Mediocrity**
0	**Absolute Mediocrity**
0.01 to 0.99	**Positive Mediocrity**
1 to 1.99	**Above Average**
2to 2.99	**Outstanding**

The colour-coded schemes will applied across all sports so there is no confusion about the characteristics of the teams. It also allows for a comparison of the teams across the sports.

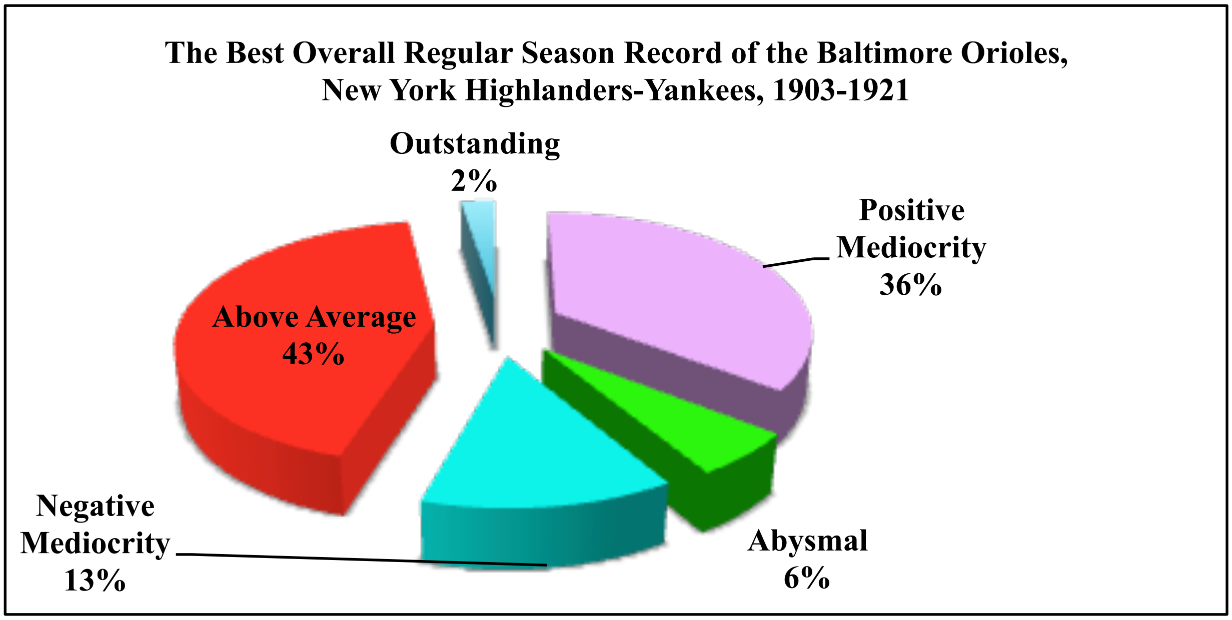
Major League Baseball

Baseball organized in 1871 under the organization of the National Association. Baseball has continued up to an including the current year (2022); this is 151 years of continuous professional baseball. All teams are considered no matter how many years a team has existed. In considering what statistic should be used to determine what the minimum number of years to be used as the criteria for a team to be considered, is any team that has played a minimum of 30 seasons. Based on the minimum of 30 seasons, a total of 31 teams meet this criterion.
The average overall z-scores are considered for these teams and the best, worst and closest team to zero are identified. These are the teams used in the evaluation of the sport. So let's start.

Da Best: to some people this would not come as any surprise. The team started out as the Baltimore Orioles, then moved to New York to become the Highlanders. They changed their

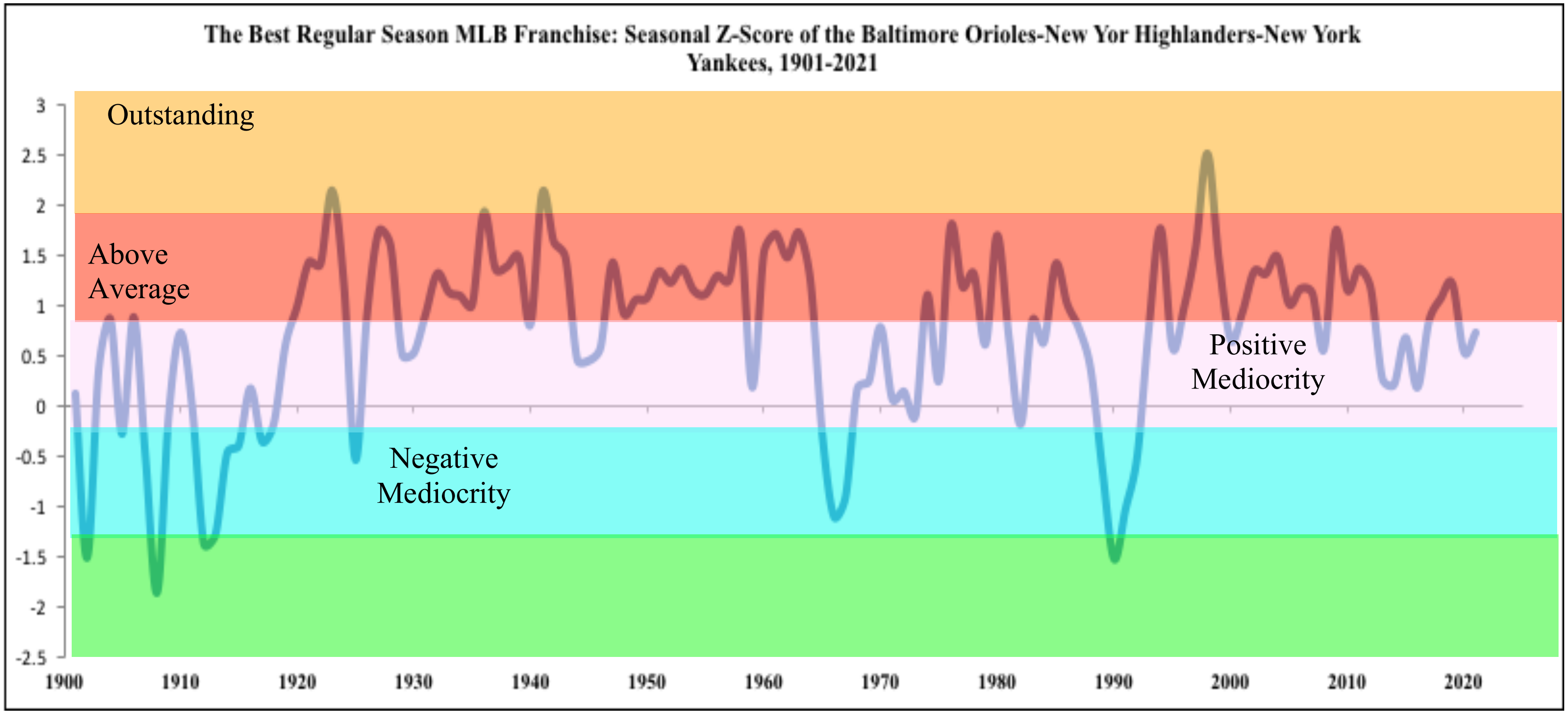
name to the Yankees in 1913 and remain the Yankees as of 2022. The current franchise is statistically the team with the top overall Z-score of 0.734. The first attribute noticed is that the Z-score is part of the positive mediocrity group. After all the pennants (40) and World Series victories (27) the franchise cannot surpass 1 standard deviation above the average.
What this tells me; is that championships are close to most teams within the 1 standard deviation of the mean. Of course, these results do not separately include the players, the management or the ownership of the club over the history of the franchise. Since the Z-score is relatively low, the other non-playing factors are important to the success of the team. That analysis is one that can be evaluated for future research.

Results:



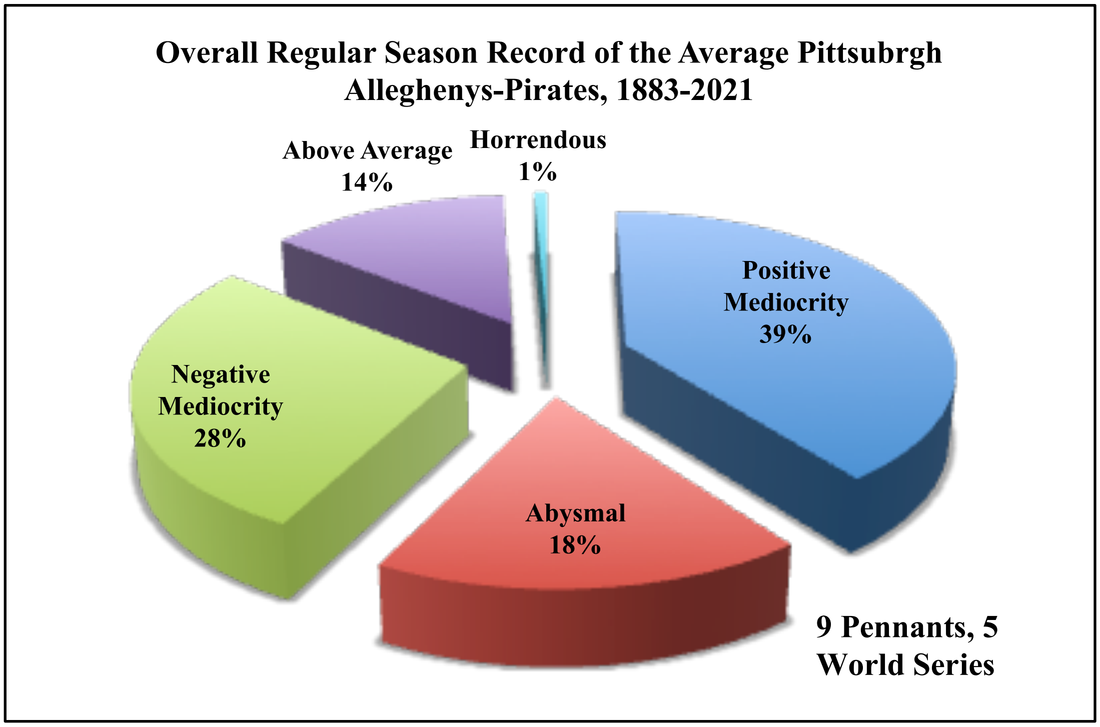
The first noticeable characteristic is that for 43% of the team's existence their record has been above average. The second most noticeable characteristic is the team finished in the positive mediocrity category with 36%. So for the history of the team, 79% of the time, the team has spent 95.59 seasons better than average. The most salient characteristic of the team is that most of the time the franchise has been better than average and therefore competitive. The team has won 40 pennants and 27 World Series. This is definitely indicative of their regular season competitiveness.

As can be seen, the amount of time spent in the abysmal category has been few and far between. The last time the team was statistically abysmal was 1990 and before that it was in 1911. The team has not gone through long periods of poor play and when it does happen it has been or short periods of time.



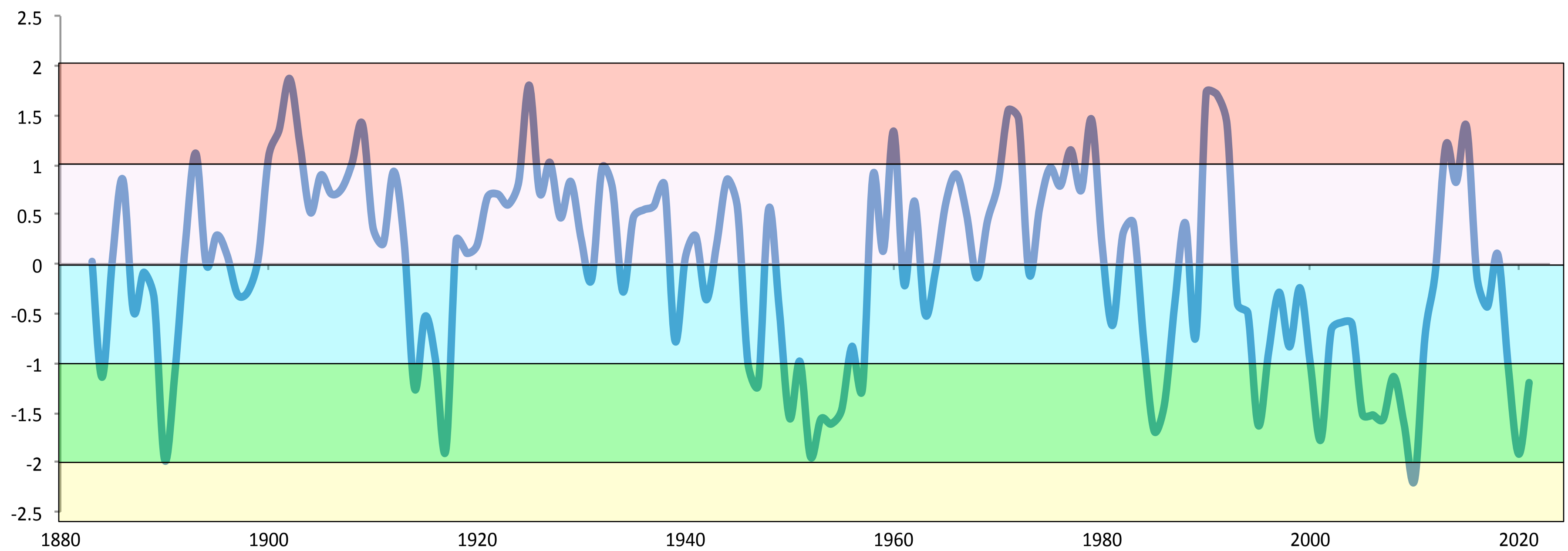
Whatever: This is the major league team that is the closest to absolute zero in its Z-score. This means that the team is average can be during the regular season. They are close to having a 50% winning record and a 50% losing record. The closest team to having such a record are the original Pittsburgh Alleghneys who later become the Pittsburgh Pirates. They have been in Major League Baseball since 1883. As of 2022, that is 138 seasons.
Their respective Z-score is -0.0136116. This is 1.36% away from having a z-score of 0 or 50% winning and losing record over 138 seasons. They appear to follow the statistical model of regression to the mean.
The team has won 9 pennants and 5 World Series in their history. So they have succeeded some of the time, but mostly are an average team. They appear to follow the statistical model of regression to the mean. They are the team that most closely follows the regression to mediocrity more closely than any other team.

Regression to the Mediocre:
Francis Galton (1822–1911) was a scientific polymath who made pioneering contributions to several fields, including the study of heredity in populations and the statistical analysis of data. Motivated by eugenic concerns, he studied patterns of human heredity. It is often claimed that while undertaking this work he discovered the statistical phenomenon of regression to the mean and that this discovery was a pivotal moment in the history of statistical thinking.[7]
https://www.sciencedirect.com/science/article/abs/pii/S0039368120302090#:~:text=Galton%20used%20regression%20towards%20mediocrity,both%20natural%20selection%20and%20eugenics.[9]
Galton produced the following paper: Regression towards Mediocrity Hereditary Stature, in The Journal of the Anthropological Institute of Great Britain and Ireland.[8]
Vol. 15 (1886), pp. 246-263 (20 pages). His result is the following:
What Galton deduced was that the offspring of parents tended to be as tall as the average person in England, not the average of the parents. This law is still applicable as it pertains to any measurement, in this case, sports winning percentage.

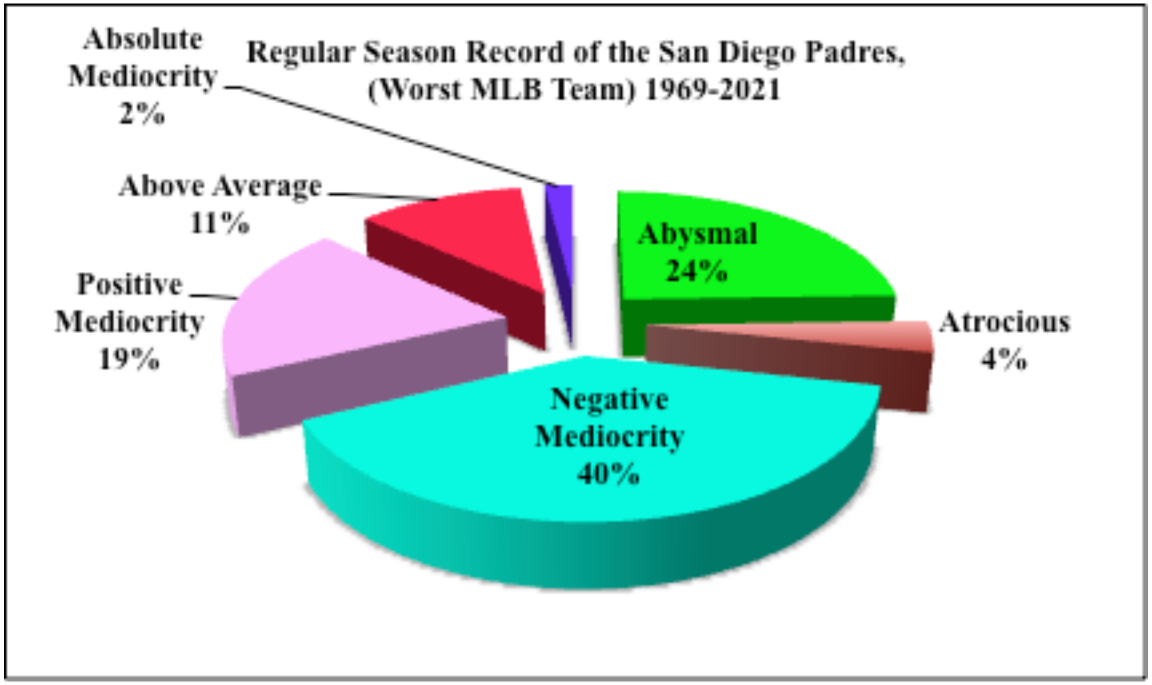


As can be seen, the positive occupies 54% and the negative is 46%. This is not exactly 50-50, but it is very close.

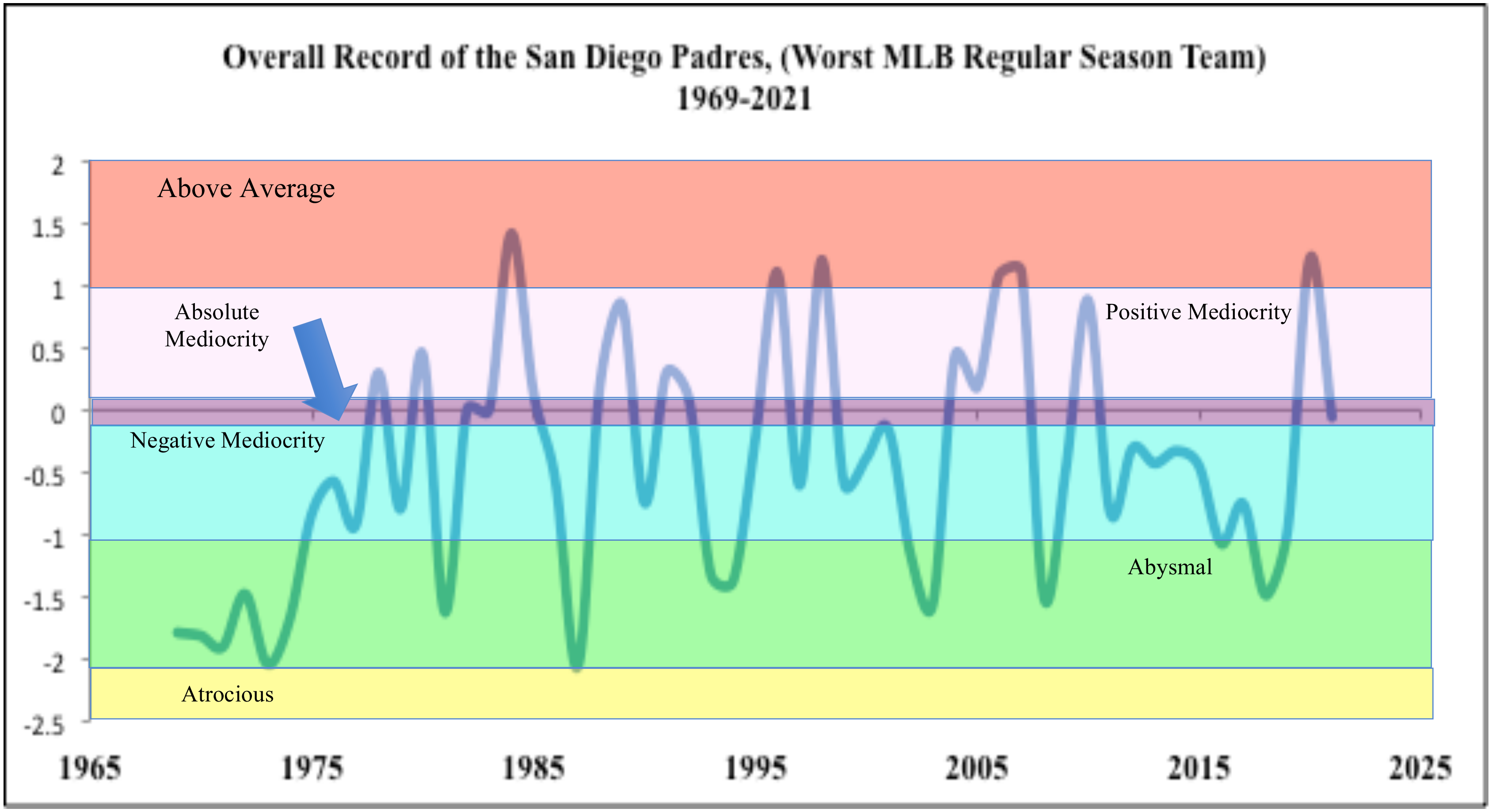
The Whatever Team: Overall Seasonal Record of the Pittsburgh Alleghenys-Pirates, 1883-2021



The Worst: The worst major league baseball team by the Z-score of the regular seasons is the San Diego Padres. The Padres have been in Major League Baseball since 1969. In that time, the team has only won 1 pennant and no world series. Their average Z-score is -0.450485233. This is not that bad and the result is only halfway between 0 and negative one. They are a mediocre team except they are in the negative category of mediocrity as opposed the positive category or close to zero.



The team has only been above absolute mediocrity 30% of the franchise's history. So for 68% of the time the team has played below absolute mediocrity. That adds up to 98%, so the remaining 2 percent was occupied by a same number of wins and losses. They are truly a franchise that plays poorly on a consistent basis. They are a bad team, but in relative terms, not that bad considering their respective Z-score.


