DECEMBER 8, 2022

NFL NEXT GEN STATS

ERIC POLVERARI
CLEMSON UNIVERSITY
STAT 8030

Introduction

Each Sunday during the NFL season, millions of people tune in to watch some of the best athletes in the world compete in what has become America's most popular TV sports event. So far in 2022, the top 14 watched sporting events involved an NFL game. However, in comparison to other popular sports like baseball and basketball, football lagged behind in analytics and have only recently started to invest in understanding the game through a different lens. In 2016, the NFL launched nextgenstats.nfl.com and the consumer finally has an opportunity to try and quantify a game that is a perfect blend of art and science. Next Generation Stats is "NFL player tracking that capture the real time location data, speed and acceleration for every player, every play on each inch of the field. Sensors throughout the stadium track tags placed on players' shoulder pads, charting individual movements within inches." Using data from the website can provide valuable insight to the nation's most popular sport.

In this study, the question of interest pertains to how the quarterback's style and tendencies have changed, if at all, for the past several seasons by examining different metrics captured by the Next Generation Stats. To address this, preliminary exploratory data analysis will be performed with summary statistics and boxplots to get a clearer illustration of the dataset before performing an ANOVA test to determine if there's reason to believe that the game's most important position has changed in their on-field behavior over the past several seasons.

Data

The data used in this study is from https://nextgenstats.nfl.com/. This study will be using next generation stats for passing metrics, as it's arguably the most valuable play in the NFL. According to their website, "Next Generation Stats (aka Next Gen Stats) tracks numerous data points to enhance our understanding of a Quarterback's style and tendencies. In 2016 we began exposing data points such as Time to Throw to indicate how quickly a passer releases a pass and various measurements of Air Yards and Air Distance to indicate how far a passer throws the ball downfield on completions. " In this dataset, only observations from completed NFL seasons from when the dataset begins will be used for this study. To ensure a reasonable sample size, the data only contains QB's that have a minimum of 105 pass attempts, calculated using 15 attempts, multiplied by the number of week and divided by 2. Each observation represents a QB's individual metrics. The dataset was clean and contained no missing values. The data was transferred to a Microsoft Excel spreadsheet and uploaded into a SAS program. To get an idea of what the data looks like, below is provided a condensed version of the final dataset looks like.

Obs	YEAR	TT	CAY	IAY	AGG	AYTS	CPAE
1	2021	2.5	5.5	8	14	-0.4	-1
2	2021	2.83	5.8	7.6	17.6	-0.9	-0.3
3	2021	2.8	6.7	8.5	12.1	0.1	-0.6
4	2021	2.84	4.8	7.3	8.7	-1.2	-2.1
5	2021	2.82	5.6	8	13.6	-0.8	2.2
6	2021	2.69	6.4	8.3	19.2	-0.6	6
7	2021	2.75	5.9	8	15.4	-0.4	2.2
8	2021	2.9	6.5	8.6	11.3	-0.2	-0.9
9	2021	2.76	6.4	8	15.9	-1.4	2.4
10	2021	2.63	4.9	7.6	16.4	-0.6	3.3
11	2021	2.77	5.7	7.2	16.3	-1.3	0.9
12	2021	2.67	5.9	7.4	15.6	-1.9	1.2
13	2021	2.71	5.7	8.1	15.4	-0.9	1.2
14	2021	2.77	5.8	7.9	12.3	-1.1	3.9
15	2021	2.38	4.4	6.7	17.9	-1.9	-3.7
16	2021	2.68	5.4	7.5	15.4	-1.2	0.3
17	2021	2.87	5.4	8	13.3	-0.8	-5.4
18	2021	2.83	6	7.9	16.5	-1.3	-2.3
19	2021	2.97	5.2	7.9	18	-1.2	-0.2
20	2021	2.77	4.3	6.6	10.7	-2.2	0.8

This dataset contains seven variables:

YEAR: the year for which the QB's season metrics took place

TT: Time to Throw measures the average amount of time elapsed from the time of snap to throw on every pass attempt for a passer (sacks excluded).

CAY: shows the average Air Yards a passer throws on completions.

IAY: shows the average Air Yards a passer throws on all attempts.

AGG: Aggressiveness tracks the amount of passing attempts a quarterback makes that are into tight coverage, where there is a defender within 1 yard or less of the receiver at the time of completion or incompletion. AGG is shown as a % of attempts into tight windows over all passing attempts.

AYTS: Air Yards to the Sticks shows the amount of Air Yards ahead or behind the first down marker on all attempts for a passer. The metric indicates if the passer is attempting his passes past the 1st down marker, or if he is relying on his skill position players to make yards after catch.

CPAE: Completion Percentage Above Expectation. A passer's actual completion percentage compared to their Expected Completion Percentage.

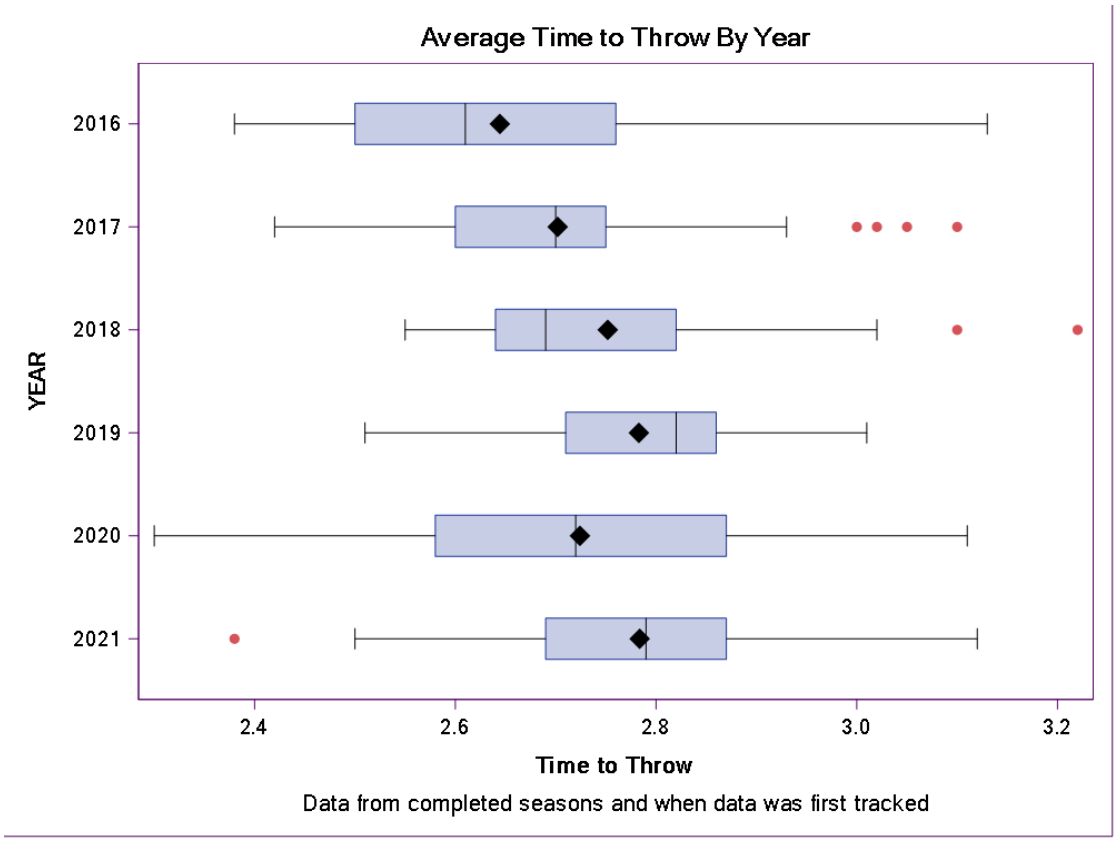
The year variable will be used as a class variable to examine how the next generation stats has changed, if at all, for the past six NFL seasons.

Data Analysis

Now, each variable will have a summary of statistics and boxplots, by year, to uncover information about the data. This will include the mean, median, minimum value, maximum value and standard deviation.

Average Time to Throw By Year Summary Statistics

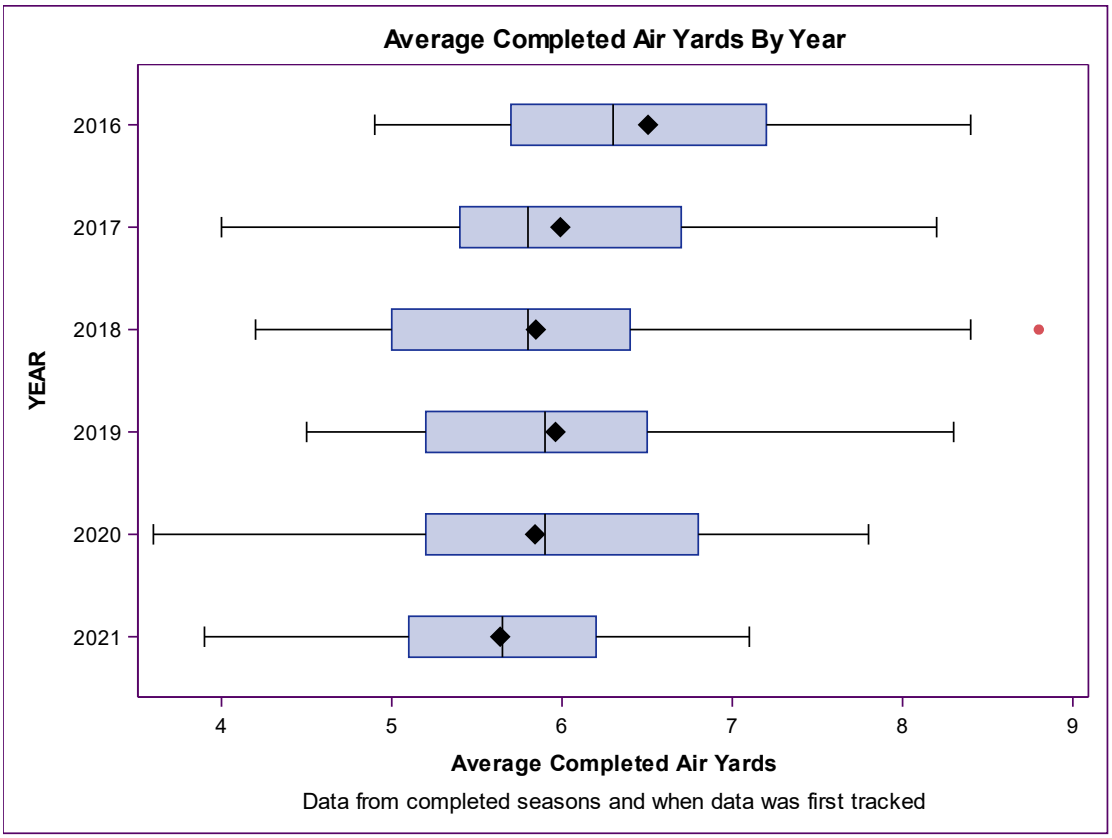
Analysis Variable : TT Time to Throw						
YEAR	N Obs	Mean	Median	Minimum	Maximum	Std Dev
2016	39	2.6443590	2.6100000	2.3800000	3.1300000	0.1702134
2017	41	2.7019512	2.7000000	2.4200000	3.1000000	0.1588430
2018	39	2.7517949	2.6900000	2.5500000	3.2200000	0.1594872
2019	39	2.7828205	2.8200000	2.5100000	3.0100000	0.1113068
2020	41	2.7241463	2.7200000	2.3000000	3.1100000	0.1840377
2021	38	2.7836842	2.7900000	2.3800000	3.1200000	0.1476376



By examining the mean for the TT variable, there is a slight trend that appears to show that from the 2016 season, QB's had an increase for about three seasons before slightly going down in 2020 and then back up in 2021. While there are extreme values for the 2017 and 2018 season, they don't deviate too far for concern. Perhaps it's no surprise that the 2020 season may be different from other seasons, as a global pandemic was a challenge for everyone.

Average Completed Air Yards Summary Statistics By Year

Analysis Variable : CAY Completed Air Yards						
YEAR	N Obs	Mean	Median	Minimum	Maximum	Std Dev
2016	39	6.5051282	6.3000000	4.9000000	8.4000000	0.9735850
2017	41	5.9902439	5.8000000	4.0000000	8.2000000	0.9096166
2018	39	5.8461538	5.8000000	4.2000000	8.8000000	0.9794859
2019	39	5.9615385	5.9000000	4.5000000	8.3000000	0.9626213
2020	41	5.8414634	5.9000000	3.6000000	7.8000000	1.0530849
2021	38	5.6368421	5.6500000	3.9000000	7.1000000	0.7528041