Overall Record of the San Diego Padres, (Worst MLB Regular Season Team)
1969-2021
Above Average
Absolute
Mediocrity
Positive Mediocrity
Negative Mediocrity
Abysmal
Atrocious
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1965
1975
1985
1995
2005
2015
2025

Statistical Analysis
I have added an additional element to the evaluation. This is a statistical analysis of the three teams. I wanted to see if there were differences between the teams. Were the Z-scores obtained by each of the teams statistically different than the other two teams? Since the average Z-Scores were all within the plus 1 and -1 Z-scores, an analysis was done. An Analysis of Variance, (ANOVA) was carried on the seasonal Z-scores of the Baltimore-Orioles-New York Highlanders-Yankees, Pittsburgh Alleghenys-Pirates and the San Diego Padres.

***Analysis of Variance*:**

What is this test for?
The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups. This guide will provide a brief introduction to the one-way ANOVA, including the assumptions of the test and when you should use this test.

What does this test do?
The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other. Specifically, it tests the null hypothesis:

Where μ = group mean and k = number of groups. If, however, the one-way ANOVA returns a statistically significant result, we accept the alternative hypothesis (H_A), which is that there are at least two group means that are statistically significantly different from each other.

At this point, it is important to realize that the one-way ANOVA is an omnibus test statistic and cannot tell you which specific groups were statistically significantly different from each other, only that at least two groups were.

To determine which specific groups differed from each other, you need to use a **post hoc test**.[10]
https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php

ANOVA Table:

The sums of squares SST and SSE previously computed for the one-way ANOVA are used to form two mean squares, one for *treatments* and the second for *error*. MST and MSE denote these mean squares, respectively. These are typically displayed in a tabular form, known as an *ANOVA Table*. The ANOVA table also shows the statistics used to test hypotheses about the population means.

When the null hypothesis of equal means is true, the two mean squares estimate the same quantity (error variance), and should be of approximately equal magnitude. In other words, their ratio should be close to 1. If the null hypothesis is false, MST should be larger than MSE.

Dividing the sum of squares by the associated degrees of freedom forms the mean squares.

Let $N=\sum ni$. Then, the degrees of freedom for treatment are
$DFT=k-1$,
and the degrees of freedom for error are
$DFE=N-k$.
The corresponding *mean squares* are:
$MST=SST/DFT$
$MSE=SSE/DFE$.

The test statistic, used in testing the equality of treatment means is: $F=MST/MSE$.
The critical value is the tabular value of the F distribution, based on the chosen α level and the degrees of freedom DFT and DFE.
The calculations are displayed in an ANOVA table, as follows:

ANOVA table

Source	**SS**	**DF**	**MS**	**F**
Treatments	**SST**	**k−1**	**SST/(k−1)**	**MST/MSE**
Error	**SSE**	**N−k**	**SSE/(N−k)**	
Total (corrected)	**SS**	**N−1**		

The word "source" stands for source of variation. Some authors prefer to use "between" and "within" instead of "treatments" and "error", respectively.

[11]https://www.itl.nist.gov/div898/handbook/prc/section4/prc433.htm#:~:text=The%20ANOVA%20table%20also%20shows,hypotheses%20about%20the%20population%20means.&text=When%20the%20null%20hypothesis%20of,be%20of%20approximately%20equal%20magnitude.

One-Way ANOVA Post Hoc Tests

Once you have determined that differences exist among the means, post hoc range tests and pairwise multiple comparisons can determine which means differ. Range tests identify homogeneous subsets of means that are not different from each other. Pairwise multiple comparisons test the difference between each pair of means and yield a matrix where asterisks indicate significantly different group means at an alpha level of 0.05.

Equal Variances Assumed

Scheffé's Method
Scheffé's method for investigating all possible contrasts of the means corresponds exactly to the *F*-test in the following sense. If the *F*-test rejects the null hypothesis at level α, then there exists at least one contrast that would be rejected using the Scheffé procedure at level α. [11] Therefore, Scheffé provides α level protection against rejecting the null hypothesis when it is true, regardless of how many contrasts of the means are tested.[12]
https://online.stat.psu.edu/stat503/lesson/3/3.3

Analysis Between Teams

Descriptive statistics of the 3 independent treatments:				
Treatment →	**A (Baltimore Orioles-New York Highlanders-Yankees)**	**B (Pittsburgh Alleghneys-Pirates)**	**C (San Diego Padres)**	**Pooled Total**
(Seasons)	121	139	53	313
Average Z-Score	0.7341	-0.0136	-0.4505	0.2015

One-way ANOVA of the 3 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	63.2815	2	31.6407	37.4969	**2.55E-15**
Error	261.585	310	0.8438		
Total	324.8665	312			

*Statistically Significant

***-** The p-value is statistically significant with the value less than .01 or a 99% confidence level that the result is valid and not due to chance. This result indicates that significance. This allows for applying the Scheffé test to see if the differences between the teams are statistically significant.

Scheffé Test

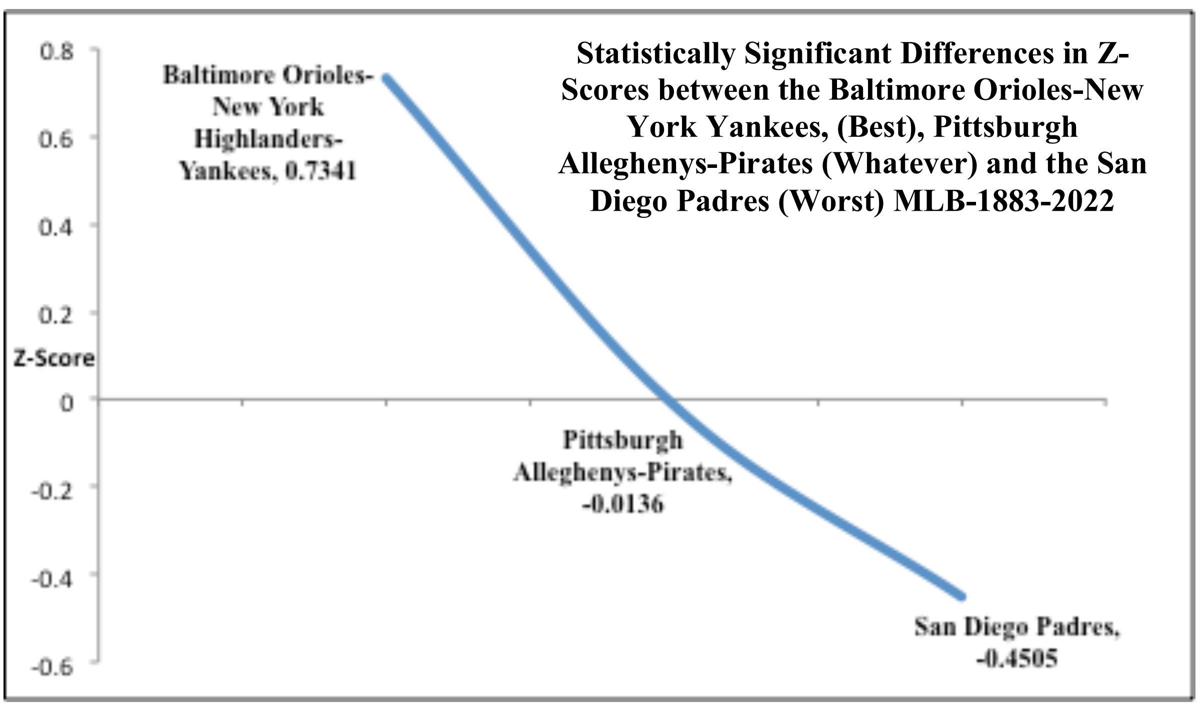
Scheffé results			
Treatment Pairs	Scheffé	Scheffé	Scheffé
	TT-statistic	p-value	Inference
A vs. B (Baltimore Orioles-New York Highlanders-New Yankees vs. Pittsburgh Alleghenys-Pirates)	6.5465	1.92E-09	**** p<0.01**
A vs. C (Baltimore Orioles-New York Highlanders-Yankees vs. San Diego Padres)	7.8287	7.18E-13	**** p<0.01**
B vs. C (Pittsburgh Alleghenys-Pirates vs. San Diego Padres)	2.9459	0.0138472	*** p<0.05**

**** 99% confidence level**
***95% confidence level**
There are statistically significant differences between the three teams based on the test.

Confidence Levels

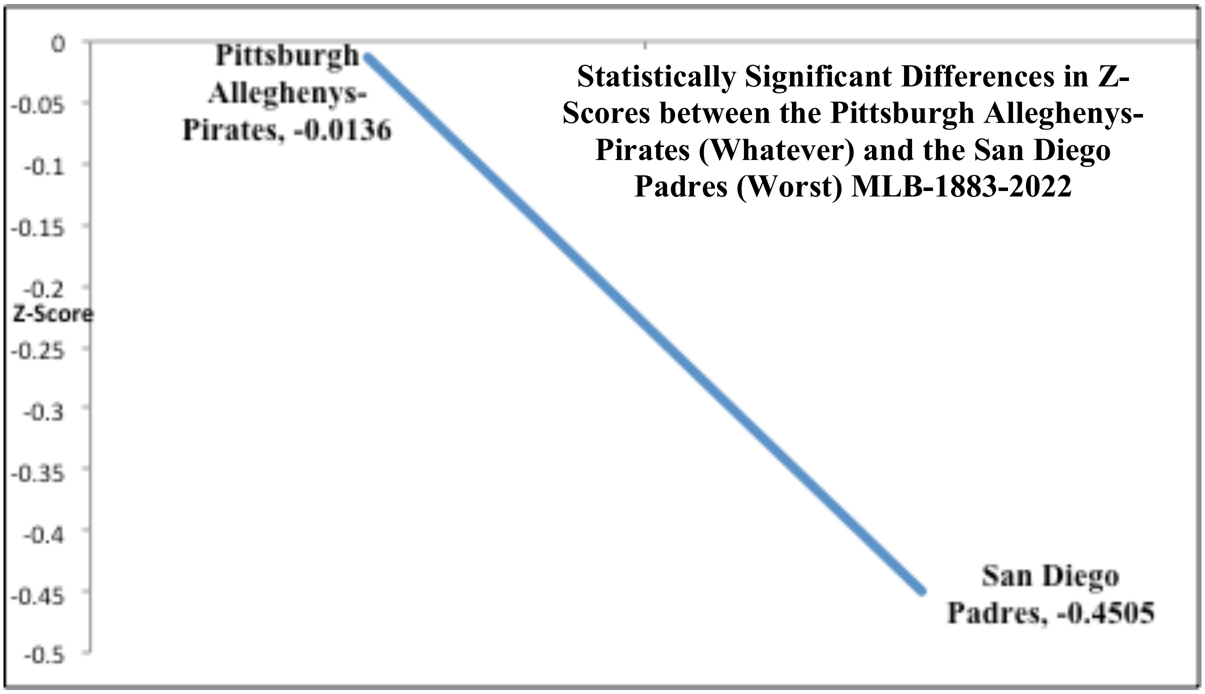
Strictly speaking a 95% confidence interval means that if we were to take 100 different samples and compute a 95% confidence interval for each sample, then approximately 95 of the 100 confidence intervals will contain the true mean value (μ). In practice, however, we select one random sample and generate one confidence interval, which may or may not contain the true mean.
The observed interval may over- or underestimate μ. Consequently, the 95% CI is the likely range of the true, unknown parameter. The confidence interval does not reflect the variability in the unknown parameter. Rather, it reflects the amount of random error in the sample and provides a range of values that are likely to include the unknown parameter. Another way of thinking about a confidence interval is that it is the range of likely values of the parameter (defined as the point estimate $\pm$ margin of error) with a specified level of confidence (which is similar to a probability).
Suppose we want to generate a 95% confidence interval estimate for an unknown population mean. This means that there is a 95% probability that the confidence interval will contain the true population mean. Thus, P([sample mean] - margin of error < μ < [sample mean] + margin of error) = 0.95. [13]
https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_confidence_intervals/bs704_confidence_intervals_print.html

Result 1: Baltimore Orioles-New York Highlanders-Yankees vs. Pittsburgh Alleghenys-Pirates and San Diego Padres.



The Yankees franchise is statistically significantly different than the Pittsburgh and San Diego franchise. The Yankees franchise has averaged a positive mediocrity regular season while the Pittsburgh and San Diego franchise have averaged a negative mediocrity Z-score for their regular season record. Therefore the Scheffé results illustrate the Yankees franchise is truly different than the other two franchises.

Result 2: Pittsburgh Alleghenys-Pirates vs. San Diego Padres.



The second result between the Pittsburgh and San Diego franchise is statistically significant. Both franchises are in the negative category, however the Pittsburgh franchise is closer to a zero average, (Absolute Mediocrity) while San Diego is closer to the halfway mark of negative mediocrity.

Within Team Analysis: As an additional analysis, an ANOVA was conducted within each team to see if there are differences between the performance categories. So the question to ask is, are there differences between the categories for each of the teams? An ANOVA will be applied to each of the teams.

Baltimore Orioles-New York Highlanders-Yankees Analysis

Descriptive statistics of the 3 independent treatments:				
Categories →	**A (Positive Mediocrity)**	**B (Negative Mediocrity)**	**C (Above Average)**	**Pooled Total**
Seasons	41	14	53	108
Average Z-Scores	0.5694	-0.344	1.3653	0.8416

One-way ANOVA of the 3 independent treatments*:*					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value **
Treatment	37.2544	2	18.6272	282.7296	**1.11E-16**
Error	6.9178	105	0.0659		
Total	44.1721	107			

*** - Significant at the 99% confidence level.**

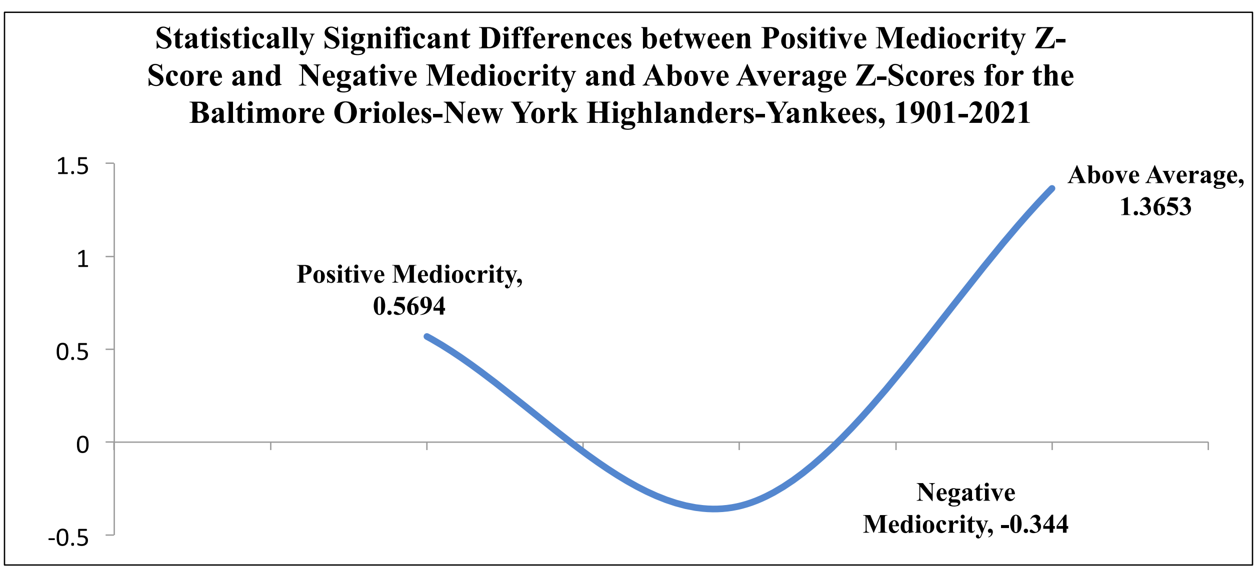
Scheffé results			
Treatment Pairs	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B	11.4965	1.11E-16	** **p<0.01**
A vs. C	14.9087	1.11E-16	** **p<0.01**
B vs. C	22.1618	1.11E-16	** **p<0.01**

****-** Scheffé test is statistically significant between the three categories.

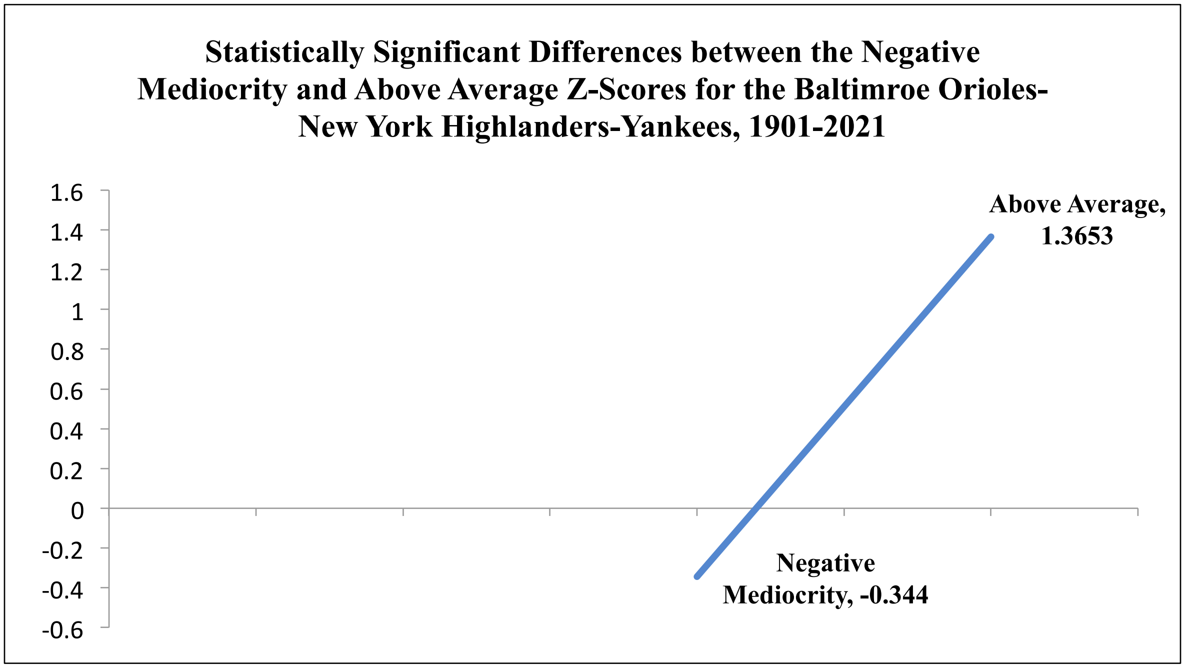
Graph Results of Between Teams Analysis:
Result 1

The Positive Mediocrity category is statistically significantly different the Abysmal, Negative Mediocrity and Above Average categories.

Graph Result 1



Graph Result 2



The Negative Mediocrity category is statistically significantly different than the Above Average category.

The franchise has occupied four categories in its history and they are distinct from each other.

Pittsburgh Alleghenys-Pirates Analysis

Descriptive statistics of the 4 independent treatments:					
Categories →	**A (Positive Mediocrity)**	**B (Negative Mediocrity)**	**C (Abysmal)**	**D (Above Average)**	**Pooled Total**
Seasons	56	38	24	19	137
Average Z-Score	0.5122	-0.466	-1.4923	1.3828	0.0104

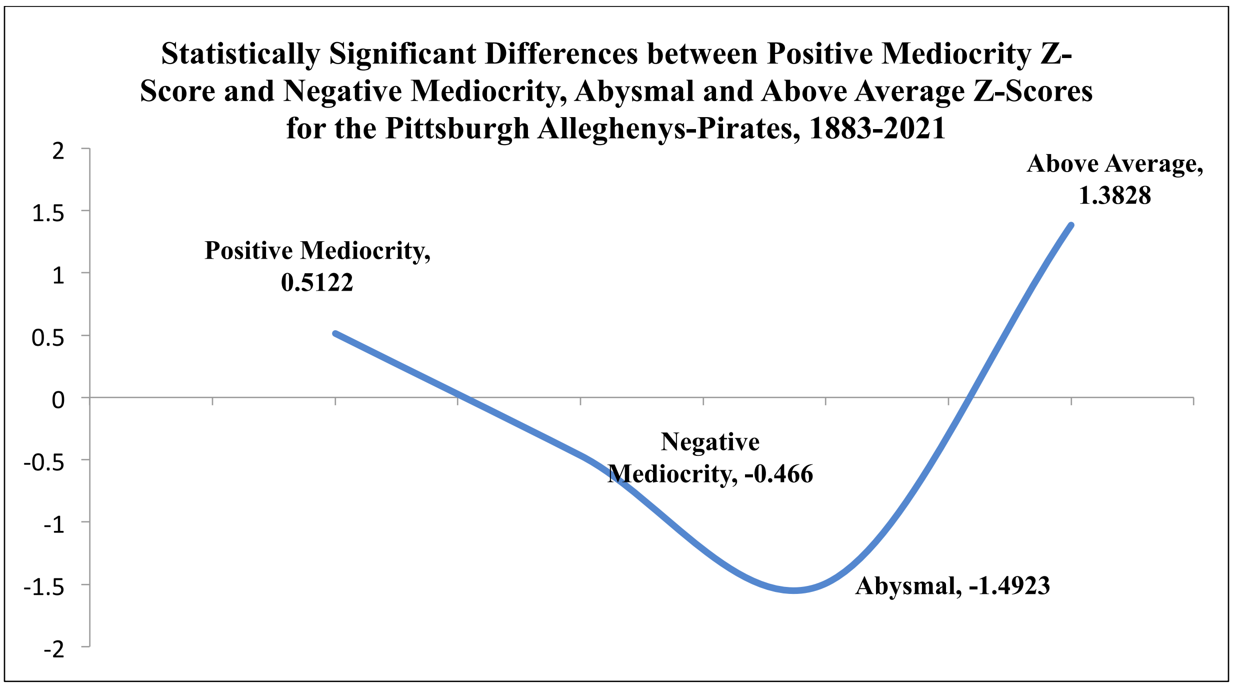
One-way ANOVA of the 4 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	112.7029	3	37.5676	460.3236	1.11E-16
Error	10.8543	133	0.0816		
Total	123.5572	136			

*- Overall the ANOVA is statistically significant at the 99% confidence level

Scheffé results			
Treatments	Scheffé	Scheffé	Scheffé
Pair	TT-statistic	p-value	Inference
A vs. B	16.292	1.11E-16	** p<0.01
A vs. C	28.7588	1.11E-16	** p<0.01
A vs. D	11.4789	1.11E-16	** p<0.01
B vs. C	13.7776	1.11E-16	** p<0.01
B vs. D	23.0332	1.11E-16	** p<0.01
C vs. D	32.7733	1.11E-16	** p<0.01

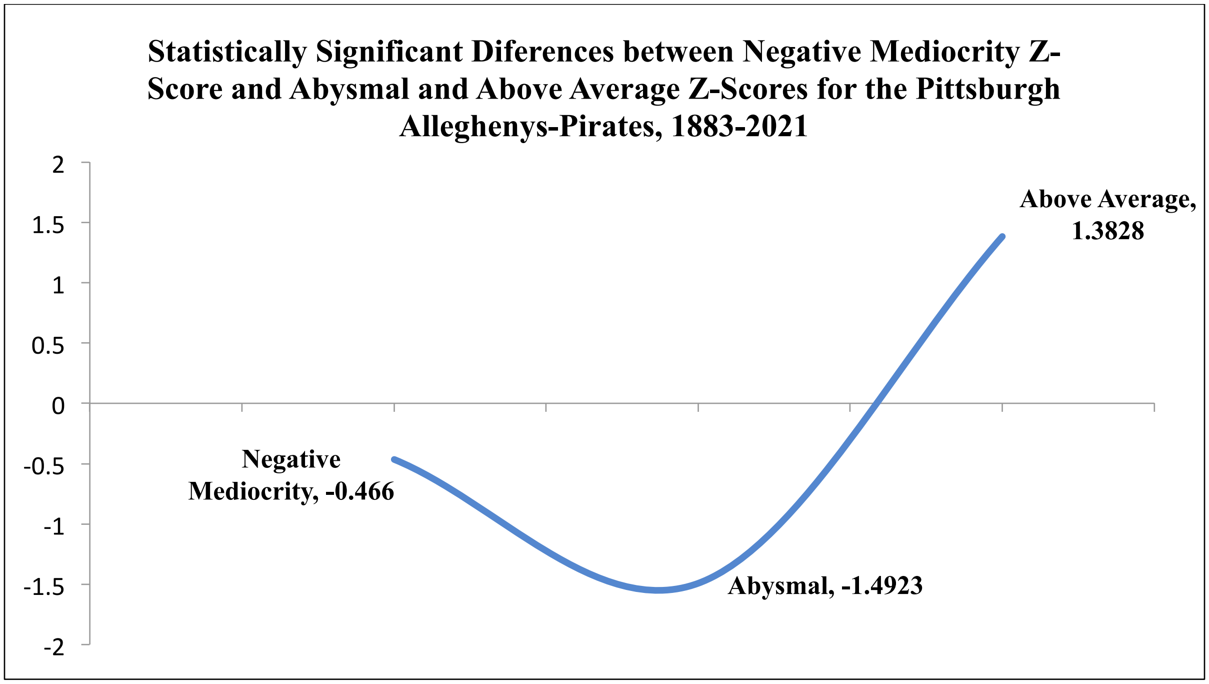
Based on the Scheffé test, all the comparisons within the Pittsburgh franchise are statistically significant at the 99% confidence level. Below are the graphs of the Scheffé results.