Data from completed seasons and when data was first tracked

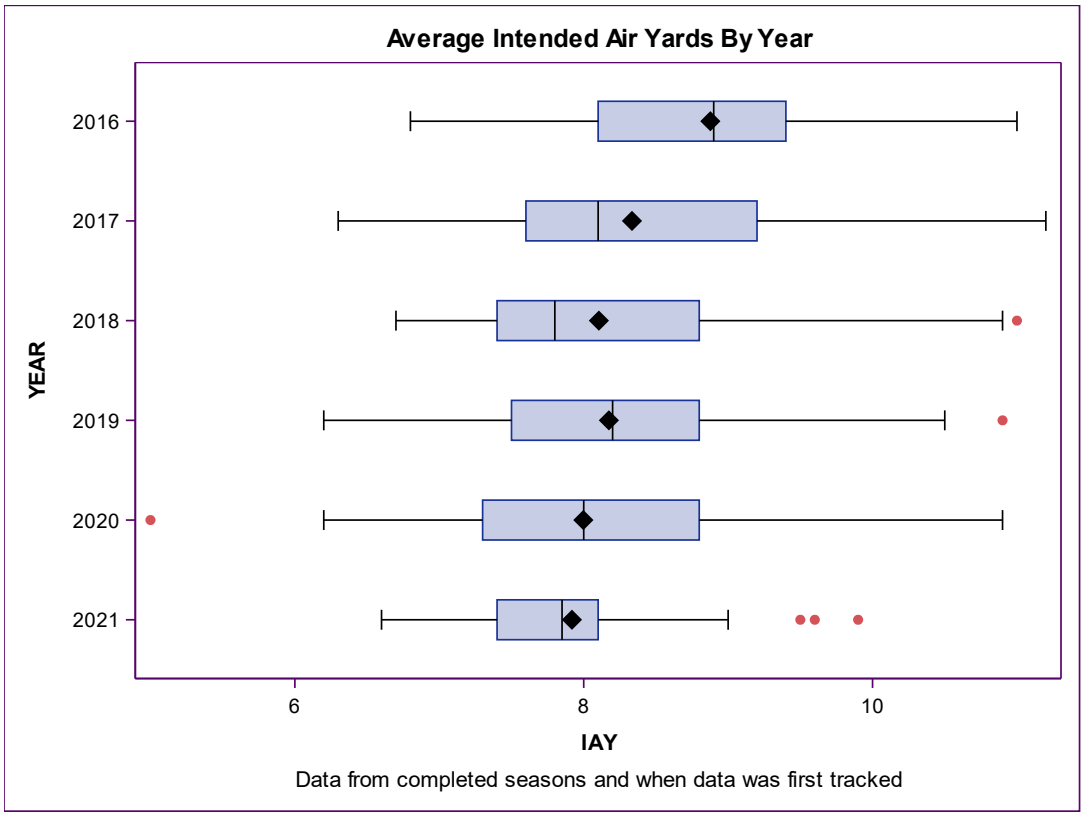
Average Completed Air Yards Summary Statistics By Year

The average CAY shows a relatively high mean for the 2016 season, before a pretty significant drop to 2017. From 2017 to 2020, there seems to be slight variation, but it's notable to see the 2021 season saw the QB's with the lowest CAY on average. The standard deviation for all years is pretty small and the data spread of the data seems to match that in the boxplot.

Data from completed seasons and when data was first tracked

Average Intended Air Yards Summary Statistics By Year

Analysis Variable : IAY Intended Air Yards						
YEAR	N Obs	Mean	Median	Minimum	Maximum	Std Dev
2016	39	8.8769231	8.9000000	6.8000000	11.0000000	0.9943158
2017	41	8.3341463	8.1000000	6.3000000	11.2000000	1.1275216
2018	39	8.1051282	7.8000000	6.7000000	11.0000000	1.0721225
2019	39	8.1743590	8.2000000	6.2000000	10.9000000	1.0732925
2020	41	7.9975610	8.0000000	5.0000000	10.9000000	1.0987465
2021	38	7.9184211	7.8500000	6.6000000	9.9000000	0.7540361



Data from completed seasons and when data was first tracked

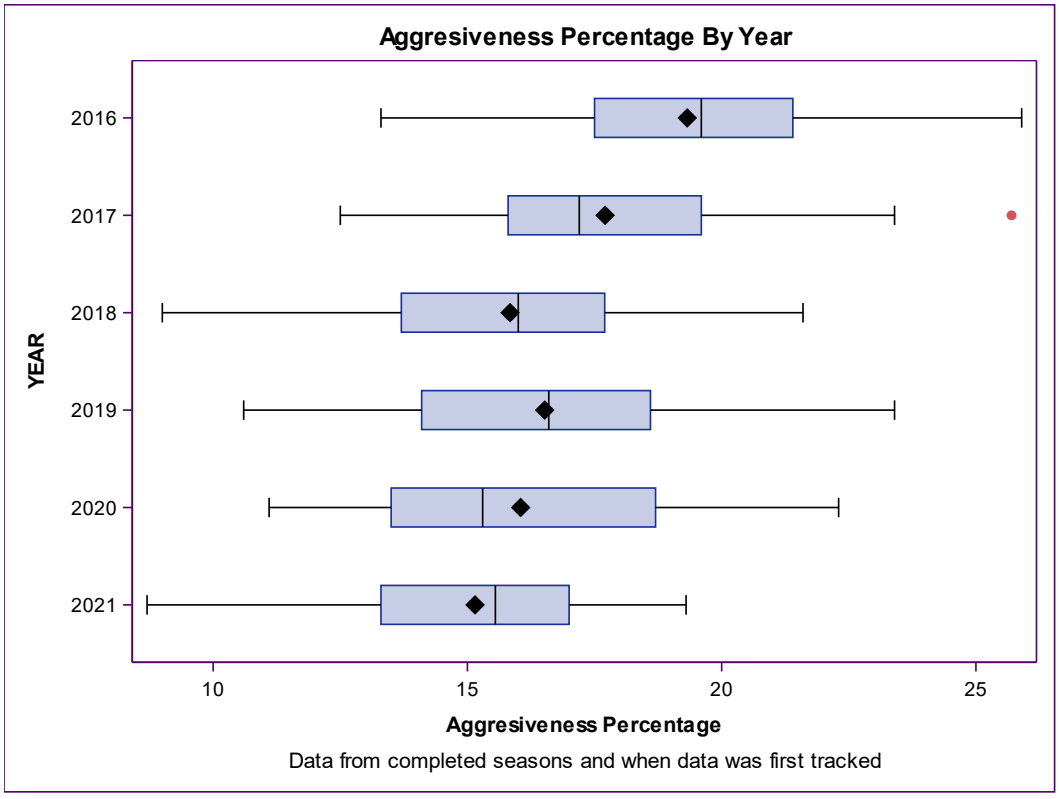
Average Intended Air Yards Summary Statistics By Year

By now, there does appear to be a pattern that the data for QBs has changed since the tracking data was first introduced. The 2016 season saw passers intending the largest average pass attempt decreasing the following seasons, though there doesn't appear to be significant changes from 2018 to 2021. The 2021 season has a bit of an unusual spread, but the values aren't too far from the rest of the data and shouldn't drastically change the results of the inference tests.