Graph Result 1



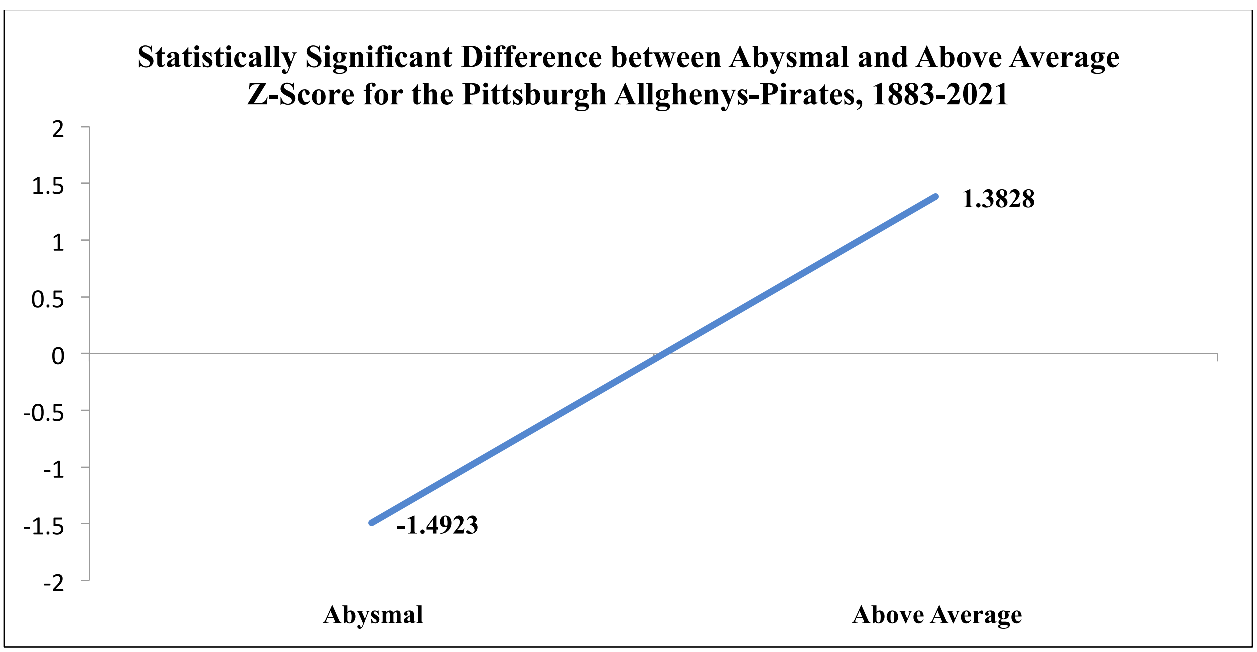
The Z-Score for Positive Mediocrity is statistically significantly different than the Negative Mediocrity, Abysmal and Above Average categories.

Graph 2 Results



The Z-Score for Negative Mediocrity is statistically different than the Abysmal and Above Average Z-Score category.

Graph 3 Results



The Abysmal Category is statistically significantly different than the Above Average Category.

San Diego Padres Analysis

Descriptive statistics of the 3 independent treatments:				
Categories →	**A (Abysmal)**	**B (Negative Mediocrity)**	**C (Positive Mediocrity)**	**Pooled Total**
Seasons	13	21	10	44
Average Z-Score	-1.5201	-0.5232	0.383	-0.6118

One-way ANOVA of 3 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value
Treatment	20.7859	2	10.393	142.6421	1.11E-16*
Error	2.9873	41	0.0729		
Total	23.7732	43			

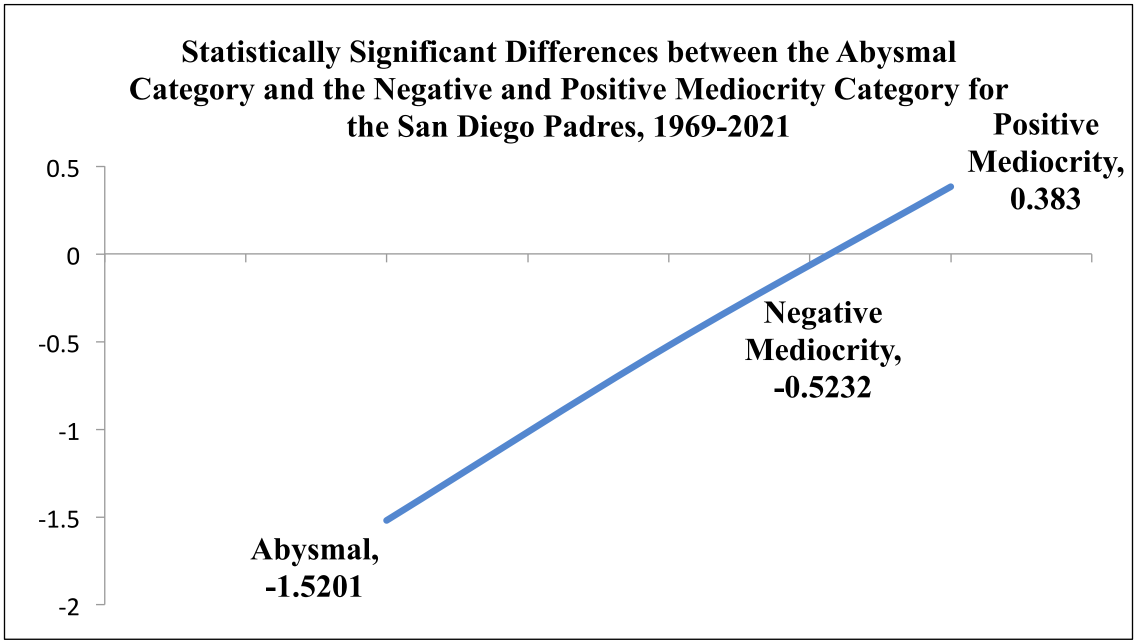
*****-Result is statistically significant at the 99% confidence level allowing for the application of the Scheffé test.

Scheffé Results

Scheffé results			
Treatment Pairs	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B	10.4652	2.64E-12	**** p<0.01**
A vs. C	16.7618	1.11E-16	**** p<0.01**
B vs. C	8.7378	4.34E-10	**** p<0.01**

The test indicates that there are statistically significant differences between the three different groups.

Graph Result 1

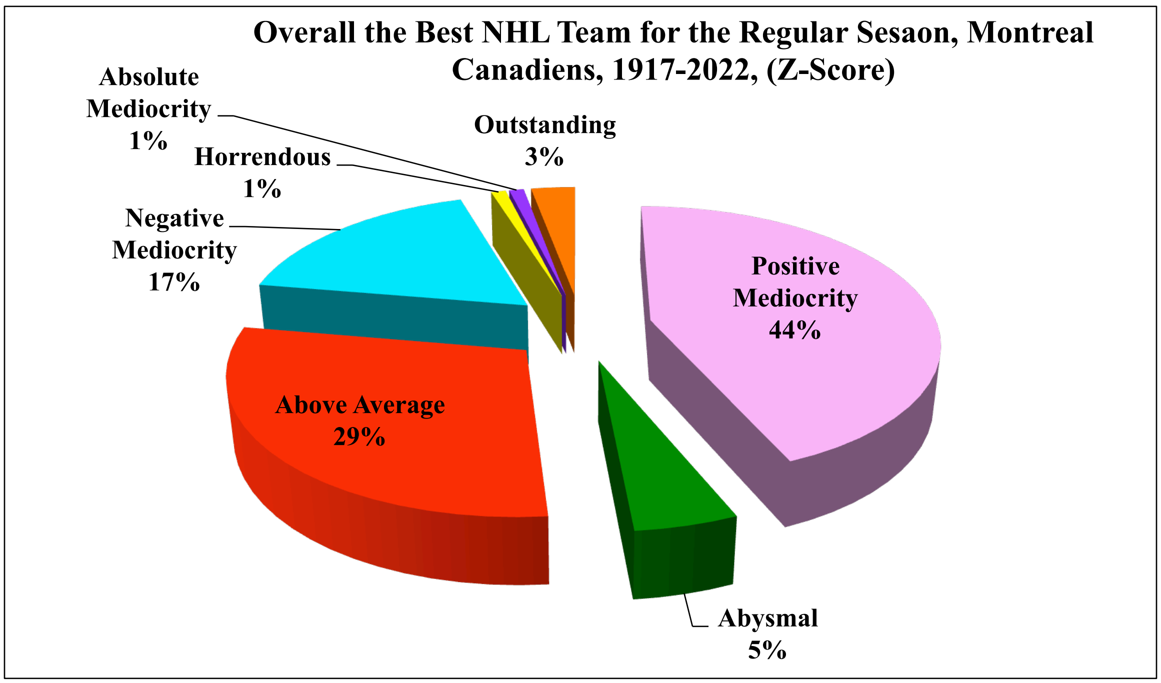


The Abysmal, Positive and Negative Mediocrity categories are statistically different than each other. The positive mediocrity category is closer to zero while the negative mediocrity category is halfway to the Abysmal category. The abysmal category is halfway to the atrocious category.

Conclusion: Based on the methodology, the three teams are in 3 distinct groups and there are statistically significant differences between and within the three groups. These three are truly the best, the worst and whatever groups of Major League Baseball.

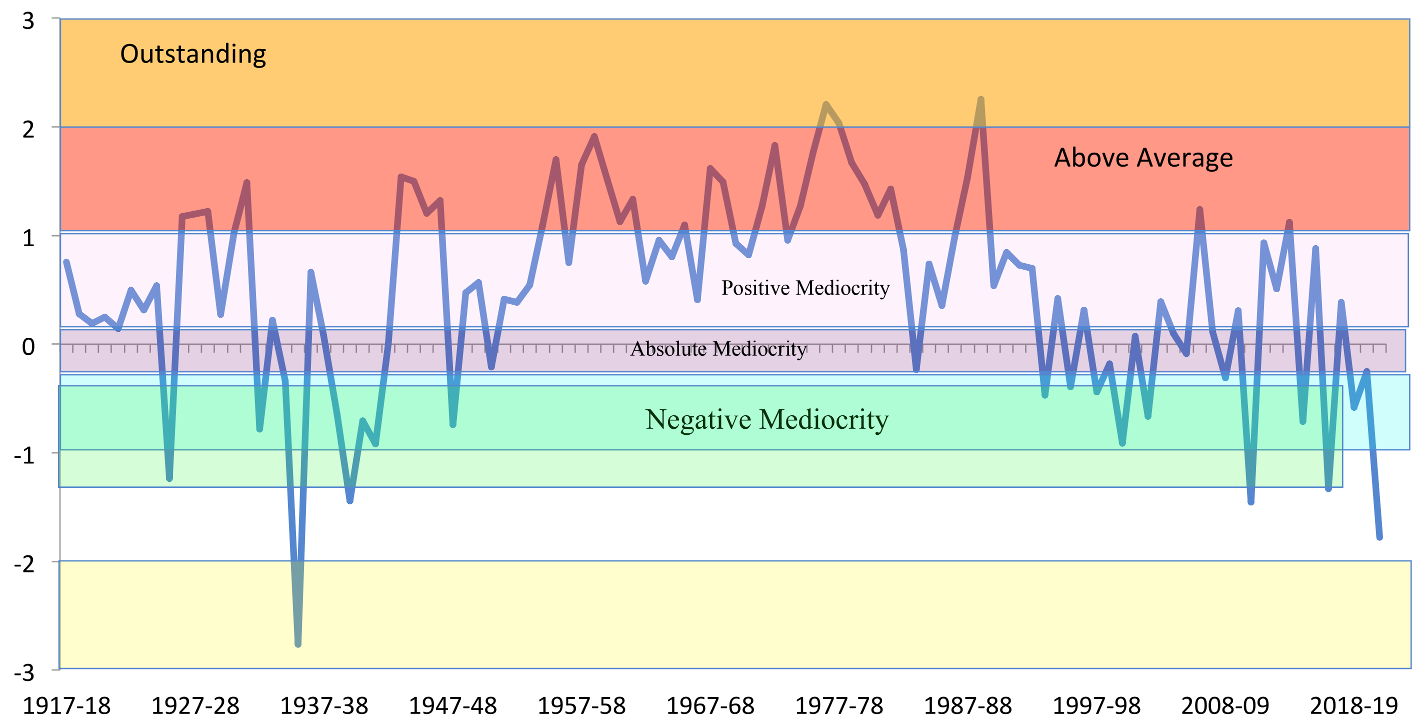
National Hockey League: The National Hockey League is the second oldest professional sports league in North America. It started in 1917, 46 years after Major League Baseball began. Let's see if this league shows any similar patterns to the MLB.
The leagues where teams are recruited are the NHL and the WHA in which six teams eventually merged with the NHL.
Based on the criteria of the mean, the minimum number of seasons is 30.72 or rounded up 31 seasons. As of 2022, this criterion yields a total of 21 teams. The team with the best overall Z-Score is the Montreal Canadiens, (0.504096292), the most average Z-Score is the Minnesota North Stars-Dallas Stars, (0.006989873) and the worst Z-Score belongs to the Winnipeg Jets-Phoenix Coyotes, (-0.32422125)

The Best: The team with the best overall Z-score is the Montreal Canadiens.



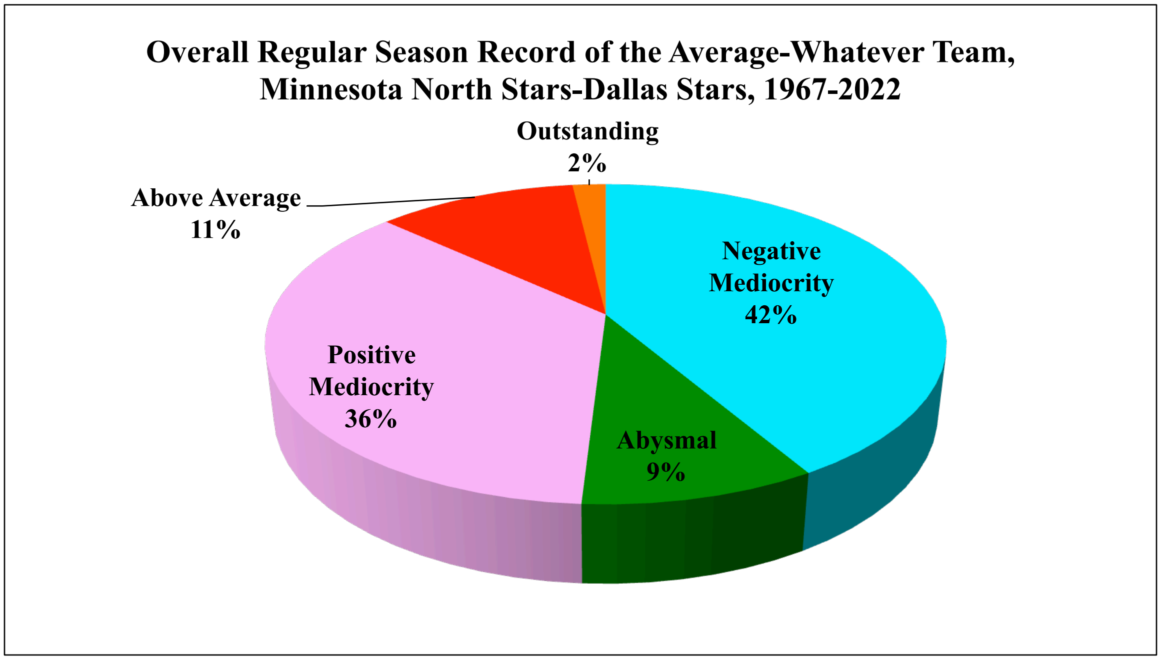
The team has encompassed all possible categories, but Positive Mediocrity accounts for almost 50% of their seasonal record.

Z-scores for the Montreal Canadiens, 1917-18 to 2021-22.

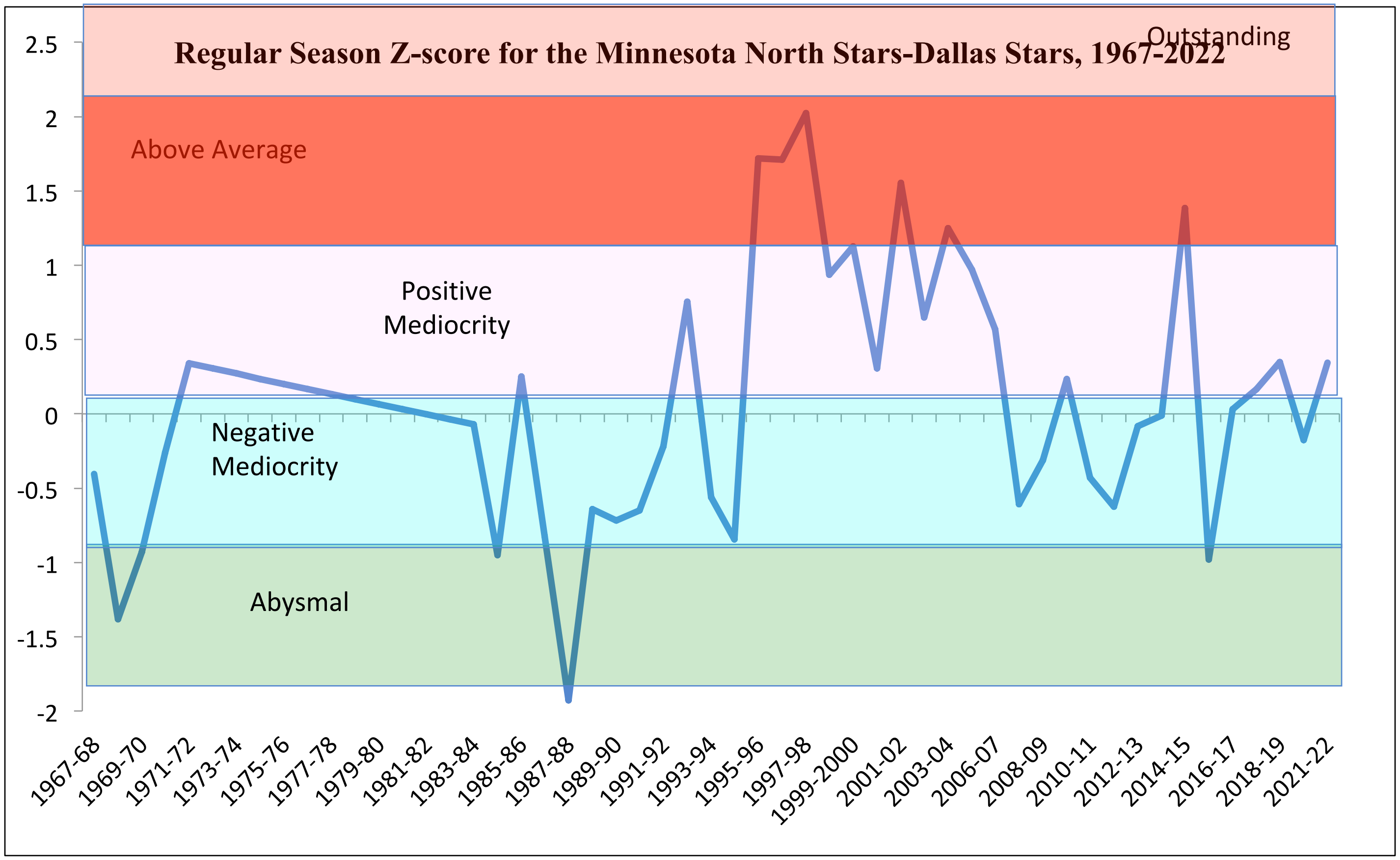


Analysis between Teams

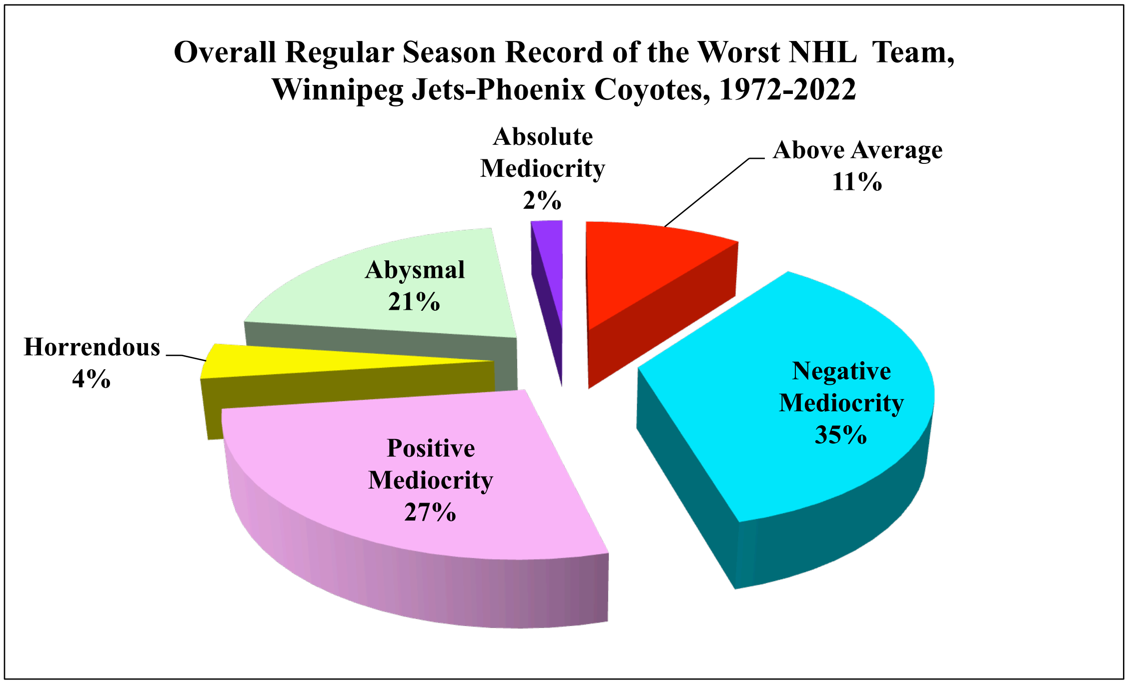
The Average-Minnesota North Stars-Dallas Stars