Data from completed seasons and when data was first tracked

Aggressiveness Percentage Summary Statistics By Year

Analysis Variable : AGG Aggressiveness						
YEAR	N Obs	Mean	Median	Minimum	Maximum	Std Dev
2016	39	19.3230769	19.6000000	13.3000000	25.9000000	2.9736971
2017	41	17.7073171	17.2000000	12.5000000	25.7000000	2.9831519
2018	39	15.8384615	16.0000000	9.0000000	21.6000000	2.7939423
2019	39	16.5205128	16.6000000	10.6000000	23.4000000	3.1136524
2020	41	16.0463415	15.3000000	11.1000000	22.3000000	3.1950820
2021	38	15.1500000	15.5500000	8.7000000	19.3000000	2.5598511

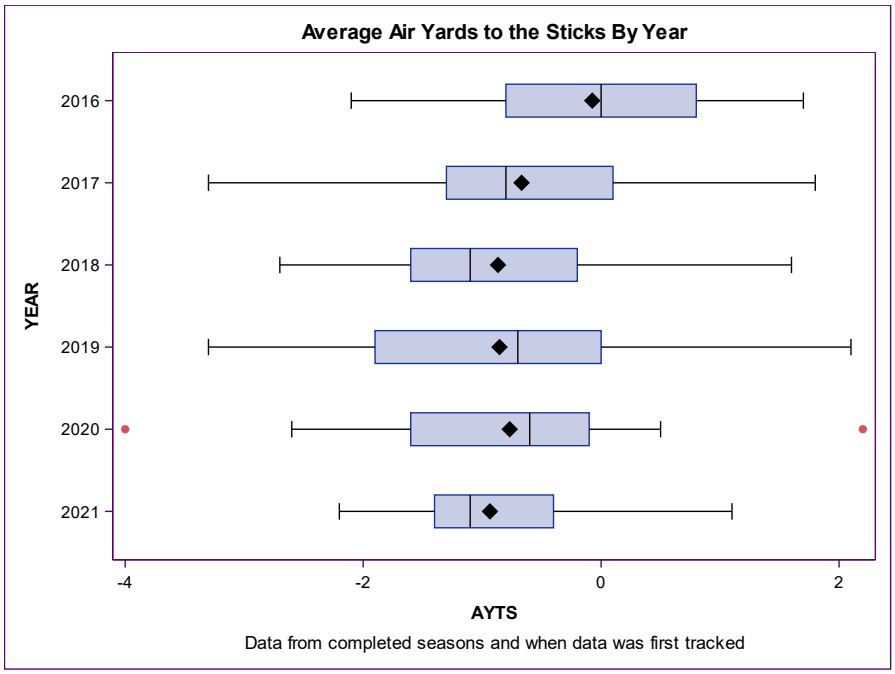


Here, aggressiveness percentage tends to show that the 2016 season mean was greater by a potentially significant amount compared to the following seasons. The boxplots show a fairly even spread between all the years. The standard deviation is relatively large for each season, but not large enough to be concerned the spread of the data is unusual.

Data from completed seasons and when data was first tracked

Air Yards to the Sticks Summary Statistics By Year

Analysis Variable : AYTS Air Yards to the Sticks						
YEAR	N Obs	Mean	Median	Minimum	Maximum	Std Dev
2016	39	-0.0743590	0	-2.1000000	1.7000000	0.9497141
2017	41	-0.6682927	-0.8000000	-3.3000000	1.8000000	1.1134269
2018	39	-0.8666667	-1.1000000	-2.7000000	1.6000000	1.0193737
2019	39	-0.8538462	-0.7000000	-3.3000000	2.1000000	1.1233985
2020	41	-0.7682927	-0.6000000	-4.0000000	2.2000000	1.1369782
2021	38	-0.9342105	-1.1000000	-2.2000000	1.1000000	0.7159482

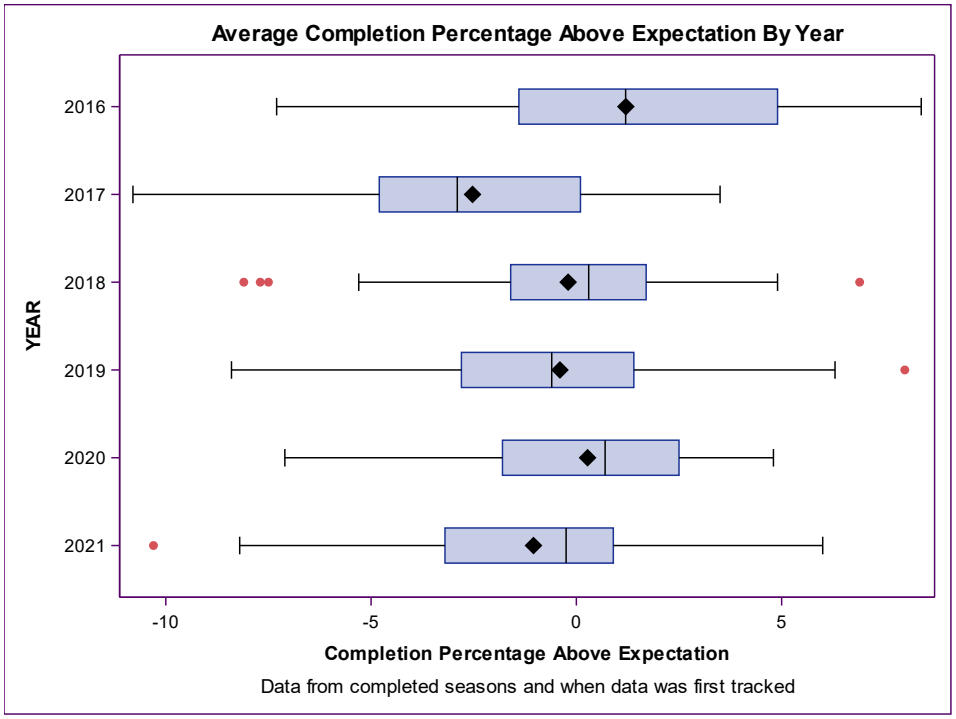


Air Yards to the Sticks Summary Statistics By Year

Once again, a variable for the 2016 season, AYTS, mean appears to deviate from the successive seasons, which is nearly a yard shorter than the 2021 season. Perhaps there was a change in philosophic approach for passer's and offense's in general that lead to the future seasons to not be as aggressive going for throws around or past the first down and sticks to more emphasis on throwing passes short of the stick to allow their playmakers to gain the necessary yards for a first down. There are a couple of outliers for the 2020 season and are somewhat concerning, but don't skew the data a significant amount to remove them, though there probably should be some skepticism for that year.