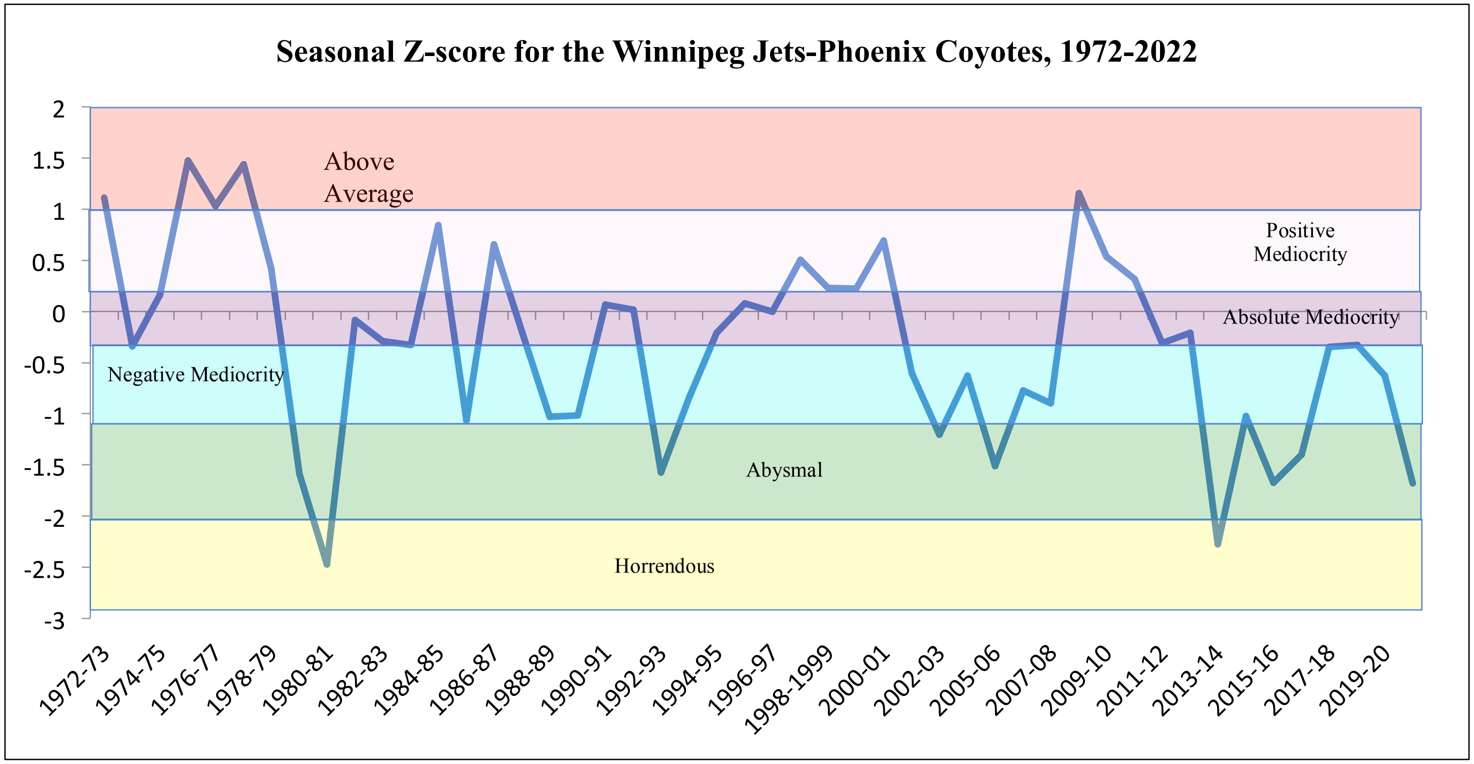
The franchise is dominated by Mediocrity, 78% of the time. This is truly a mediocre team.


Regular Season Z-score for the Minnesota North Stars-Dallas Stars, 1967-2022
Outstanding
Above Average
Positive
Mediocrity
Negative
Mediocrity
Abysmal
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1967-68
1969-70
1971-72
1973-74
1975-76
1977-78
1979-80
1981-82
1983-84
1985-86
1987-88
1989-90
1991-92
1993-94
1995-96
1997-98
1999-2000
2001-02
2003-04
2006-07
2008-09
2010-11
2012-13
2014-15
2016-17
2018-19
2021-22

The Worst Team-Winnipeg Jets-Phoenix Coyotes, 1972-2022



The team occupies six different categories, with negative categories occupying 50% .


Seasonal Z-score for the Winnipeg Jets-Phoenix Coyotes, 1972-2022
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
Above Average
Positive Mediocrity
Absolute Mediocrity
Negative Mediocrity
Abysmal
Horrendous
1972-73
1974-75
1976-77
1978-79
1980-81
1982-83
1984-85
1986-87
1988-89
1990-91
1992-93
1994-95
1996-97
1998-1999
2000-01
2002-03
2005-06
2007-08
2009-10
2011-12
2013-14
2015-16
2017-18
2019-20

Analysis Between Teams

Descriptive statistics of the 3 independent treatments:				
Teams →	A (Montreal Canadiens)	B (Minnesota-Dallas Stars)	C (Winnipeg Jets-Phoenix Coyotes)	Pooled Total
Seasons	103	53	48	204
Average Z-Score	0.5041	0.007	-0.3242	0.18

One-way ANOVA of the k***=3 independent treatments:***					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	24.6089	2	12.3044	14.509	1.30E-06
Error	170.4592	201	0.8481		
Total	195.0681	203			

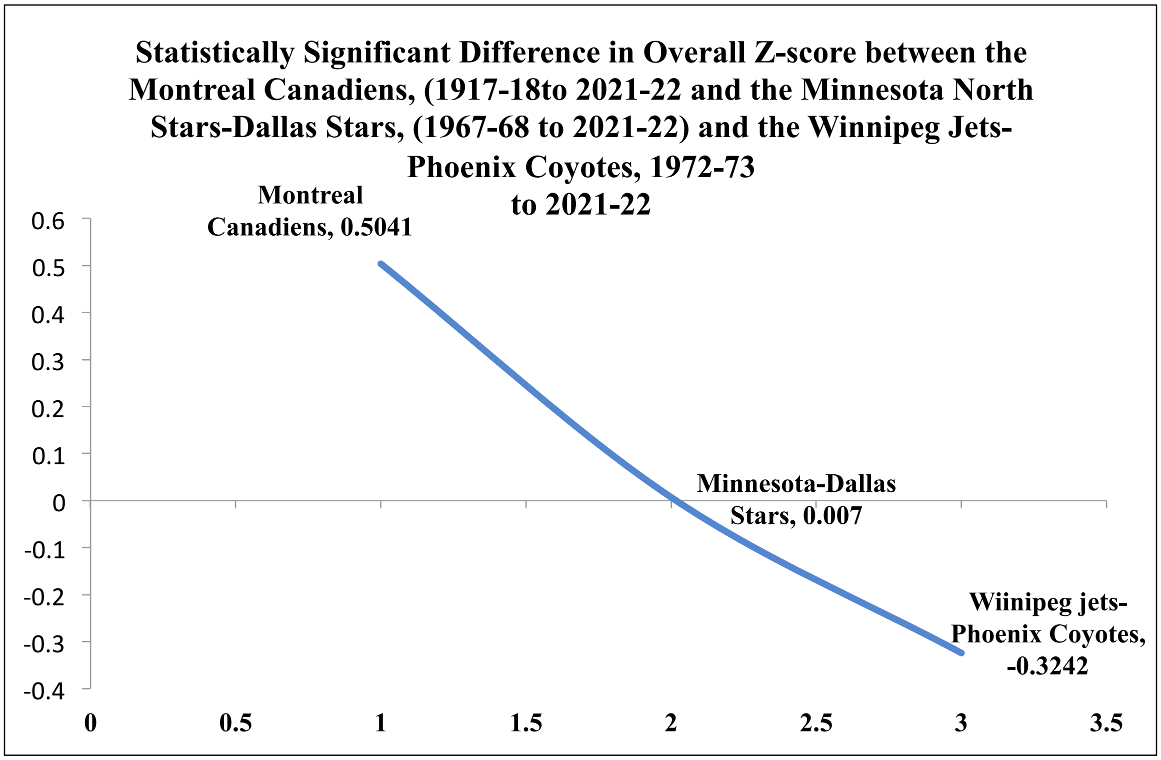
* - The overall ANOVA is statistically significant allowing for the post hoc test.

Scheffé Analysis

Scheffé results			
Treatment Pairs	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B (Montreal vs. Minnesota)	3.1932	0.0069204	**** p<0.01**
A vs. C (Montreal vs. Winnipeg)	5.1468	3.95E-06	**** p<0.01**
B vs. C (Minnesota vs. Winnipeg)	1.8051	0.1986821	Insignificant

The Scheffé test indicates that the Montreal Canadiens are statistically significantly different than the Minnesota-Dallas franchise and the Winnipeg-Phoenix franchise. There is no statistically significant difference between the Minnesota-Dallas franchise and the Winnipeg-Phoenix franchise.

Graph 1 Result:



The Montreal Canadiens are just above the halfway value for positive mediocrity, the Minnesota-Dallas franchise is 7/1000 away from the average score, or in other words an almost perfect average team. The Winnipeg-Phoenix franchise is part of the negative mediocrity group but below the hallway mark in the negative mediocrity group. Statistically, Minnesota-Dallas franchise is statistically the same as the Winnipeg-Phoenix franchise.

Within Team Analysis:
Montreal Canadiens

Descriptive statistics of the 3 independent treatments:				
Categories →	**A (Positive Mediocrity)**	**B (Above Average)**	**C (Negative Mediocrity)**	**Pooled Total**
Seasons	44	30	18	92
Average Z-Score	0.4804	1.4007	-0.4993	0.5888

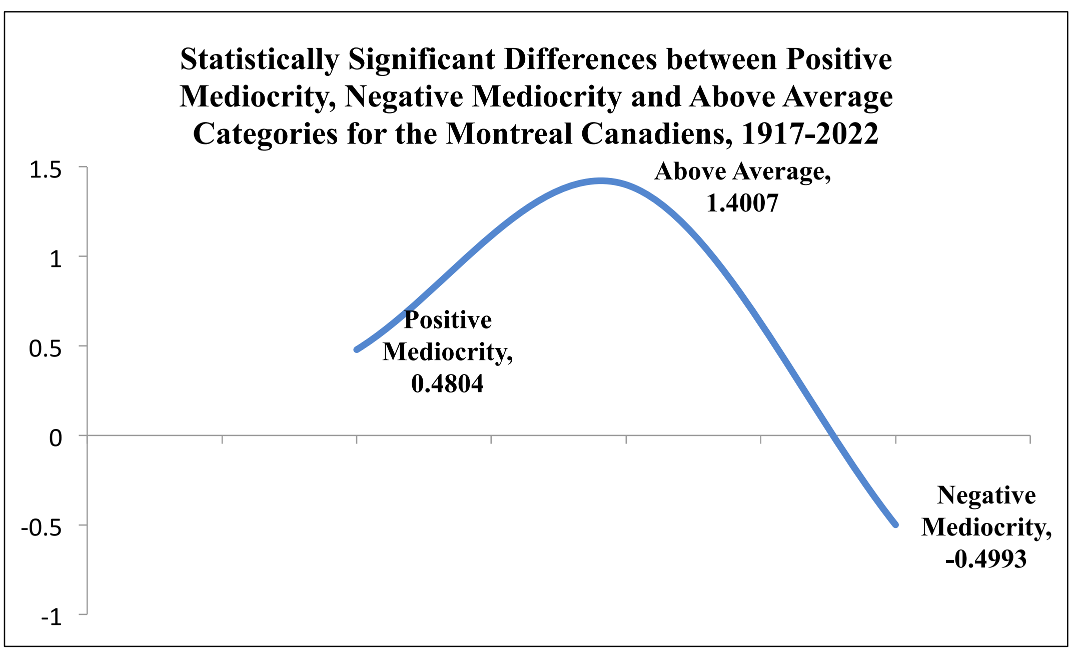
One-way ANOVA of 3 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	41.6051	2	20.8026	254.5424	1.11E-16
Error	7.2736	89	0.0817		
Total	48.8787	91			

*-ANOVA is statistically significant allowing for Scheffé post hoc test.

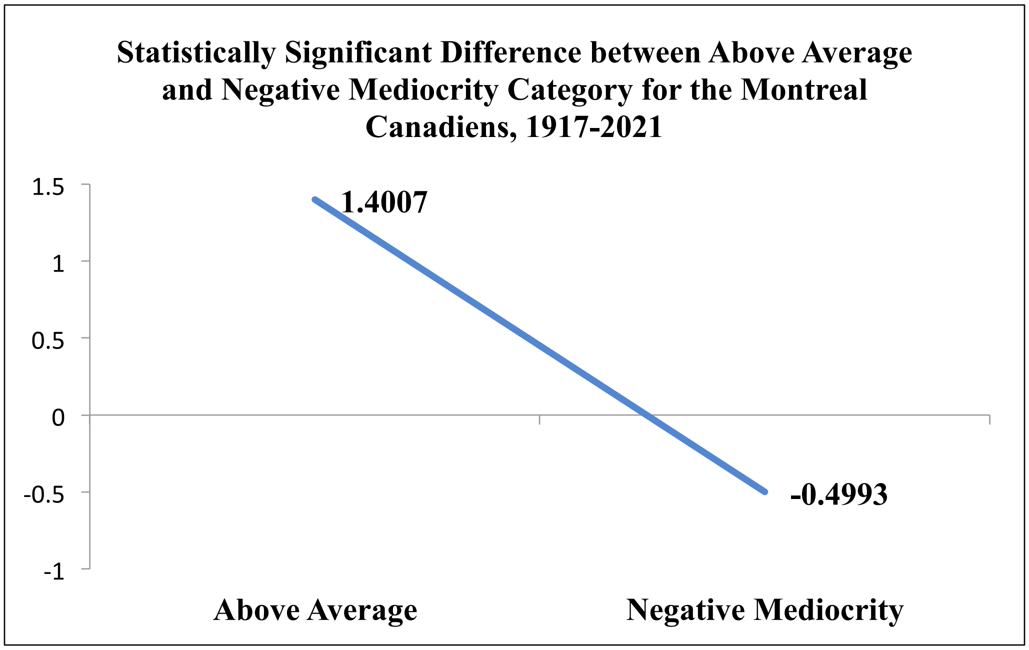
Scheffé results			
Treatment pairs	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B (Positive Mediocrity vs. Above Average)	13.5971	1.11E-16	** p<0.01
A vs. C (Positive Mediocrity vs. Negative Mediocrity)	12.2481	1.11E-16	** p<0.01
B vs. C (Above Average vs. Negative Mediocrity)	22.2924	1.11E-16	** p<0.01

The Scheffé tests indicates that there statistically significant differences between Positive Mediocrity and Negative Mediocrity and Above Average categories. In addition, the Above Average category is significantly different than Negative Mediocrity.

Graph 1 Result:



Graph 2 Results:



Within Team Analysis:
Minnesota North Stars-Dallas Stars

Descriptive statistics of the 2 independent treatments:			
Category →	**A (Negative Mediocrity)**	**B (Positive Mediocrity)**	**Pooled Total**
Seasons	22	19	41
Average Z-Score	-0.4399	0.3258	-0.0851

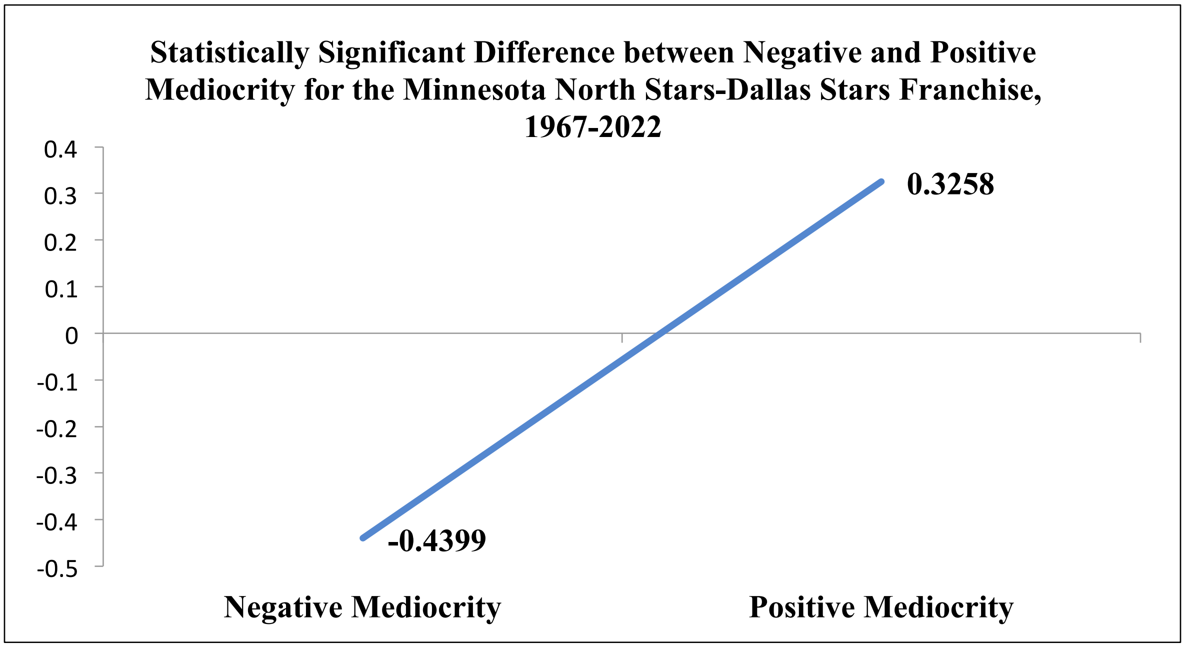
One-way ANOVA of the 2 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean square MS	F statistic	p-value*
Treatment	5.9768	1	5.9768	48.5571	2.37E-08
Error	4.8005	39	0.1231		
Total	10.7773	40			

*Overall, the significance value is at the 99% confidence level allowing for the Scheffé test.

Scheffé results			
Treatment	Scheffé	Scheffé	Scheffé
Pair	TT-statistic	p-value	Inference
A vs. B (Negative Mediocrity vs. Positive Mediocrity)	6.9683	2.37E-08	** $p<0.01$

**-The Scheffé test indicates a statistically significant result at the 99% confidence level.

Graph 1 Result:



There is a statistically significant result between Negative Mediocrity and Positive Mediocrity.

Within Team Analysis
Winnipeg Jets-Phoenix Coyotes

Categories →	**A (Negative Mediocrity)**	**B (Positive Mediocrity)**	**C (Abysmal)**	**Pooled Total**
Seasons	17	13	10	40
Average Z-Score	-0.4331	0.365	-1.3164	-0.3946

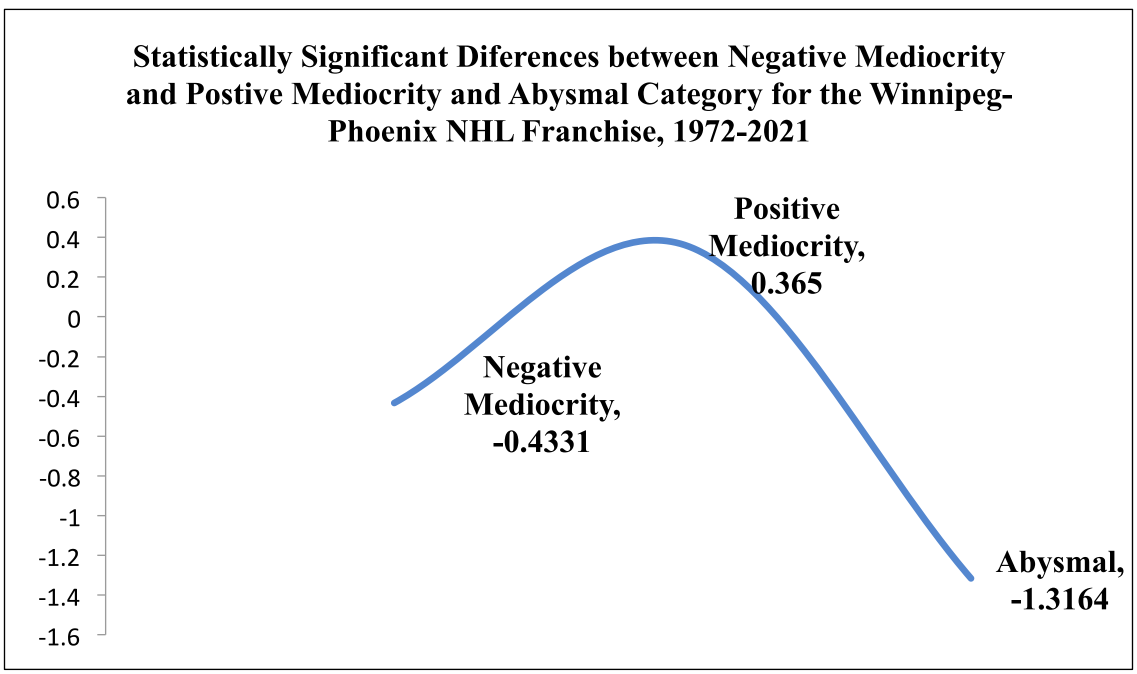
One-way ANOVA of the 3 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	16.0236	2	8.0118	118.0932	1.11E-16
Error	2.5102	37	0.0678		
Total	18.5338	39			

- The overall ANOVA is statistically significant allowing for the Scheffé post hoc test.