Data

Analysis Variable : CPAE Completion %% Above Expectation						
YEAR	N Obs	Mean	Median	Minimum	Maximum	Std Dev
2016	39	1.2051282	1.2000000	-7.3000000	8.4000000	4.0068986
2017	41	-2.5292683	-2.9000000	-10.8000000	3.5000000	3.3363336
2018	39	-0.2000000	0.3000000	-8.1000000	6.9000000	3.4279041
2019	39	-0.4000000	-0.6000000	-8.4000000	8.0000000	3.2895368
2020	41	0.2731707	0.7000000	-7.1000000	4.8000000	2.8674923
2021	38	-1.0447368	-0.2500000	-10.3000000	6.0000000	3.3049265



The 2016 season, once again, shows a potentially significant difference in CPAE mean compared to the seasons after 2016. CPAE is a very interesting statistic and it can show if passers are performing at an exceptional level or they're just attempting easier passes. As we saw from IAY, it doesn't appear the latter is the case, as passers were more willing to try longer passes than the other seasons. Perhaps this was just an exceptional year of QB play that is very difficult to replicate. The 2018 season does have several outliers, but not significant enough to be removed from this analysis. The standard deviation also shows a wider spread compared to the other statistics, though not large enough to be concerned.

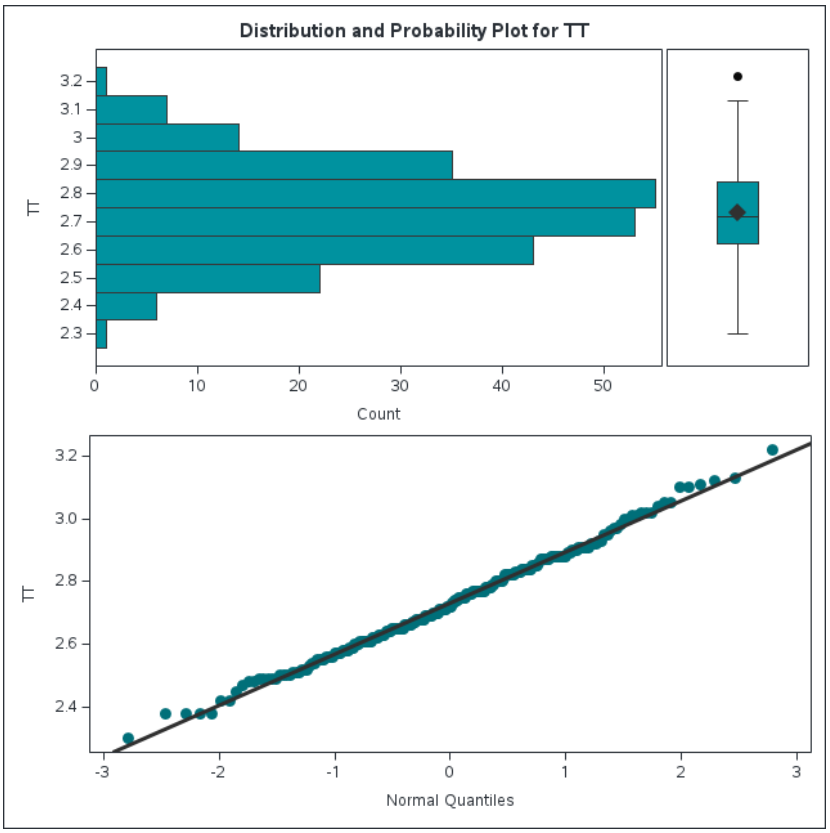
Data from completed seasons and when data was first tracked

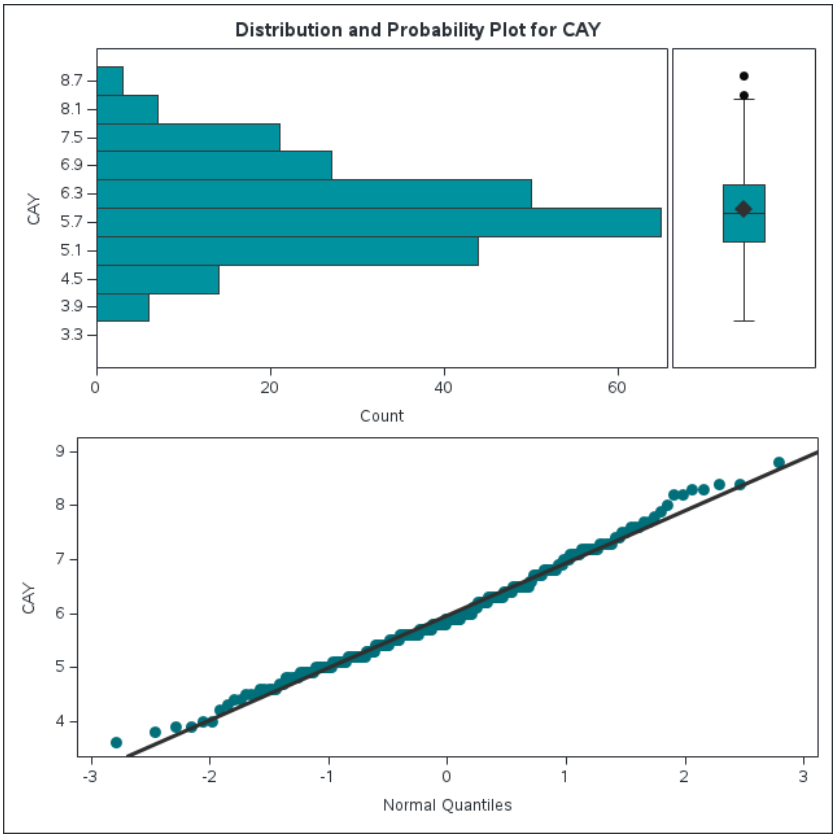
Data

From the preliminary exploratory data analysis, there appears to be some potential significant differences in QB's style and tendencies, especially for the 2016 season compared to the rest of the seasons. Now, an ANOVA test will be applied to each variable with the class of year to determine if there's a significant difference between any years. For each variable, two of the three assumptions are satisfied: the distributions have essentially the same variance evident by the standard deviation ratio falling within 0.5 to 2 between the smallest and largest sample standard deviation. The data are independent as they were selected from independent years. While the sample size meets the general threshold of 30, the highest number of observations for a year is 41, so normality plots for each variable will be provided below to assure we can proceed with an ANOVA test. Each variable has a normal population distribution, so we can confidently conduct the inference test. There will also be a post-hoc Tukey HSD to determine which years, if any, differ significantly from one another.