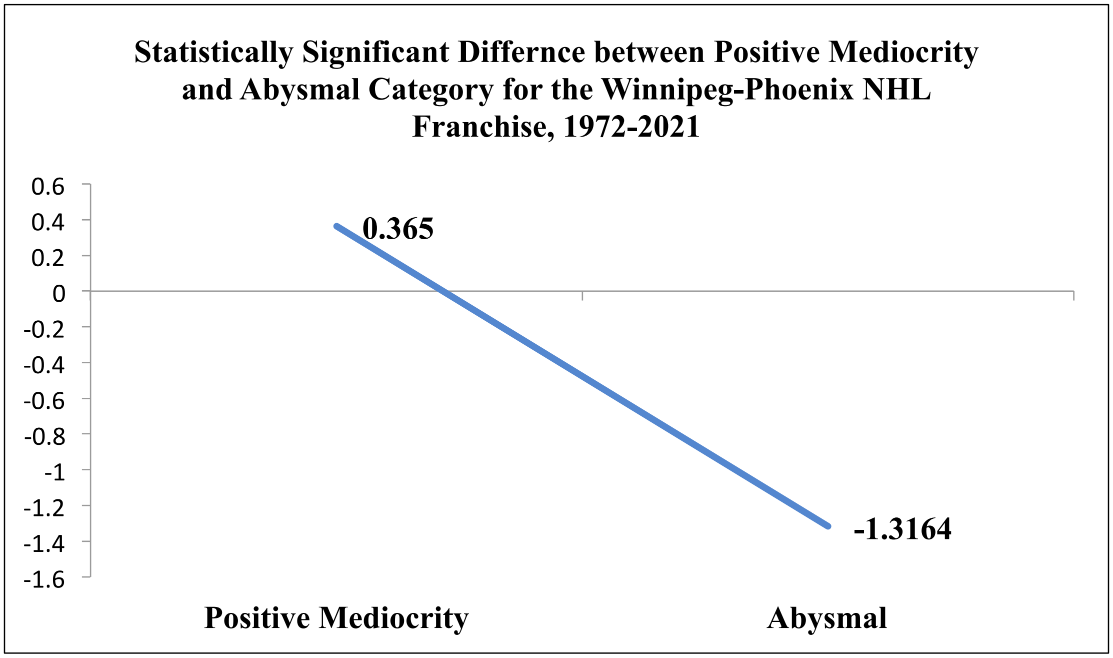
Scheffé Test:

Scheffé results			
Treatment	Scheffé	Scheffé	Scheffé
pairs	TT-statistic	p-value	Inference
A vs. B (Negative Mediocrity vs. Positive Mediocrity)	8.3168	3.39E-09	**** p<0.01**
A vs. C (Negative Mediocrity vs. Abysmal)	8.5093	1.95E-09	**** p<0.01**
B vs. C (Positive Mediocrity vs. Abysmal)	15.3473	1.11E-16	**** p<0.01**

** - The Scheffé results indicate that all categories are statistically significant with each other at the 99% confidence level.



Negative Mediocrity is statistically different than Positive Mediocrity and the Abysmal category.



The Positive Mediocrity category is statistically different than the Abysmal category.

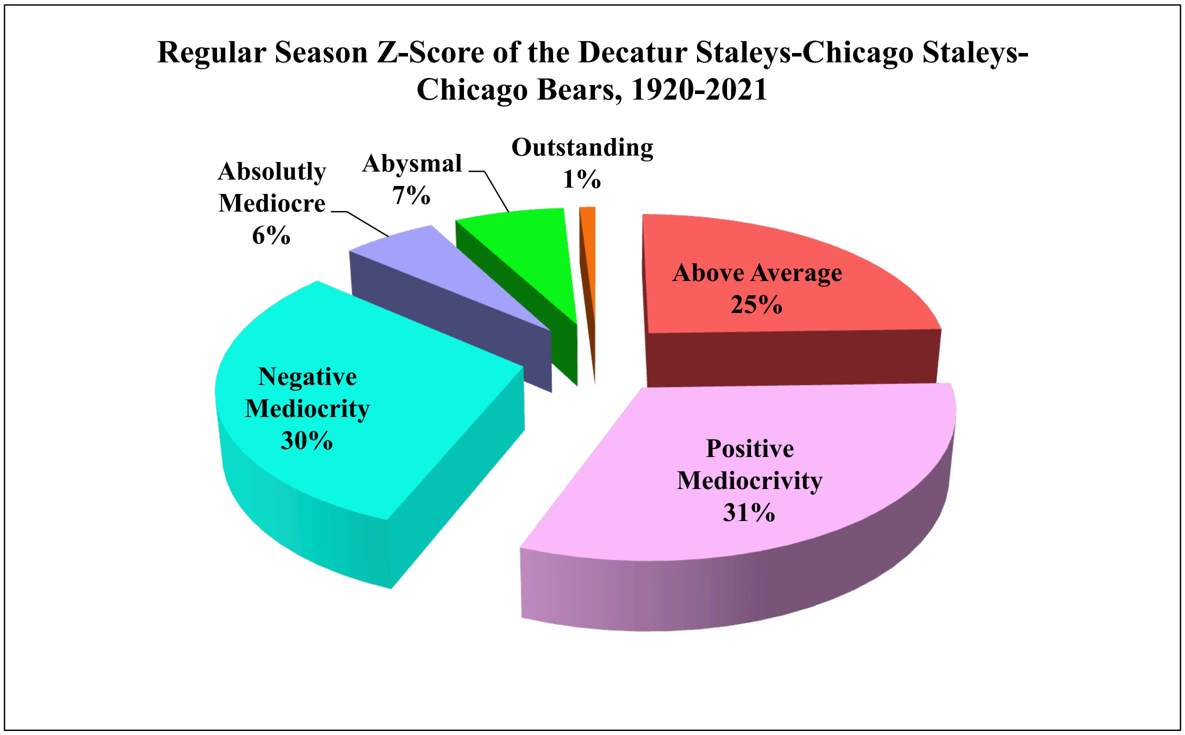
Between Teams Analysis-Positive Mediocrity: There is no statistically significant difference between the three teams.

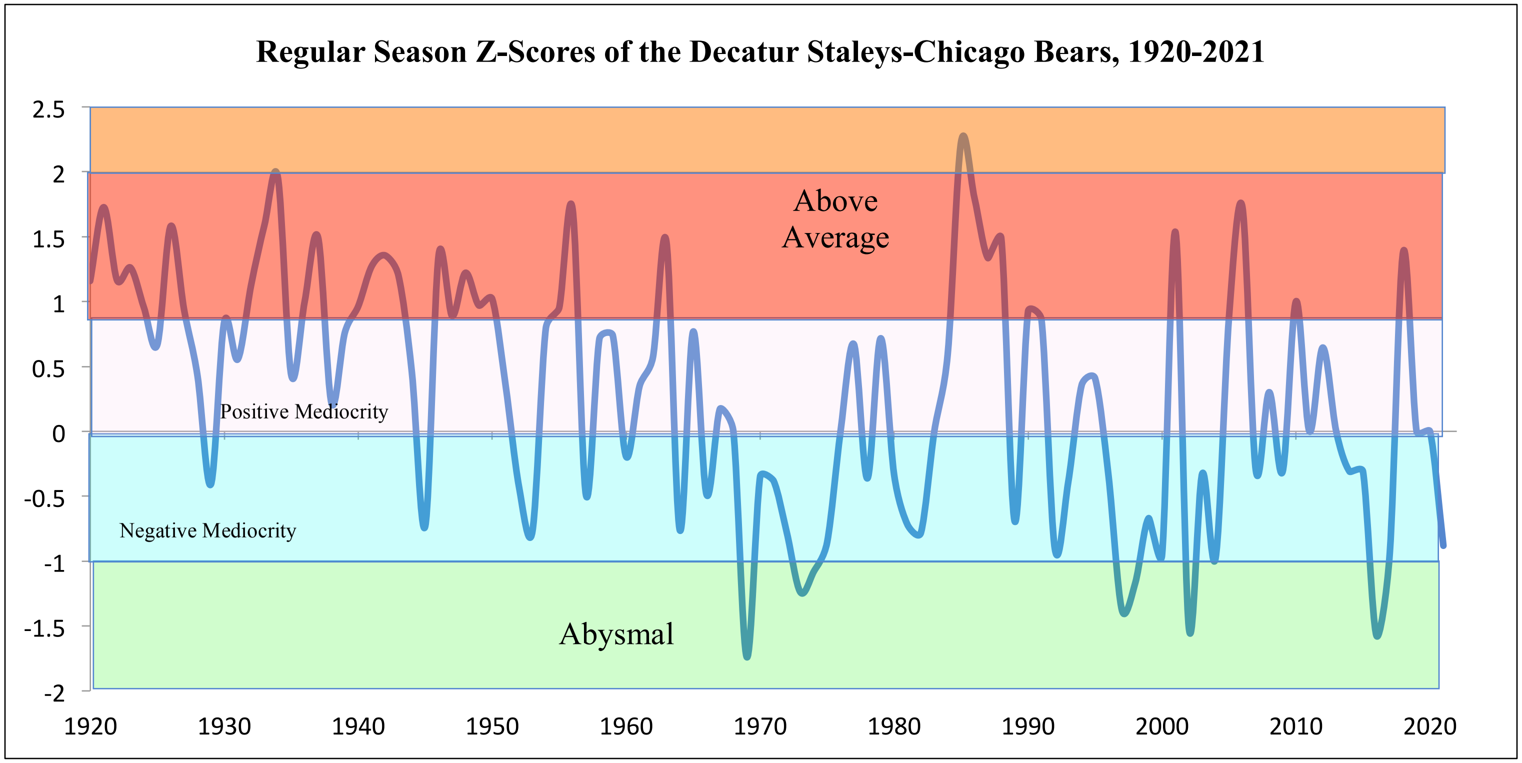
Between Teams Analysis – Negative Mediocrity: There is no statistically significant difference between the three teams.

Conclusion: Overall the Montreal Canadiens are statistically different than the Minnesota-Dallas and the Winnipeg-Phoenix franchise. There is no statistically significant difference between the Minnesota-Dallas franchise and the Winnipeg-Phoenix franchise. They are both within the same category.

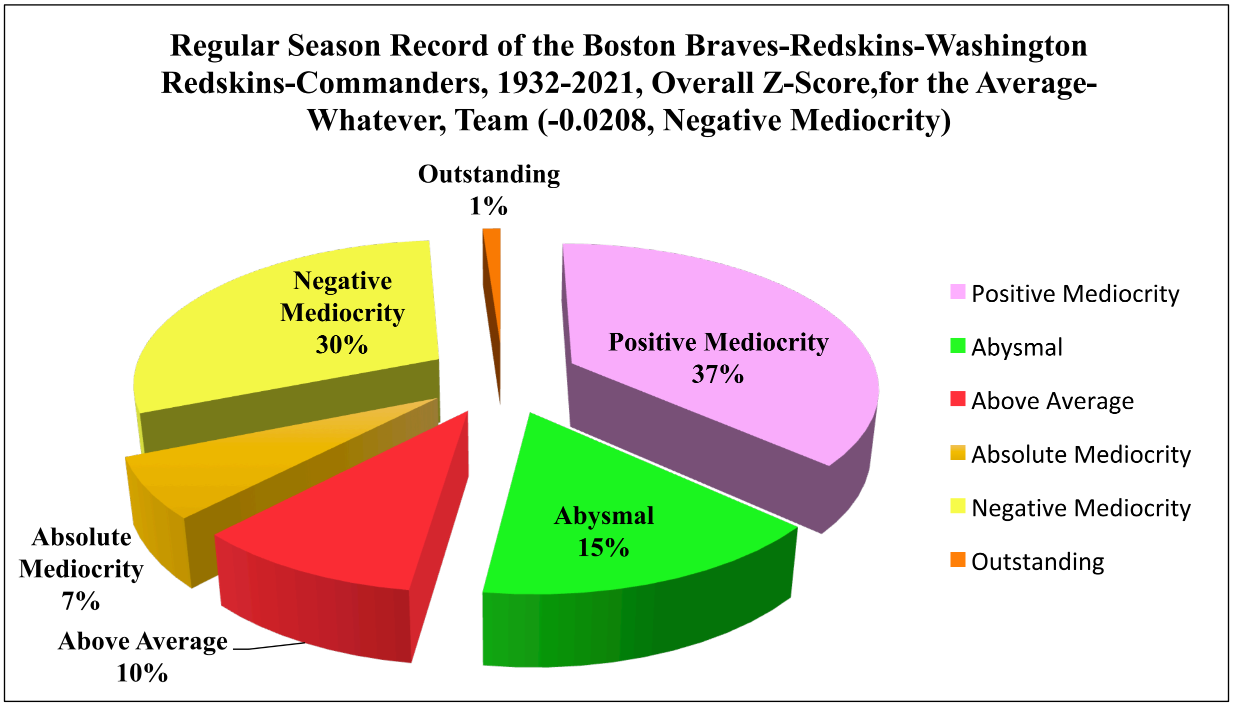
NFL: The NFL officially came into to existence after 2 seasons as the AAPC in 1922. We will see if the same pattern continue for the best, worst and average teams across the existence of the league. The data from the AAPC will be taken into consideration for the analysis.

NFL Best Overall Team: Decatur Staleys-Chicago Bears

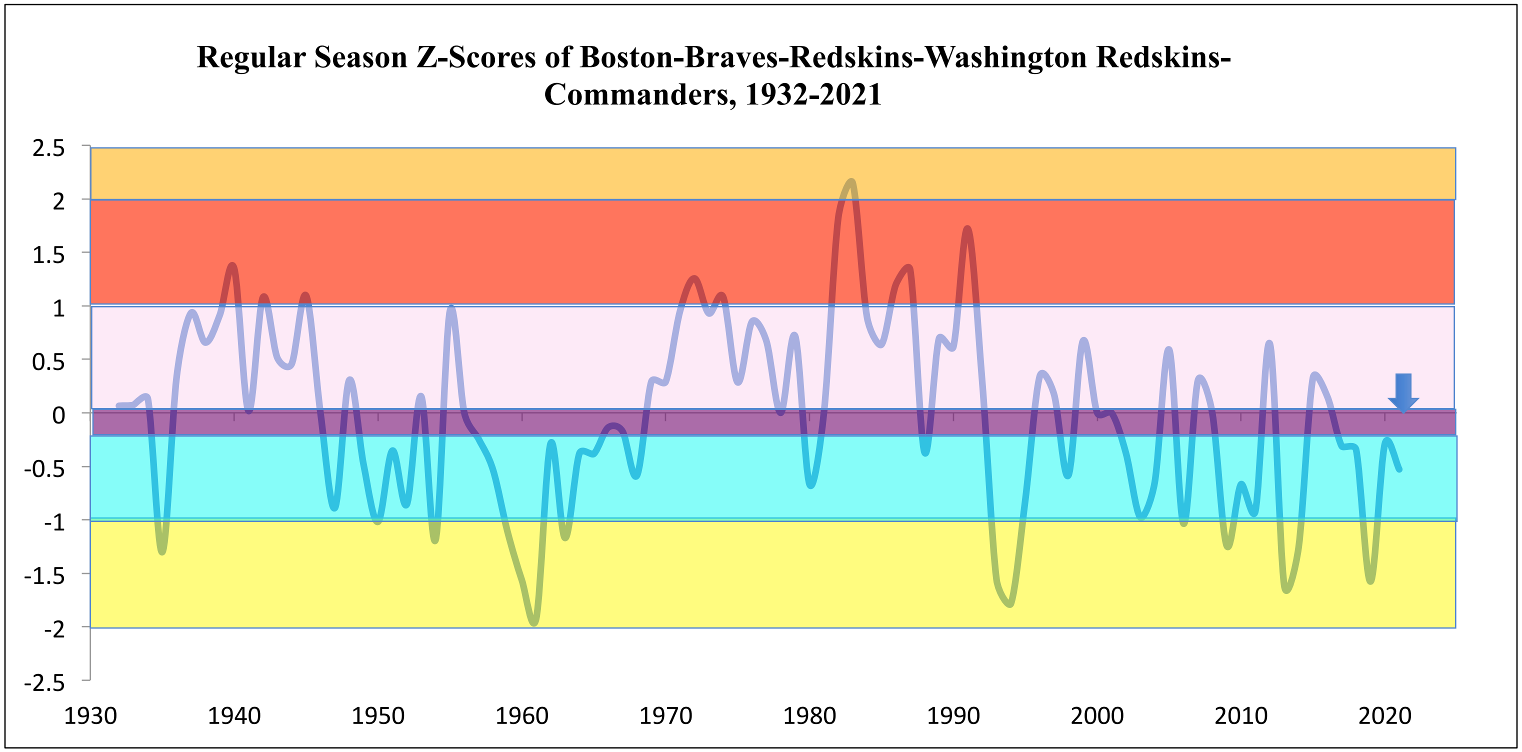



Regular Season Z-Scores of the Decatur Staleys-Chicago Bears, 1920-2021
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Above Average
Positive Mediocrity
Negative Mediocrity
Abysmal

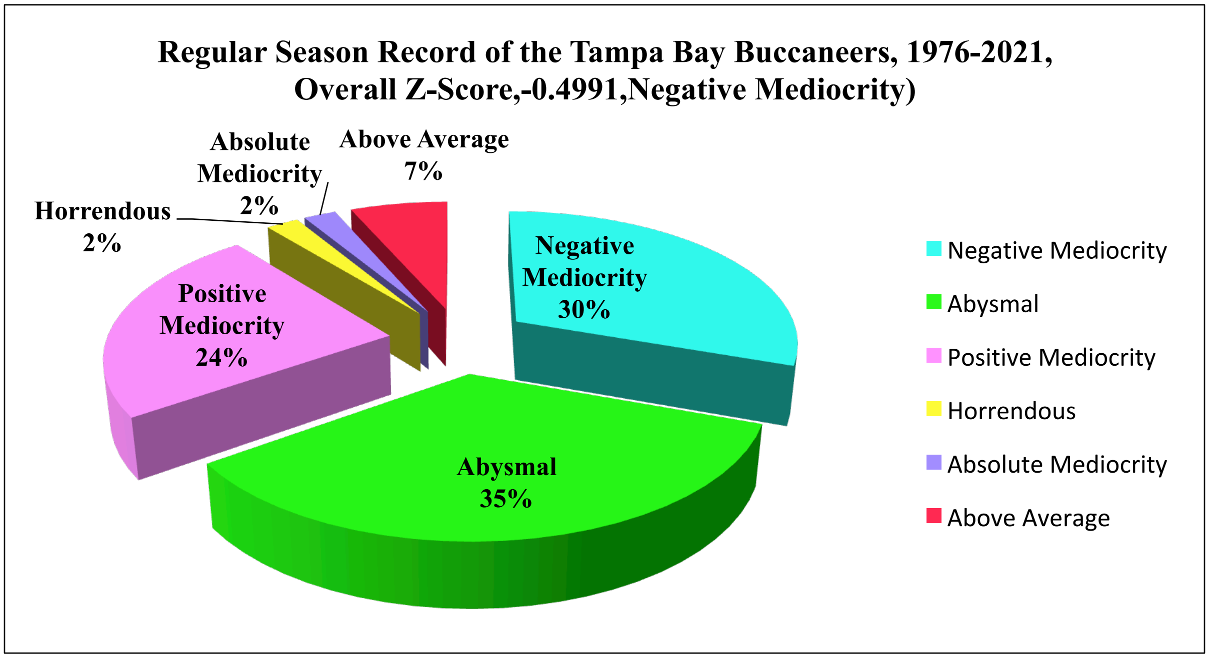
The Average-Whatever NFL Team: Boston Braves-Redskins-Washington Redskins-Commanders



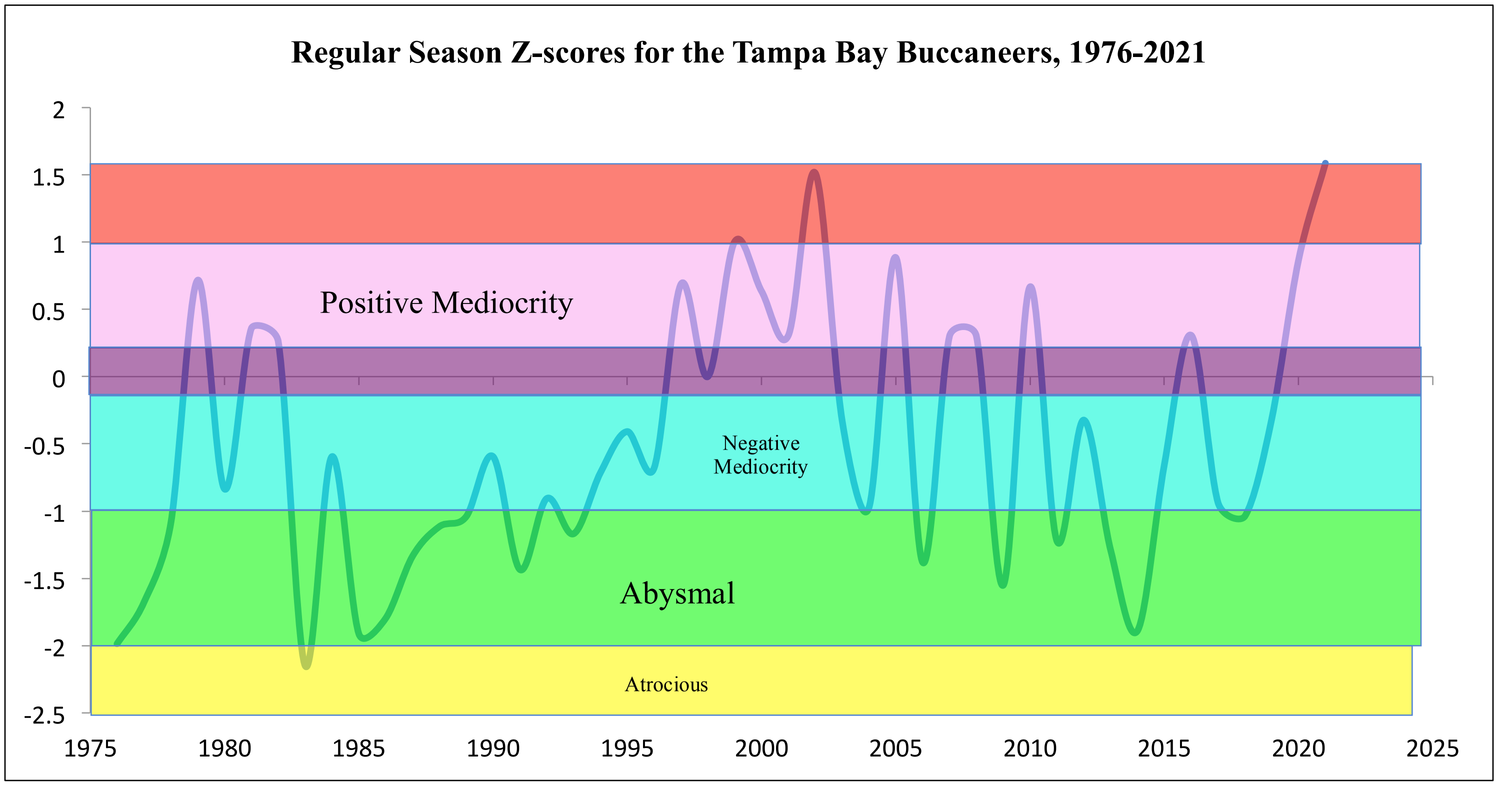
The average Washington team has occupied six categories with mediocrity occupying 67.