Data from completed seasons and when data was first tracked

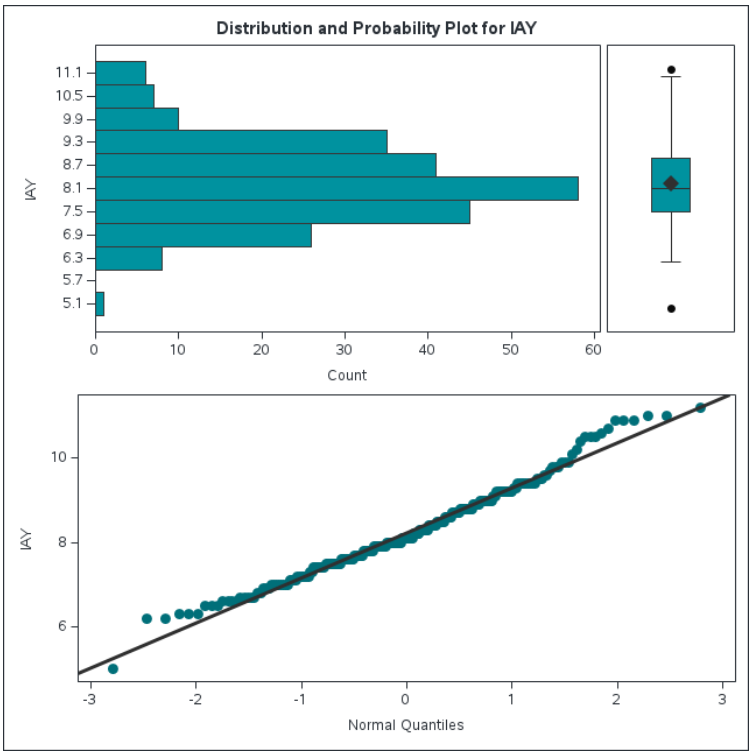
Data

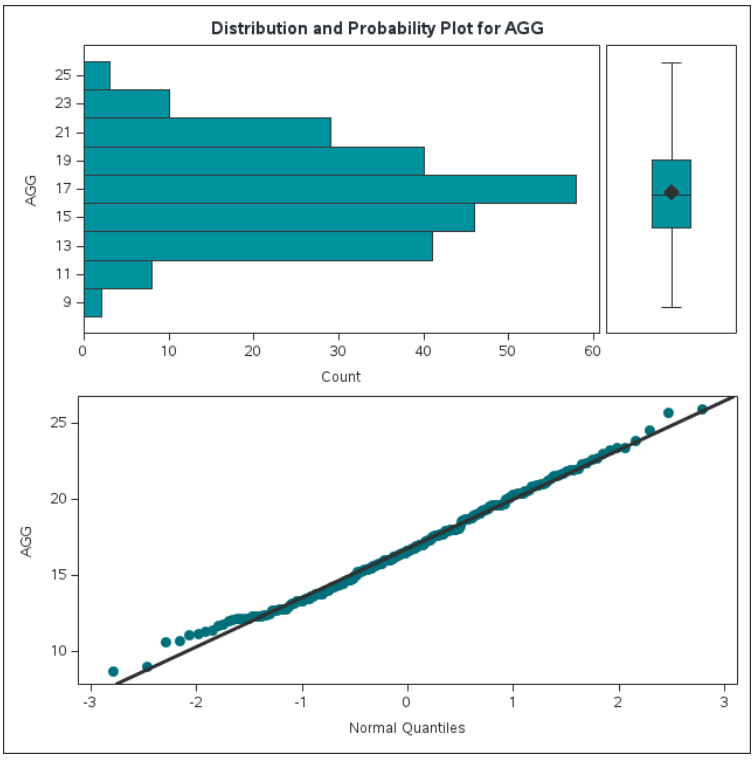




Data from completed seasons and when data was first tracked

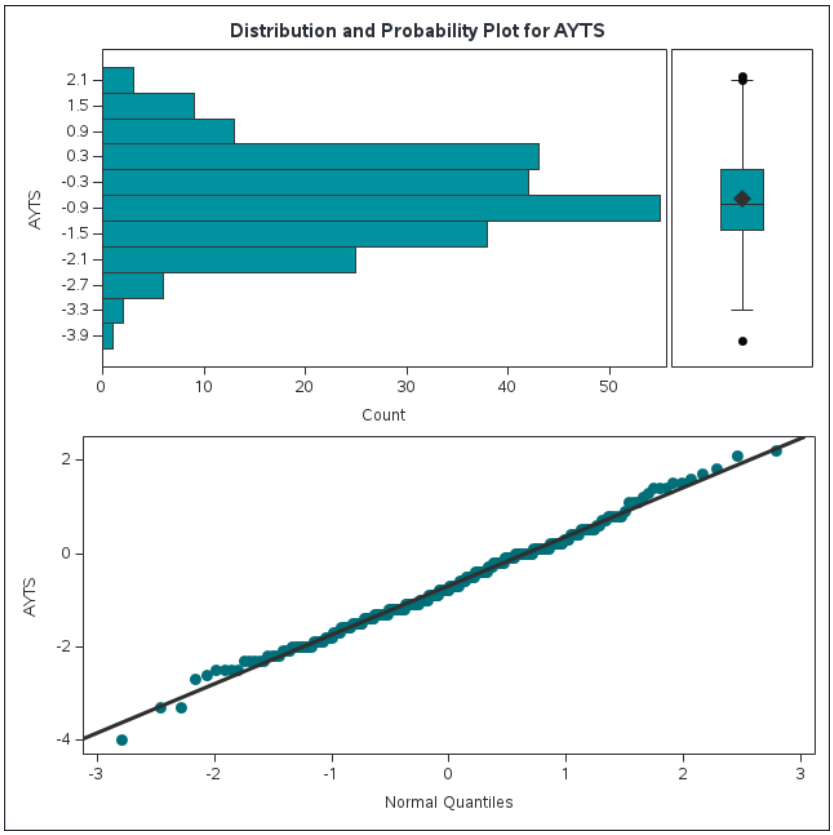
Data

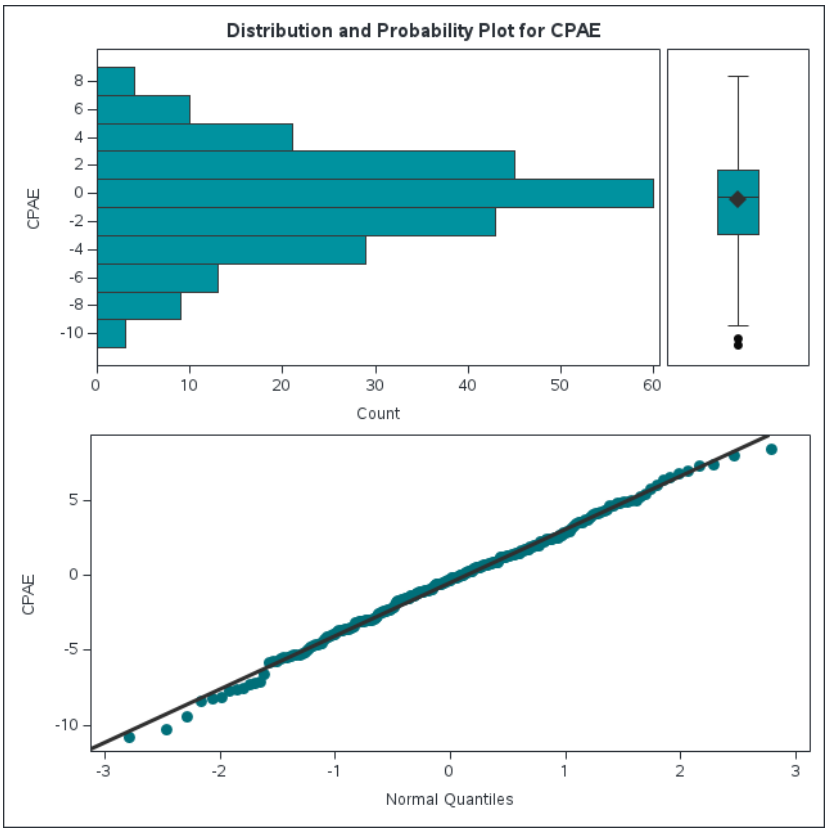




Data from completed seasons and when data was first tracked

Data





Data from completed seasons and when data was first tracked

ANOVA Test For Time to Throw Differences By Year

Class Level Information		
Class	**Levels**	**Values**
YEAR	6	2016 2017 2018 2019 2020 2021

Number of Observations Read	237
Number of Observations Used	237

H0: M1=M2=M3=M4=M5=M6
Ha: at least one mu is not equal

Significance level: 0.05

ANOVA Test For Time to Throw Differences By Year

Dependent Variable: TT
TT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.55634947	0.11126989	4.50	0.0006
Error	231	5.70884631	0.02471362		
Corrected Total	236	6.26519578			

R-Square	Coeff Var	Root MSE	TT Mean
0.088800	5.756492	0.157206	2.730928

Source	DF	Type I SS	Mean Square	F Value	Pr > F