Regular Season Z-Scores of Boston-Braves-Redskins-Washington Redskins-Commanders, 1932-2021
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020

The Overall Worst NFL Team: Tampa Bay Buccaneers



The Tampa Bay franchise has spent 69% of their existence at zero or below negative mediocrity. Thank heaven for Tom Brady!


Regular Season Z-scores for the Tampa Bay Buccaneers, 1976-2021
Positive Mediocrity
Negative Mediocrity
Abysmal
Atrocious
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025

Analysis between the Three NFL groups:

Descriptive statistics of the 3 independent treatments:				
Categories →	A (Decatur Staleys-Chicago Staleys-Bears	B (Boston Braves-Redskins,Washington Redskins-Commanders)	C (Tampa Bay Buccaneers)	Pooled Total
Seasons	102	90	46	238
Average Z-Score	0.2993	-0.0209	-0.4991	0.0239

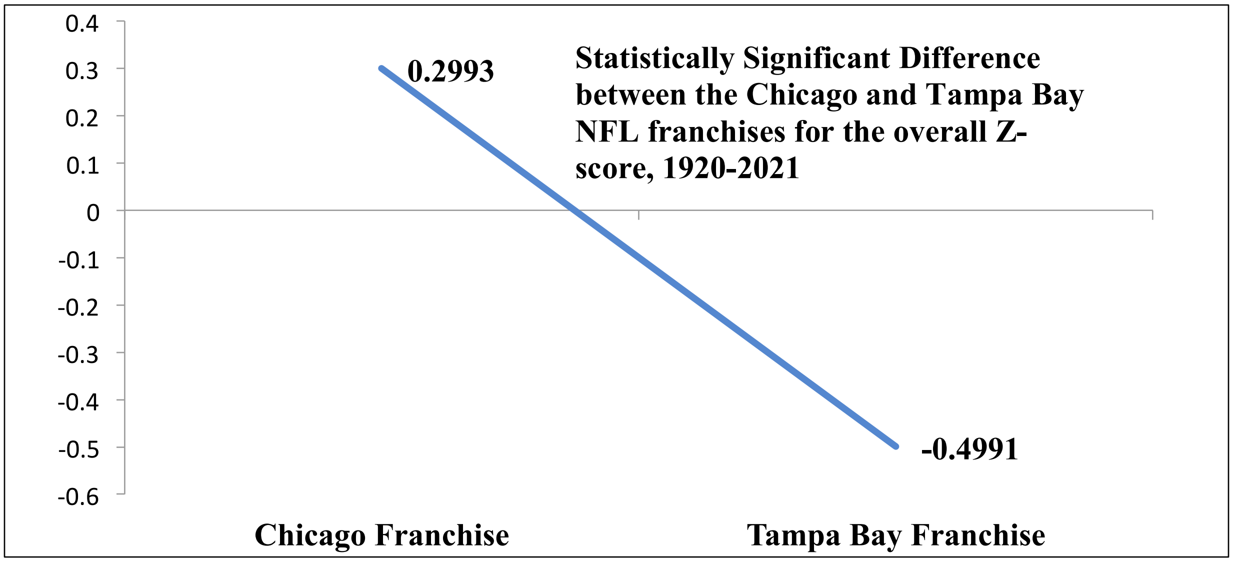
One-way ANOVA of the 3 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	20.497	2	10.2485	12.1004	9.96E-06
Error	199.0339	235	0.847		
Total	219.5309	237			

*-The overall ANOVA is statistically significant allowing for the post hoc evaluation.

Scheffé results			
Treatment	Scheffé	Scheffé	Scheffé
Pairs	TT-statistic	p-value	Inference
A vs. B (Chicago vs. Washington)	2.4051	0.0574302	Insignificant
A vs. C (Chicago vs. Tampa Bay)	4.8845	1.16E-05	** $p<0.01$
B vs. C (Washington vs. Tampa Bay)	2.8673	0.0175896	* $p<0.05$

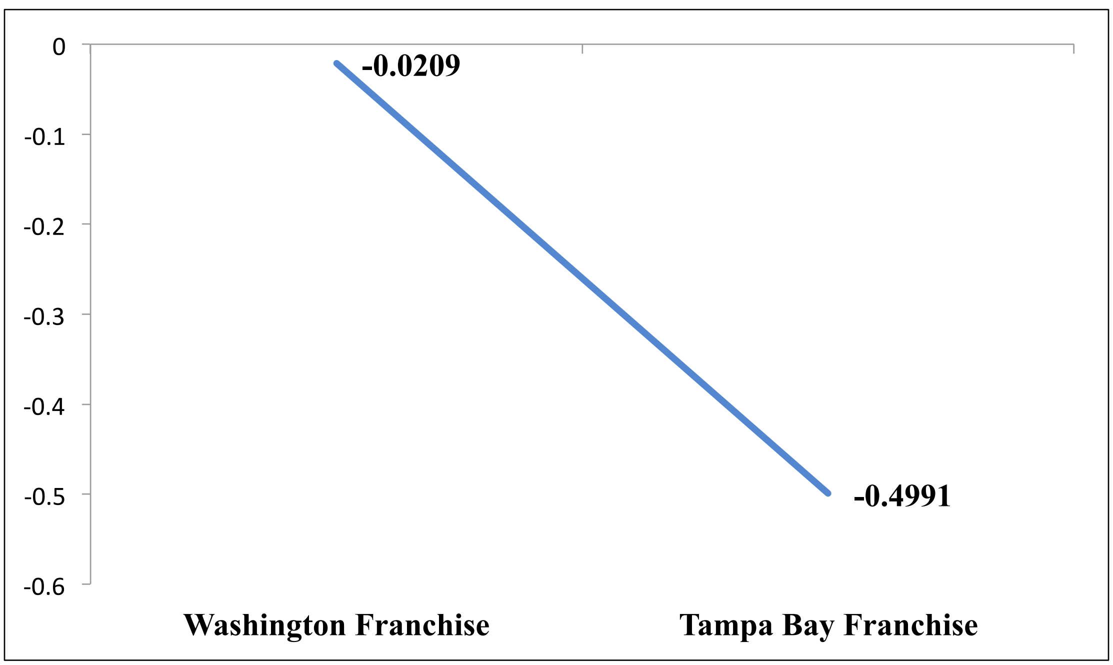
**- There is a statistically significant result between the Chicago franchise and the Tampa Bay franchise at the 99% confidence level.
*-There is a statistically significant result between the Washington franchise and the Tampa Bay franchise at the 95% confidence level
The results allow for the post hoc analysis.

Graph 1 Result:



Both teams are situated in the mediocre category however the Chicago franchise is in the positive area of mediocrity and the Tampa Bay franchise is close to the halfway mark of negative mediocrity.

Graph 2 Results:



The Washington franchise is statistically different than the Tampa franchise. The Washington franchise is only 2% from absolute zero while the Tampa Bay franchise is close to -5.

Within Team Analysis: Decatur-Chicago Franchise

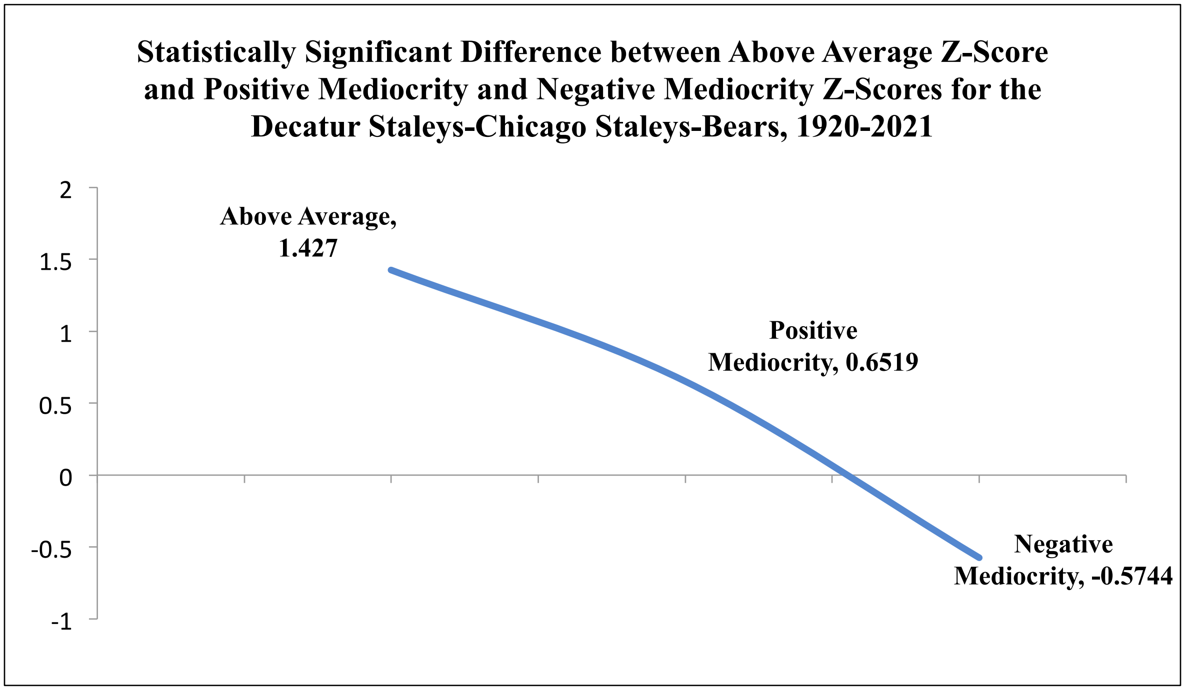
Descriptive statistics of the 3 independent treatments:				
Categories →	**A (Above Average)**	**B (Positive Mediocrity)**	**C (Negative Mediocrity)**	**Pooled Total**
Seasons	26	32	32	90
Average Z-Score	1.427	0.6519	-0.5744	0.4398

One-way ANOVA of the 3 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	59.6967	2	29.8483	450.4879	1.11E-16
Error	5.7644	87	0.0663		
Total	65.4611	89			

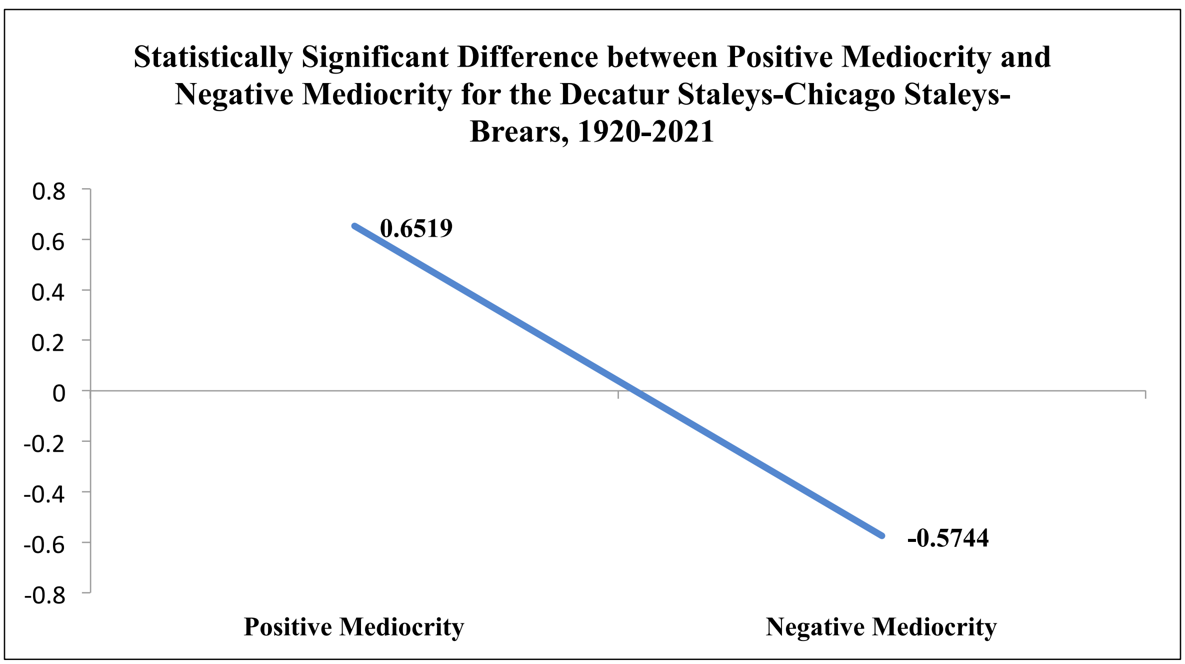
Scheffé results			
Treatment Pairs	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B (Above Average vs. Positive Mediocrity)	11.4051	1.11E-16	** **p<0.01**
A vs. C (Above Average vs. Negative Mediocrity)	29.4493	1.11E-16	** **p<0.01**
B vs. C (Positive Mediocrity vs. Negative Mediocrity)	19.0568	1.11E-16	** **p<0.01**

** - There are statistically significant results between each of the categories at the 99% confidence level.

Graph 1 Result:



Graph 2 Results:



Within Team Analysis-Washington Franchise

Descriptive statistics of the 3 independent treatments:				
Categories →	**A (Positive Mediocrity)**	**B (Abysmal)**	**C (Negative Mediocrity)**	**Pooled Total**
Seasons	29	14	27	70
Average Z-Score	0.5077	-1.3816	-0.4641	-0.245

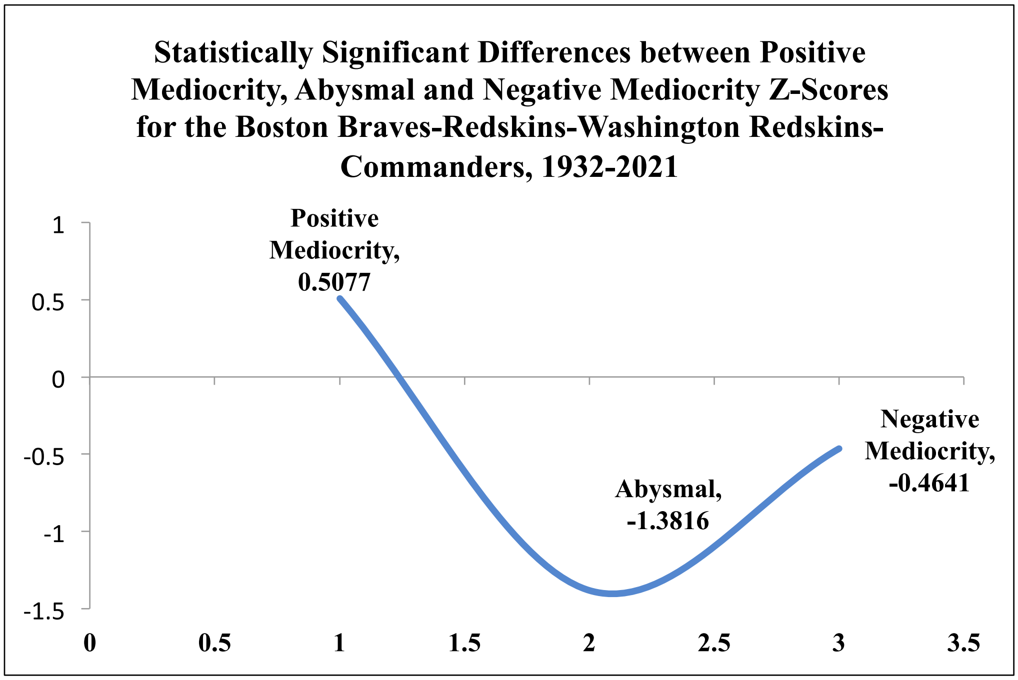
One-way ANOVA of the 3 independent treatments:					
	Sum of	Degrees of	Mean Square		
Source	Squares SS	Freedom df	MS	F statistic	p-value*
Treatment	35.8129	2	17.9064	201.4176	1.11E-16
Error	5.9564	67	0.0889		
Total	41.7693	69			

* - P value is less than .05 indicating significance and allowing for post hoc tests.

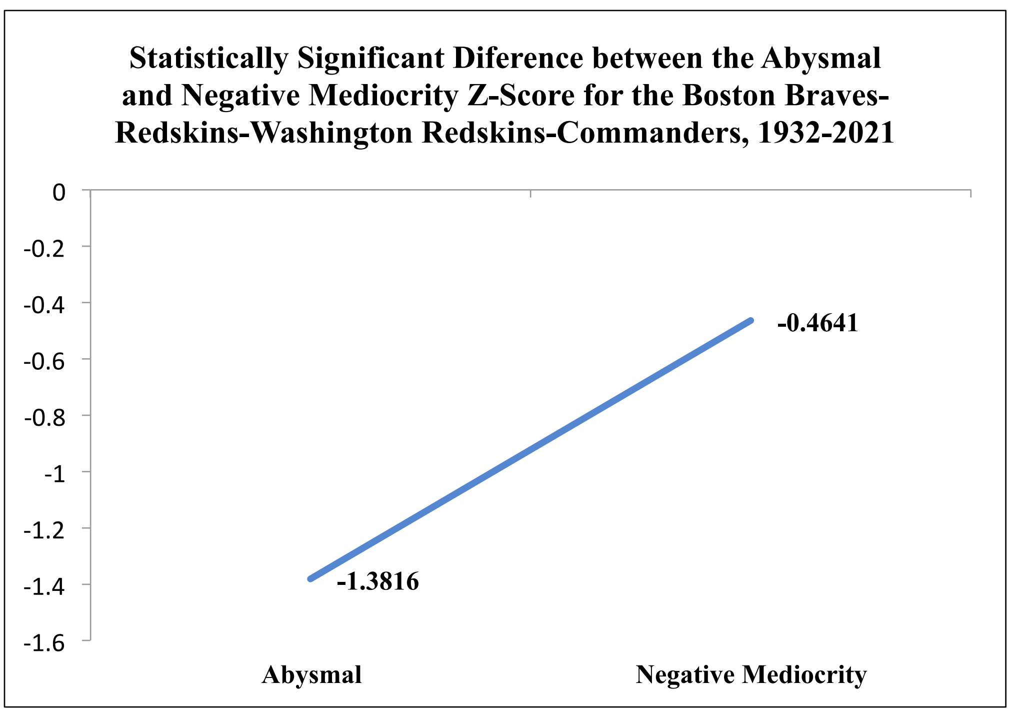
Scheffé results			
Treatment	Scheffé	Scheffé	Scheffé
Pairs	TT-statistic	p-value	Inference
A vs. B (Positive Mediocrity vs. Abysmal)	19.4706	1.11E-16	** p<0.01
A vs. C (Positive Mediocrity vs. Negative Mediocrity)	12.187	1.11E-16	** p<0.01
B vs. C (Abysmal vs. Negative Mediocrity)	9.3439	7.29E-13	** p<0.01

** - Comparisons between the categories are statistically significant at the 99% confidence level.

Graph Result 1:



Graph Results 2:



Within Team Analysis – Tampa Bay Buccaneers

Descriptive statistics of the 3 independent treatments:				
Categories →	**A (Abysmal)**	**B (Positive Mediocrity)**	**C (Negative Mediocrity)**	**Pooled Total**
Seasons	16	11	13	40
Average Z-Score	-1.4329	0.4937	-0.6356	-0.644

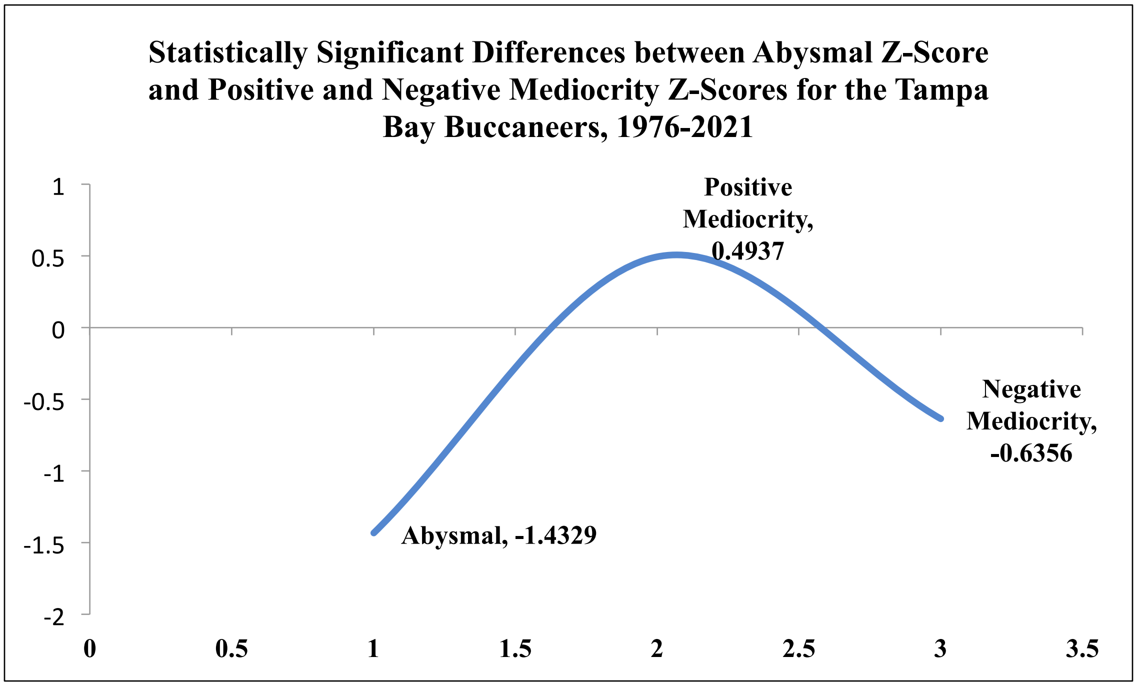
One-way ANOVA of the 3 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	24.1975	2	12.0987	158.9633	1.11E-16
Error	2.8161	37	0.0761		
Total	27.0136	39			