YEAR	5	0.55634947	0.11126989	4.50	0.0006

Source	DF	Type III SS	Mean Square	F Value	Pr > F
YEAR	5	0.55634947	0.11126989	4.50	0.0006

The F Value is 4.50 and the p-value is 0.0006, which is significantly less than the significance level of 0.05. Therefore, there is strong evidence to conclude that at least one year is different from another.

ANOVA Test For Time to Throw Differences By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

YEAR	TT LSMEAN	LSMEAN Number
2016	2.64435897	1
2017	2.70195122	2
2018	2.75179487	3
2019	2.78282051	4
2020	2.72414634	5
2021	2.78368421	6

Least Squares Means for effect YEAR
Pr > |t| for H0: LSMean(i)=LSMean(j)

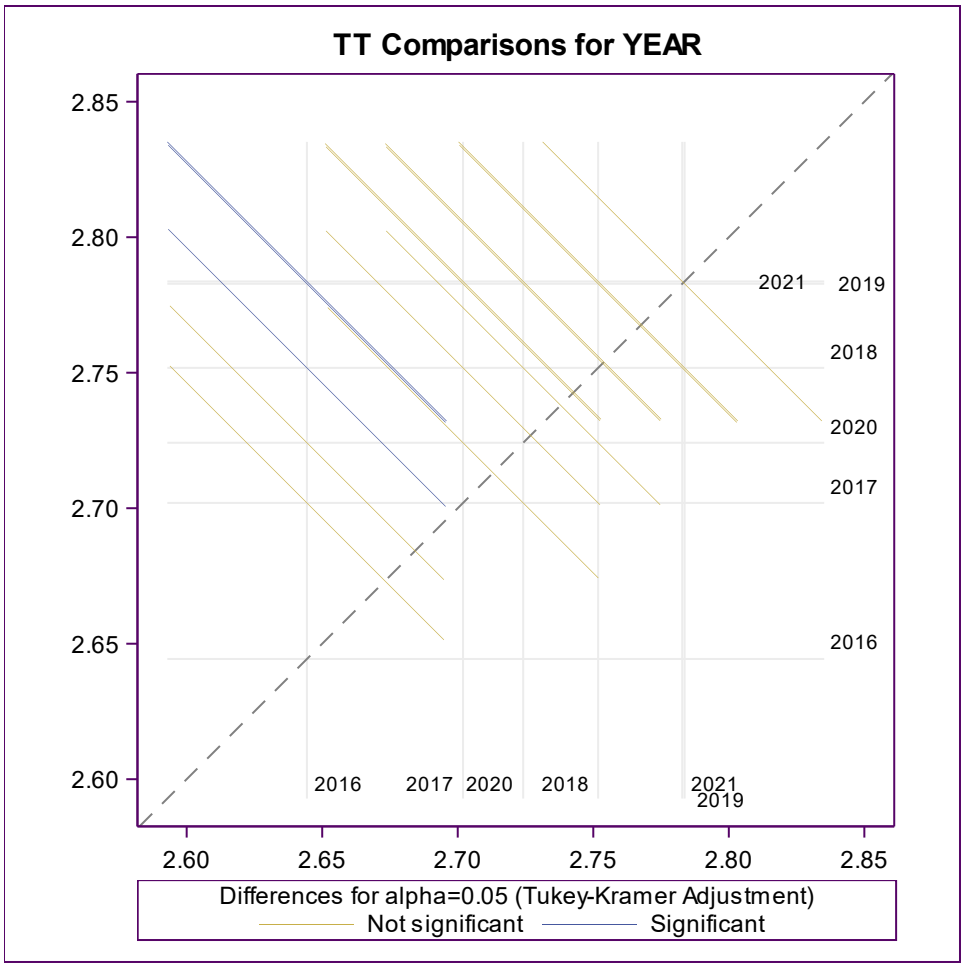
Dependent Variable: TT

i/j	1	2	3	4	5	6
1		0.5743	0.0333	0.0018	0.2111	0.0018
2	0.5743		0.7164	0.1983	0.9879	0.1947
3	0.0333	0.7164		0.9529	0.9696	0.9486
4	0.0018	0.1983	0.9529		0.5540	1.0000
5	0.2111	0.9879	0.9696	0.5540		0.5452
6	0.0018	0.1947	0.9486	1.0000	0.5452	

From the Tukey HSD, we see the 2016 season is significantly different than the 2018, 2019 and 2021 seasons.