*-p value is less than .05 and therefore the post hoc test can be applied

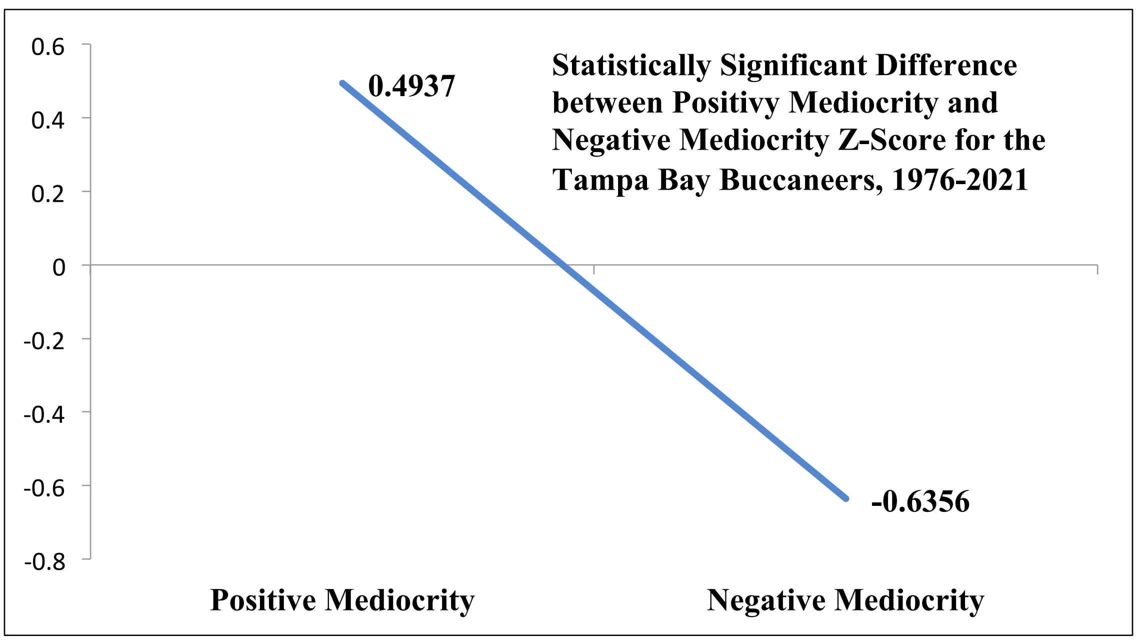
Scheffé results			
Treatment Pairs	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B (Abysmal vs. Positive Mediocrity)	17.83	1.11E-16	**** p<0.01**
A vs. C (Abysmal vs. Negative Mediocrity)	7.7402	1.84E-08	**** p<0.01**
B vs. C (Positive Mediocrity vs. Negative Mediocrity)	9.9919	3.10E-11	**** p<0.01**

** Post hoc test is significant at the 99% confidence level indicating statistical significance for all comparisons.

Graph Result 1:



Graph Results 2:



The Positive Mediocrity Z-score is close to the halfway mark for positive mediocrity and the Negative Mediocrity score is more than halfway towards the Abysmal Category.

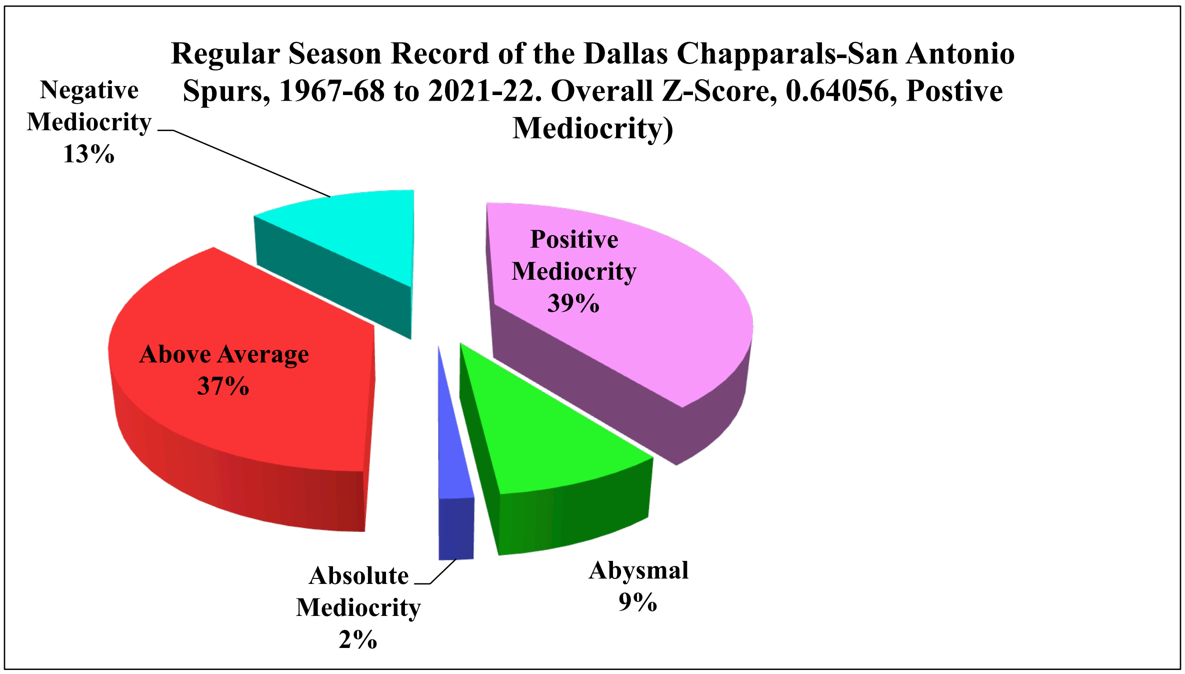
Conclusion: The three selected teams are distinctly different than each other despite the lower score for the Chicago franchise. The teams show similar patterns to the MLB and the NHL.

NBA: We now look at the youngest of the North American Professional team sports, the NBA. The NBA was a result of the merger of two other professional basketball leagues and came into existence in 1947. Let's see if the NBA follows the same as the other three leagues.

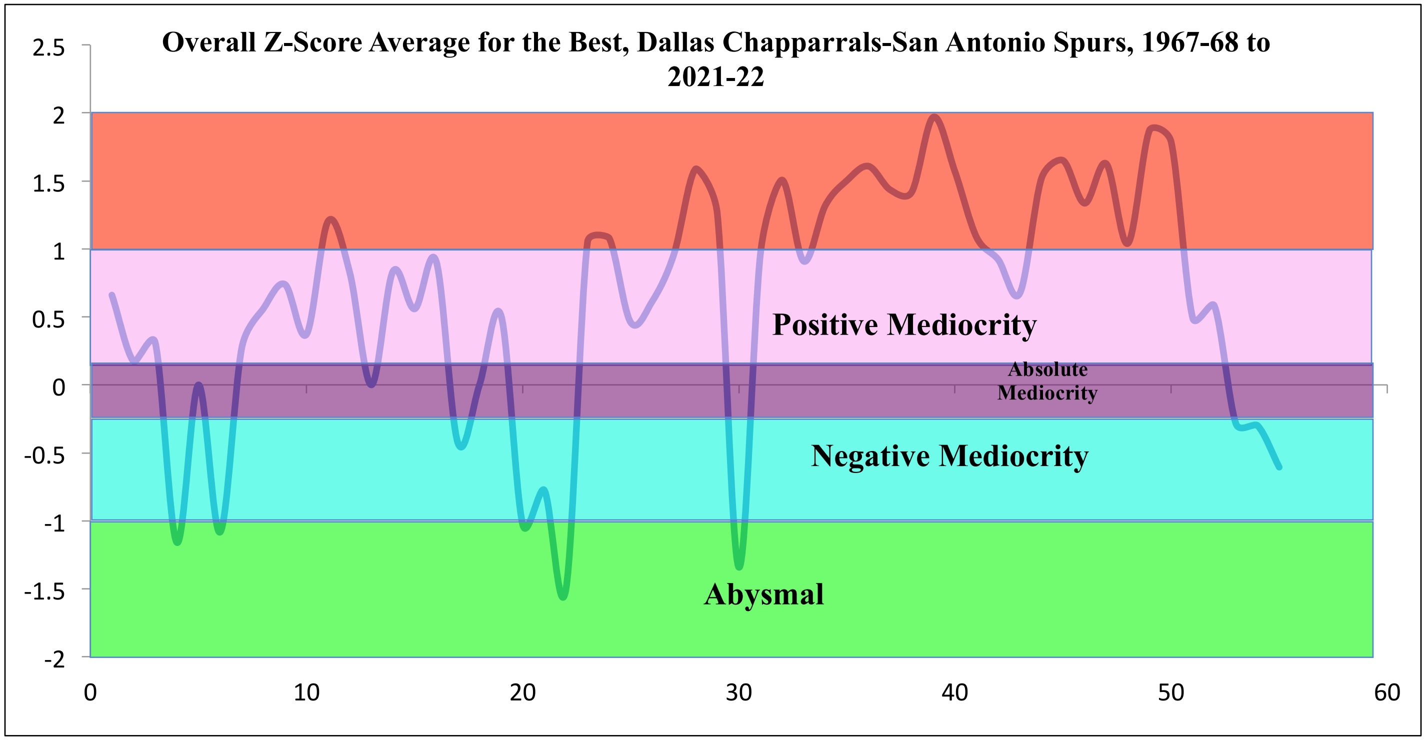
Analysis of the Best, the Worst and Average Basketball Teams

Using the sample size of 30 for the number of seasons played, a total of 27 teams qualify. The top franchise are the Dallas Chaparral's-San Antonio Spurs with a Z-score of 0.640559687, this is the highest o all teams from all leagues. The average team is the Chicago Bulls with a Z-score of -0.004845998. The team is 4/1000 away from the perfect average score. The team categorized as the worst overall team is the Minnesota Timberwolves with a Z-score of -0.616641093. They are almost a full standard deviation away from the best team.

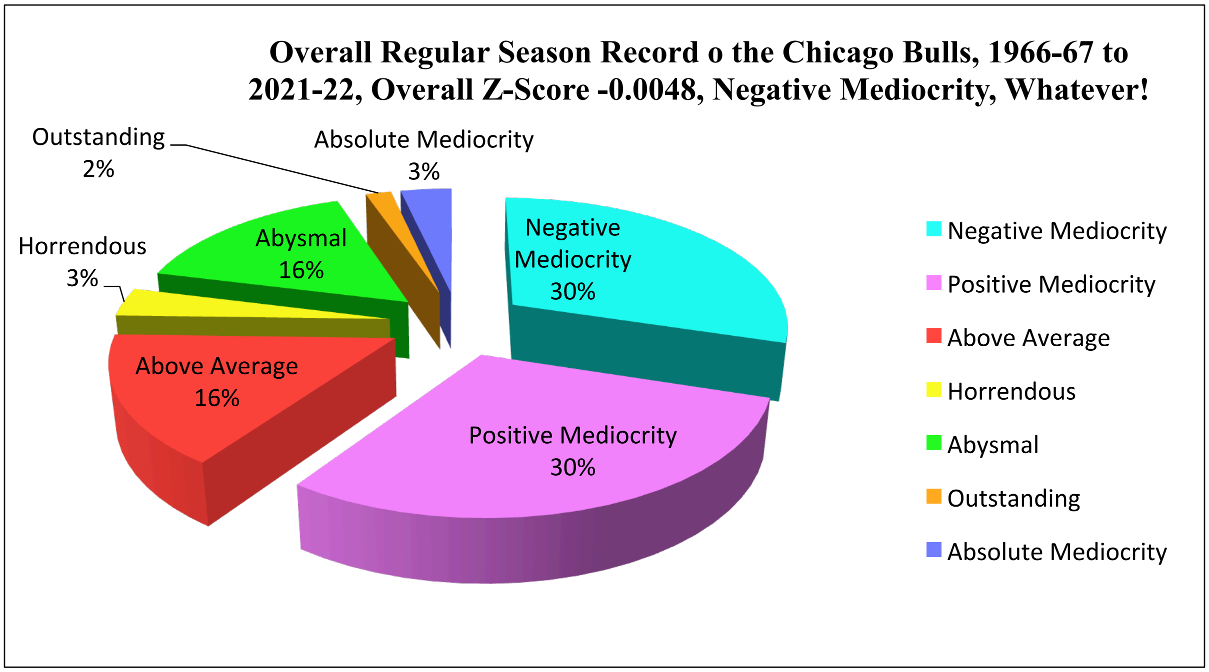
The Best - Dallas Chaparrals-San Antonio Spurs



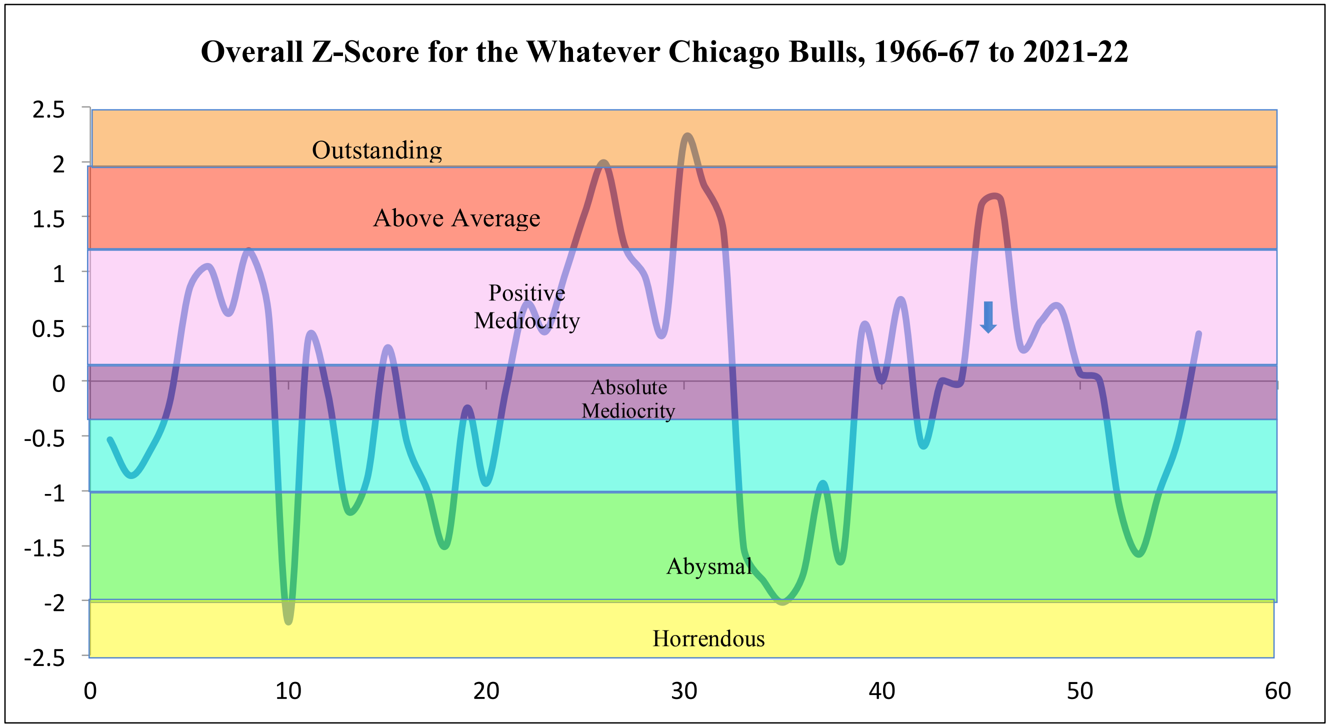
The franchise has played at 76% above mediocrity for the franchise.


Overall Z-Score Average for the Best, Dallas Chapparrals-San Antonio Spurs, 1967-68 to 2021-22
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Positive Mediocrity
Absolute Mediocrity
Negative Mediocrity
Abysmal

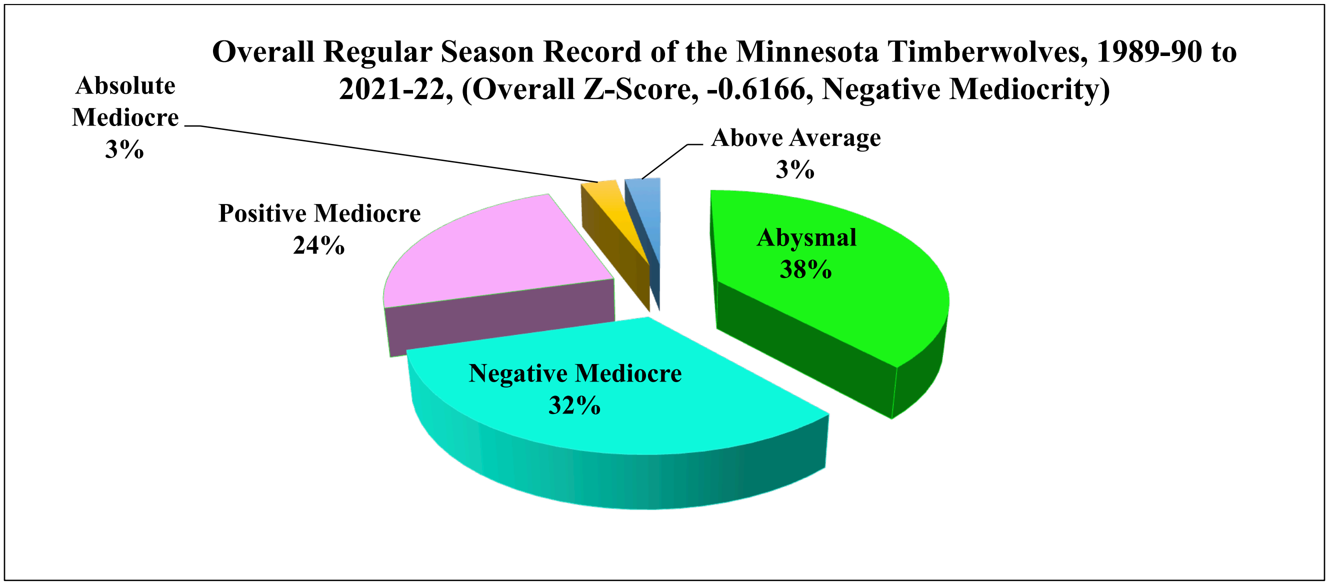
The Whatever, Most Average NBA Franchise-Chicago Bulls



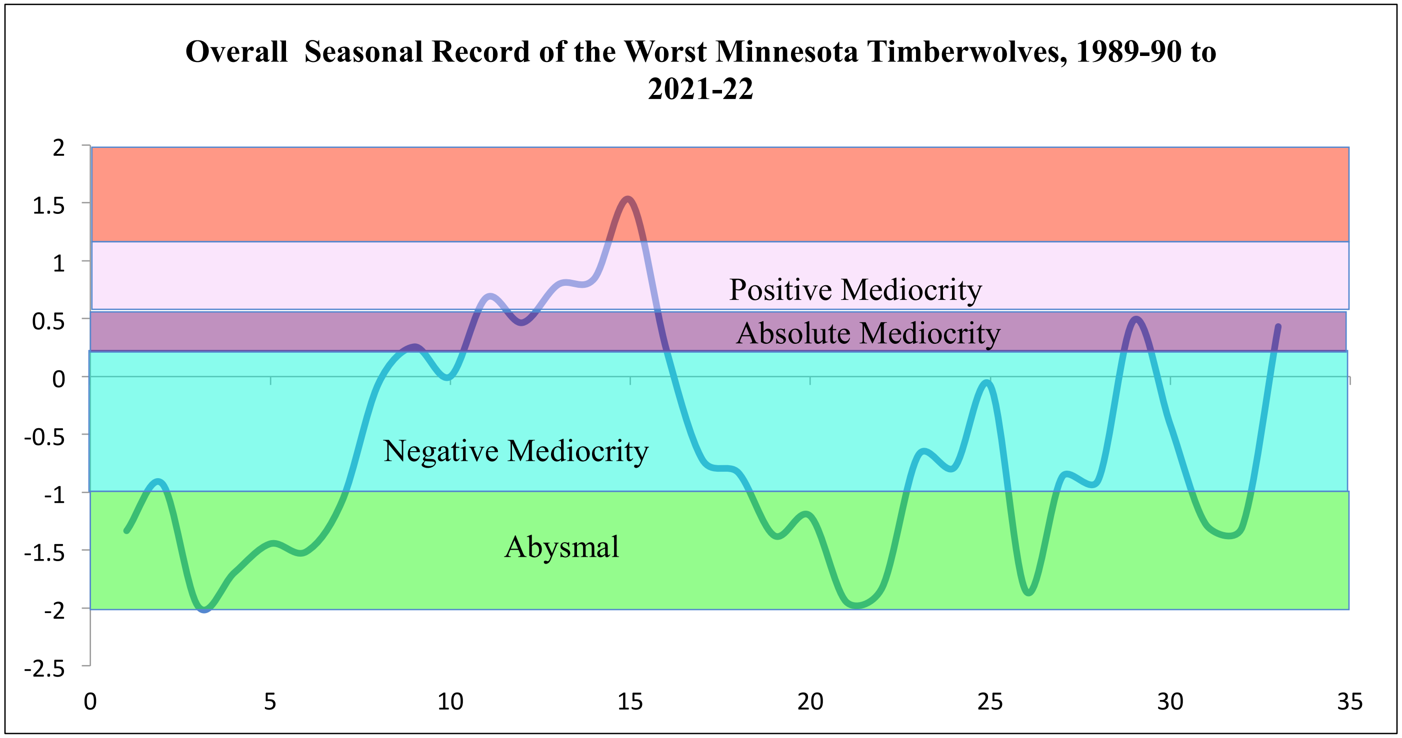
The Whatever Chicago Bulls have in their existence occupied 7 different performance categories. 52% of their existence has been in the negative performance. For 48% of the time, the team has been in the positive performance categories.


Overall Z-Score for the Whatever Chicago Bulls, 1966-67 to 2021-22
Outstanding
Above Average
Positive Mediocrity
Absolute Mediocrity
Abysmal
Horrendous
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
10
20
30
40
50
60

The Worst NBA Franchise, Minnesota Timberwolves



The franchise that has not won a championship or finished in first place has spent70% below the average.


Overall Seasonal Record of the Worst Minnesota Timberwolves, 1989-90 to 2021-22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
5
10
15
20
25
30
35
Positive Mediocrity
Absolute Mediocrity
Negative Mediocrity
Abysmal

Analysis of the Best, Worst and Whatever Teams

Descriptive statistics of the 3 independent treatments:				
Categories →	**A (Dallas Chaparrals-San Antonio Spurs)**	**B (Chicago Bulls)**	**C (Minnesota Timberwolves)**	**Pooled Total**
Seasons	56	57	34	147
Average Z-Score	0.6406	-0.0048	-0.6092	0.1012

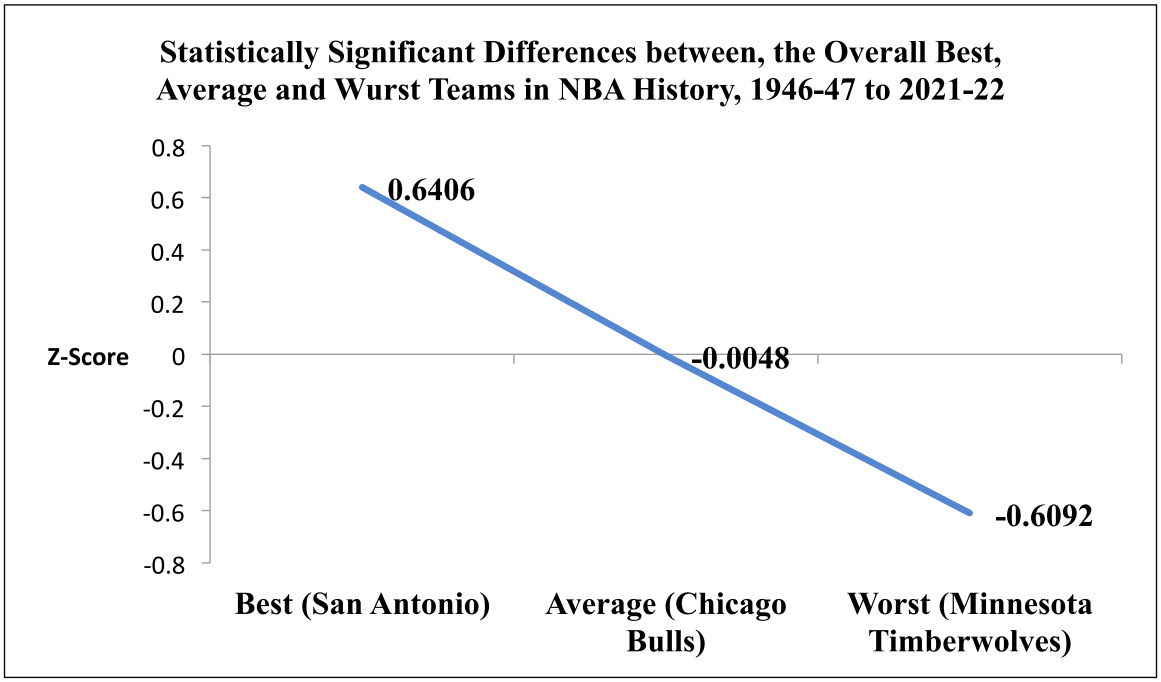
One-way ANOVA of the 3 independent treatments:					
Source	**Sum of Squares SS**	**Degrees of Freedom df**	**Mean Square MS**	**F statistic**	**p-value***
Treatment	34.091	2	17.0455	17.9953	1.06E-07
Error	136.4	144	0.9472		
Total	170.491	146			

*- The p-value is less than .01 indicating an overall statistical significance and allowing for the post hoc test.

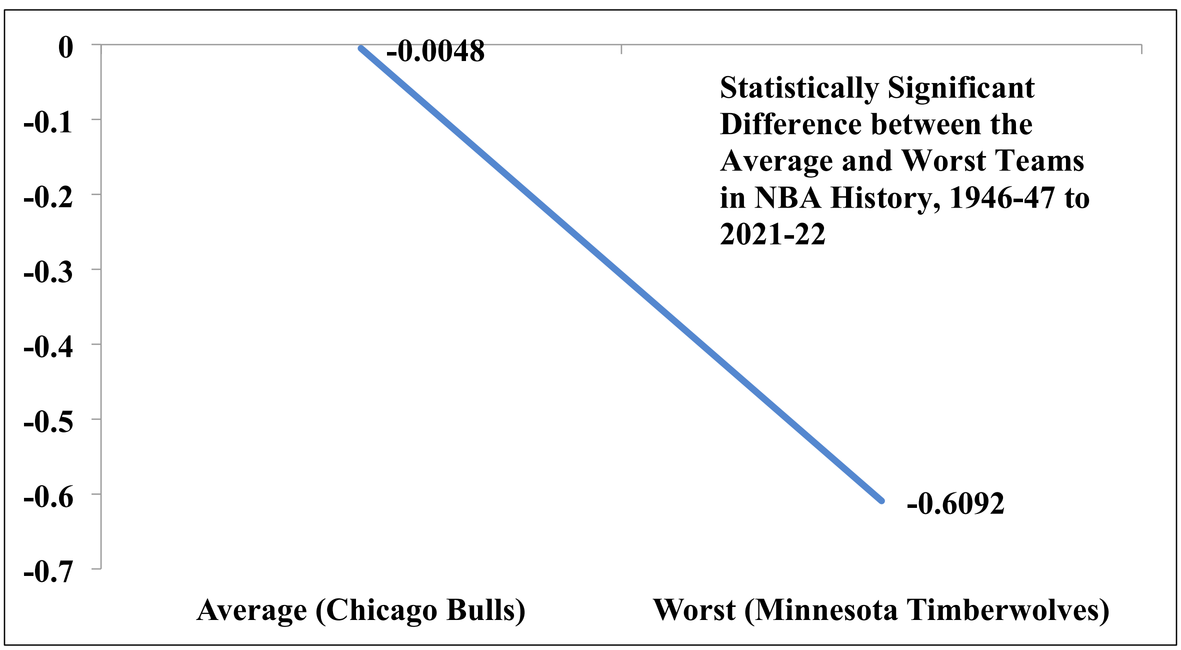
Scheffé results			
Treatment Pairs	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B (Dallas-San Antonio vs. Chicago)	3.5245	0.0025859	**** p<0.01**
A vs. C (Dallas-San Antonio vs. Minnesota)	5.9063	1.65E-07	**** p<0.01**
B vs. C (Chicago vs. Minnesota)	2.8657	0.0184397	*** p<0.05**

** - p value is less than .01 indicating a 99% confidence level.
*- p value is less than .05 indicating a 95% confidence level

Graph Result 1:



Graph Results 2:



Within Teams Analysis-Dallas Chaparrals-San Antonio Spurs

Descriptive statistics of 2 independent treatments:			
Categories →	A (Positive Mediocrity)	B (Above Average)	Pooled Total
Seasons	21	21	42
Average Z-Score	0.6379	1.4491	1.0435

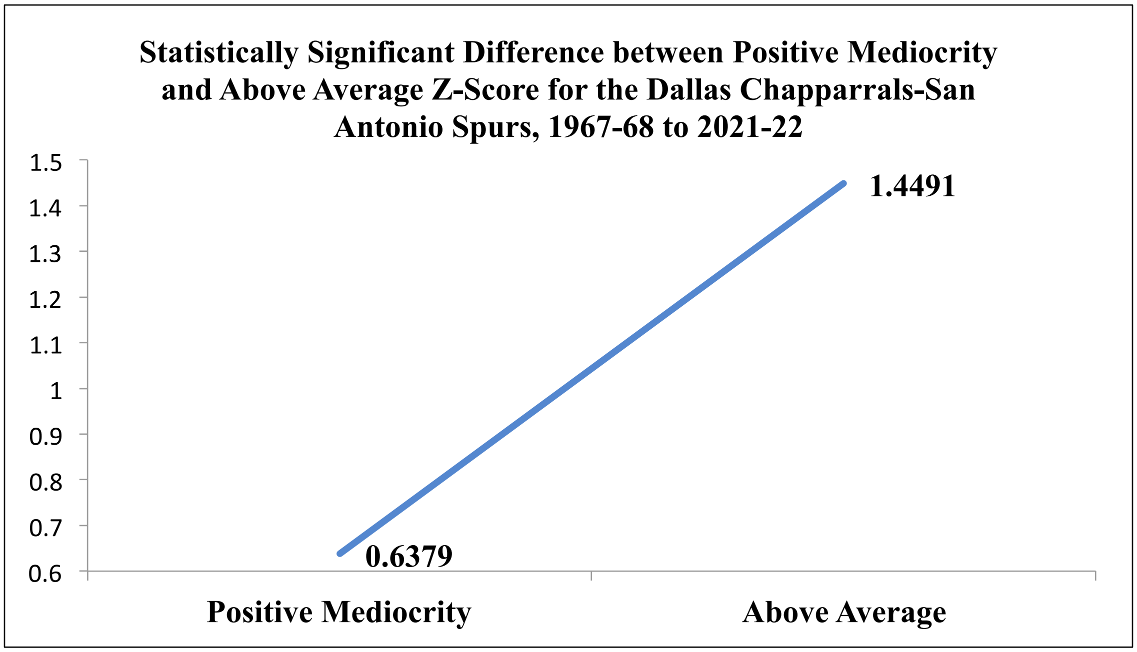
One-way ANOVA of 2 independent treatments:					
Source	Sum of Squares SS	Degrees of Freedom df	Mean Square MS	F statistic	p-value*
Treatment	6.91	1	6.91	101.154	1.64E-12
Error	2.7325	40	0.0683		
Total	9.6425	41			

***-**p value is statistically significant so the post hoc test can be applied.

Scheffé results			
Treatment Pair	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B Positive Mediocrity vs. Above Average	10.0575	1.64E-12	** p<0.01

** - p value indicates that the comparison is statistically significant at the 99% confidence level

Graph Result:



Within Team Analysis – Chicago Bulls

Descriptive statistics of the 2 independent treatments:			
Categories →	**A (Negative Mediocrity)**	**B (Positive Mediocrity)**	Pooled Total
Seasons	17	14	31
Average Z-Score	-0.4735	0.5623	-0.0057

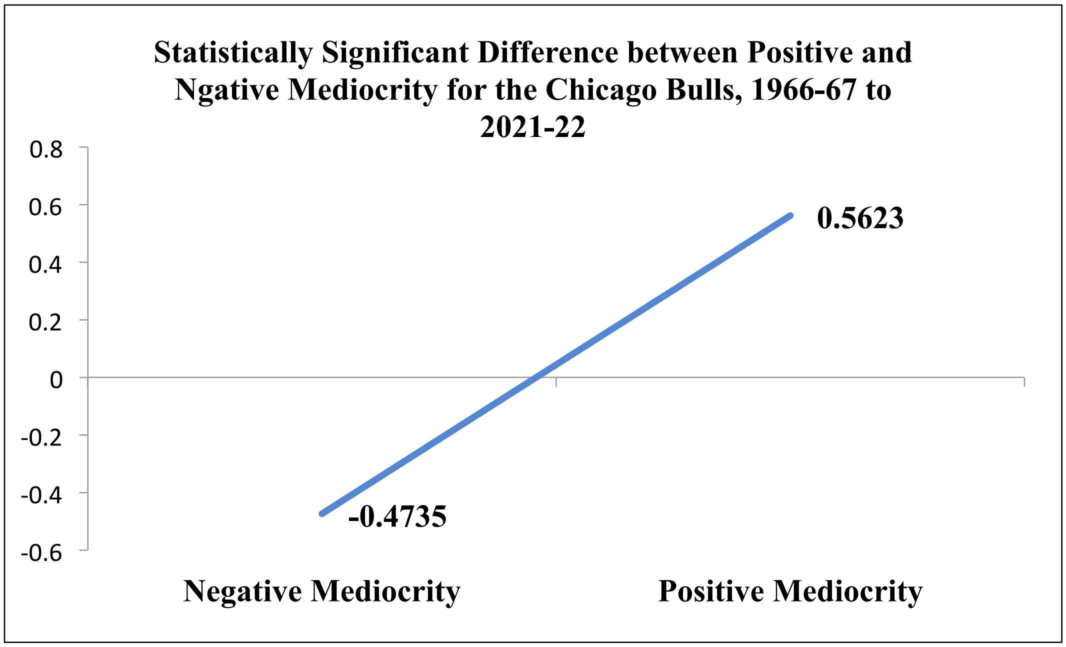
One-way ANOVA of 2 independent treatments:					
Source	**Sum of Squares SS**	**Degrees of Freedom df**	**Mean Square MS**	**F statistic**	**p-value***
Treatment	8.2372	1	8.2372	79.3557	8.49E-10
Error	3.0102	29	0.1038		
Total	11.2474	30			

***-** p value of the ANOVA is at the 99% confidence level allowing the post hoc test

Scheffé results			
Treatment	Scheffé	Scheffé	Scheffé
Pair	TT-statistic	p-value	Inference
A vs. B (Negative Mediocrity vs. Positive Mediocrity)	8.9082	8.49E-10	**** p<0.01**

**** -** p value indicates statistical significance at the 99% confidence level.

Graph Result 1:



Within Team Analysis – Minnesota Timberwolves

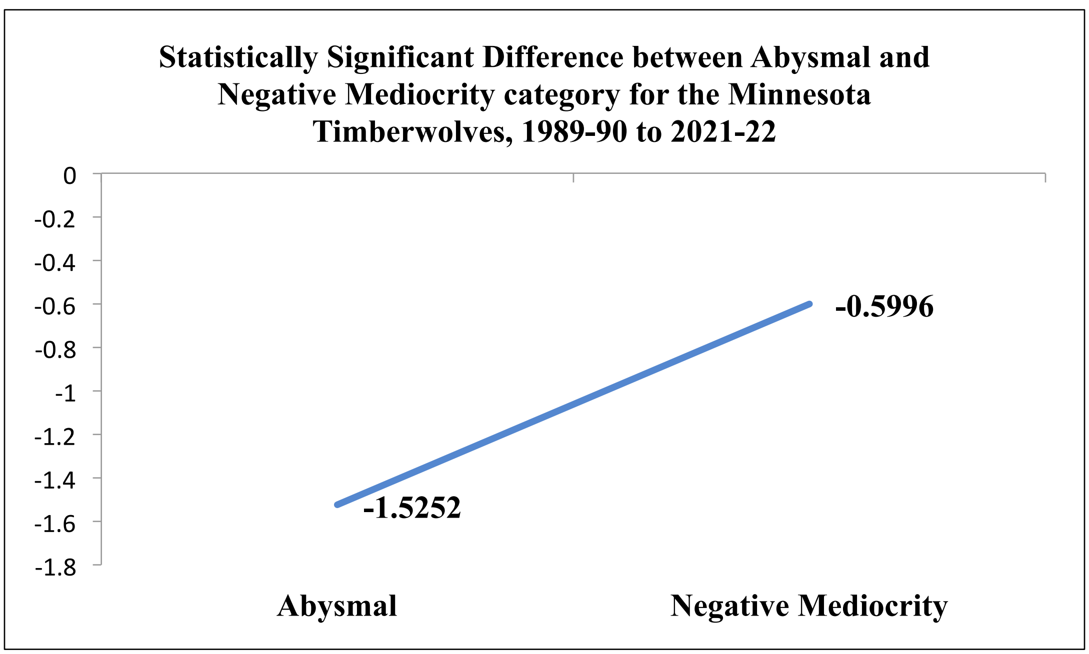
Categories →	**A (Abysmal)**	**B (Negative Mediocrity)**	**Pooled Total**
Seasons	13	11	24
Average Z-Score	-1.5252	-0.5996	-1.101

One-way ANOVA of the 2 independent treatments:					
Source	**Sum of Squares SS**	**Degrees of Freedom df**	**Mean Square MS**	**F statistic**	**p-value***
Treatment	5.1051	1	5.1051	52.6055	2.89E-07
Error	2.135	22	0.097		
Total	7.24	23			

- **Overall ANOVA indicates statistical significance allowing the post hoc test.**

Scheffé results			
Treatment	Scheffé	Scheffé	Scheffé
Pair	TT-statistic	p-value	Inference
A vs. B (Abysmal vs. Negative Mediocrity)	7.253	2.89E-07	** $p<0.01$

** - p value is statistically significant at the 99% confidence level allowing for the use of the post hoc test.



Between Teams Analysis
Between Dallas-San Antonio vs. Chicago (Positive Mediocrity)
There was no statistical significance for this category between the teams.

Between Chicago vs. Minnesota (Negative Mediocrity)
There was no statistical significance for this category between the teams.

Concluding Analysis

Since the data has been converted to Z-Scores, we are able to compare the categorized teams from each of the leagues. There will be three analyses, the best, worst and average teams across the four leagues to see how the teams from each compare to each other.

The Best Team Analysis
We start with the best of each league against each other. Is there one dominant team that is different than the other best rated teams?

Descriptive statistics of the 4 independent treatments:					
Best Teams →	**A (Yankees)**	**B (Canadiens)**	**C (Bears)**	**D (Spurs)**	**Pooled Total**
Playing Seasons	121	103	102	55	381
Overall Z-Score	0.7341	0.5041	0.3058	0.6406	0.5437

One-way ANOVA of the 4 independent treatments:					
Source	Sum of squares SS	Degrees of Freedom df	Mean square MS	F statistic	p-value**
Treatment	10.837	3	3.6123	4.5152	0.004
Error	301.6122	377	0.8		
Total	312.4492	380			

** - p value is less than .01 at the 99% confidence level so the post hoc analysis can be applied.

Scheffé results			
Treatment pairs	Scheffé TT-statistic	Scheffé p-value	Scheffé Inference
A vs. B	1.9179	0.2999481	insignificant
A vs. C (Yankees vs. Bears)	3.5623	0.0058586	**** $p<0.01$**
A vs. D	0.6429	0.9374079	insignificant
B vs. C	1.5873	0.4726885	insignificant
B vs. D	0.9136	0.8411359	insignificant
C vs. D	2.2374	0.1732989	insignificant

The post hoc test yields only one statistically significant comparison. The New York Yankees are statistically different than the Chicago Bears. There are no other statistically significant comparisons. If the teams are ranked they are as follows:
New York franchise (Baseball) The Yankees is the most consistent winning sports franchise.
San Antonio franchise (Basketball)
Montreal Canadiens (Hockey)
Chicago franchise (Football)

The Whatever Analysis, (The most average Sports franchise)

The analysis between the 'whatever' teams from the 4 leagues yields no significant difference between the four teams. This translates to the four teams are similar in their respective regular seasons. They are the epitome of mediocrity.

The Worst Analysis (The 4 worst teams with the most consistent poor performance).

The analysis between the 'worst' teams from the four leagues did not yield a significant difference between the four teams. Once again the analysis illustrates that the worst teams from the four leagues are similar to each other and are really the worst of their respective league.

Thus concludes the analysis based on the characteristics of the regular season play of teams in the four professional leagues in North America.

Glossary

ANOVA: a statistical method of studying the variation in responses of two or more groups on a dependent variable. ANOVAs test for significant differences among the mean response values of the groups and can be used to isolate both the joint interaction effects and the separate main effects of independent variables upon the dependent variable.

Central Tendency: the middle or center point of a set of scores. The central tendency of a sample data set, for instance, may be estimated by a number of different statistics (e.g., mean, median, mode).

Chebyshev's Theorem: According to Chebyshev's inequality, the probability that a value will be more than two standard deviations from the mean (k = 2) cannot exceed 25 percent.

Confidence Level: The confidence level tells you how sure you can be. It is expressed as a percentage and represents how often the true percentage of the population who would pick an answer that lies within the confidence interval. The 95% confidence level means you can be 95% certain; the 99% confidence level means you can be 99% certain. Most researchers work for a 95% confidence level.

Interquartile Rage: an index of the dispersion within a data set: the difference between the 75th and 25th percentile scores (also known as the upper and lower hinges) within a distribution.

Mean: The arithmetic mean (usually synonymous with average) represents a point about which the numbers balance. For example, if unit masses are placed on a line at points with coordinates x1, x2, …, xn, then the arithmetic mean is the coordinate of the centre of gravity of the system. In statistics, the arithmetic mean is commonly used as the single value typical of a set of data. (https://www.britannica.com/science/mean-median-and-mode)

Median: The median is the middle value in a list ordered from smallest to largest. The mode is the most frequently occurring value on the list. ((https://www.britannica.com/science/mean-median-and-mode)

Mode: The mode is the most frequently occurring value on the list. (https://www.britannica.com/science/mean-median-and-mode)

µ: (the greek letter "mu") is used to denote the population mean. The population mean is worked out in exactly the same way as the sample mean: add all of the scores together, and divide the result by the total number of scores.

Normal Distribution: normal distribution, also called Gaussian distribution, the most common distribution function for independent, randomly generated variables. Its familiar bell-shaped curve is ubiquitous in statistical reports, from survey analysis and quality control to resource allocation.

Standard Deviation: (symbol: SD) a measure of the variability of a set of scores or values within a group, indicating how narrowly or broadly they deviate from the mean. A small standard deviation indicates data points that cluster around the mean, whereas a large standard deviation indicates data points that are dispersed across many different values.

Variability: the degree to which members of a group or population differ from each other, as measured by statistics such as the range, standard deviation, and variance.

Z-Score: the standardized score that results from applying a z-score transformation to raw data. For purposes of comparison, the data set is converted into one having a distribution with a mean of 0 and a standard deviation of 1. For example, consider a person who scored 30 on a 40-item test having a mean of 25 and standard deviation of 5, and 40 on an 80-item test having a mean of 50 and a standard deviation of 10. The resulting z scores would be +1.0 and –1.0, respectively. Thus, the individual performed better on the first test, on which he or she was one standard deviation above the mean, than on the second test, on which he or she was one standard deviation below the mean.

Bibliogrpahy

American Psychological Association: https://dictionary.apa.org/

Boston University School of Public Health: https://www.bu.edu/sph/

Elsievier 2022: https://www.elsevier.com/

Encyclopaedia Britannica Inc, 2022: britannica.com

JSTOR, One Liberty Plaza, New York, 2022-08-08: jstor.org

Information Technology Laboratory, Lund Research Ltd, 2020:
https://statistics.laerd.com/contactus.php

National Institute of Standards Technology, US Department of Commerce,
https://www.itl.nist.gov/

National Library of Medicine: https://www.nih.gov/

Penn Stat Eberly College of Science, STAT503, Design of Experiments:
https://online.stat.psu.edu/

Statistics Canada: https://www.statcan.gc.ca

University of Sussex: https://www.sussex.ac.uk/

University of Utah: https://www.utah.edu