ANOVA Test For Time to Throw Differences By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer



Data from completed seasons and when data was first tracked

ANOVA Test For Completed Air Yards By Year

Class Level Information		
Class	**Levels**	**Values**
YEAR	6	2016 2017 2018 2019 2020 2021

Number of Observations Read	237
Number of Observations Used	237

H0: M1=M2=M3=M4=M5=M6
Ha: at least one mu is not equal

Significance level: 0.05

Data from completed seasons and when data was first tracked

ANOVA Test For Completed Air Yards By Year

Dependent Variable: CAY
CAY

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	16.6729117	3.3345823	3.74	0.0028
Error	231	206.1122359	0.8922608		
Corrected Total	236	222.7851477			

R-Square	Coeff Var	Root MSE	CAY Mean
0.074839	15.83793	0.944596	5.964135

Source	DF	Type I SS	Mean Square	F Value	Pr > F
YEAR	5	16.67291174	3.33458235	3.74	0.0028

Source	DF	Type III SS	Mean Square	F Value	Pr > F
YEAR	5	16.67291174	3.33458235	3.74	0.0028

The F Value is 3.47 and the p-value is 0.0028, less than the significance level of 0.05. Therefore, we can conclude there is strong evidence that at least one of the seasons for the average CAY is significantly different.

ANOVA Test For Completed Air Yards By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

YEAR	CAY LSMEAN	LSMEAN Number
2016	6.50512821	1
2017	5.99024390	2
2018	5.84615385	3
2019	5.96153846	4
2020	5.84146341	5
2021	5.63684211	6

Least Squares Means for effect YEAR
Pr > |t| for H0: LSMean(i)=LSMean(j)

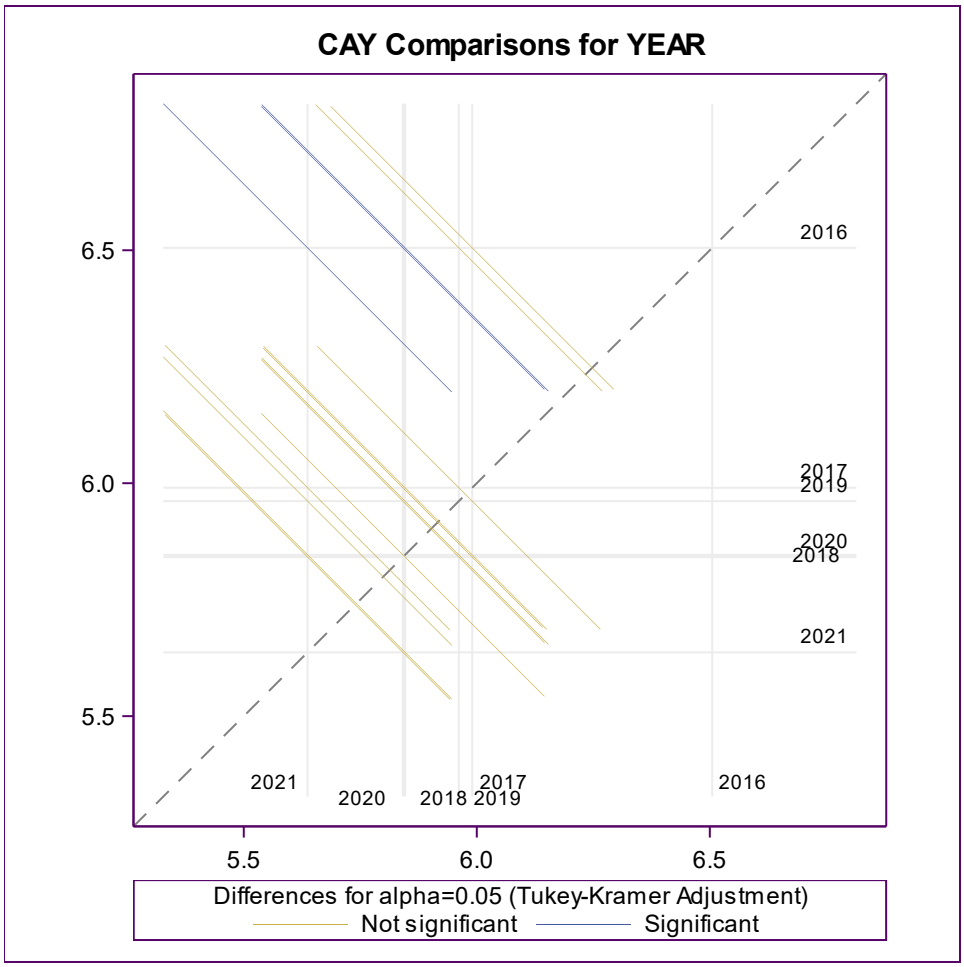
Dependent Variable: CAY

i/j	1	2	3	4	5	6
1		0.1480	0.0277	0.1166	0.0231	0.0010
2	0.1480		0.9838	1.0000	0.9802	0.5587
3	0.0277	0.9838		0.9945	1.0000	0.9263
4	0.1166	1.0000	0.9945		0.9930	0.6595
5	0.0231	0.9802	1.0000	0.9930		0.9294
6	0.0010	0.5587	0.9263	0.6595	0.9294	

The 2016 season average CAY was significantly different that than 2018, 2020 and 2021 seasons. The 2017 to 2021 seasons do not differ significantly from one another.

ANOVA Test For Completed Air Yards By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer



Data from completed seasons and when data was first tracked

ANOVA Test Average Intended Air Yards Differences By Year

Class Level Information		
Class	**Levels**	**Values**
YEAR	6	2016 2017 2018 2019 2020 2021

Number of Observations Read	237
Number of Observations Used	237

H0: M1=M2=M3=M4=M5=M6
Ha: at least one mu is not equal

Significance level: 0.05

ANOVA Test Average Intended Air Yards Differences By Year

Dependent Variable: IAY
IAY

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	23.3946663	4.6789333	4.41	0.0007
Error	231	245.2016206	1.0614789		
Corrected Total	236	268.5962869			

R-Square	Coeff Var	Root MSE	IAY Mean
0.087100	12.51161	1.030281	8.234599

Source	DF	Type I SS	Mean Square	F Value	Pr > F
YEAR	5	23.39466633	4.67893327	4.41	0.0007

Source	DF	Type III SS	Mean Square	F Value	Pr > F
YEAR	5	23.39466633	4.67893327	4.41	0.0007

The F Value is 4.41 and the p-value is 0.0007, less than the significance level of 0.05. Therefore, there is strong evidence that at least one season mean IAY is significantly different from another.

ANOVA Test Average Intended Air Yards Differences By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

YEAR	IAY LSMEAN	LSMEAN Number
2016	8.87692308	1
2017	8.33414634	2
2018	8.10512821	3
2019	8.17435897	4
2020	7.99756098	5
2021	7.91842105	6

Least Squares Means for effect YEAR
Pr > |t| for H0: LSMean(i)=LSMean(j)

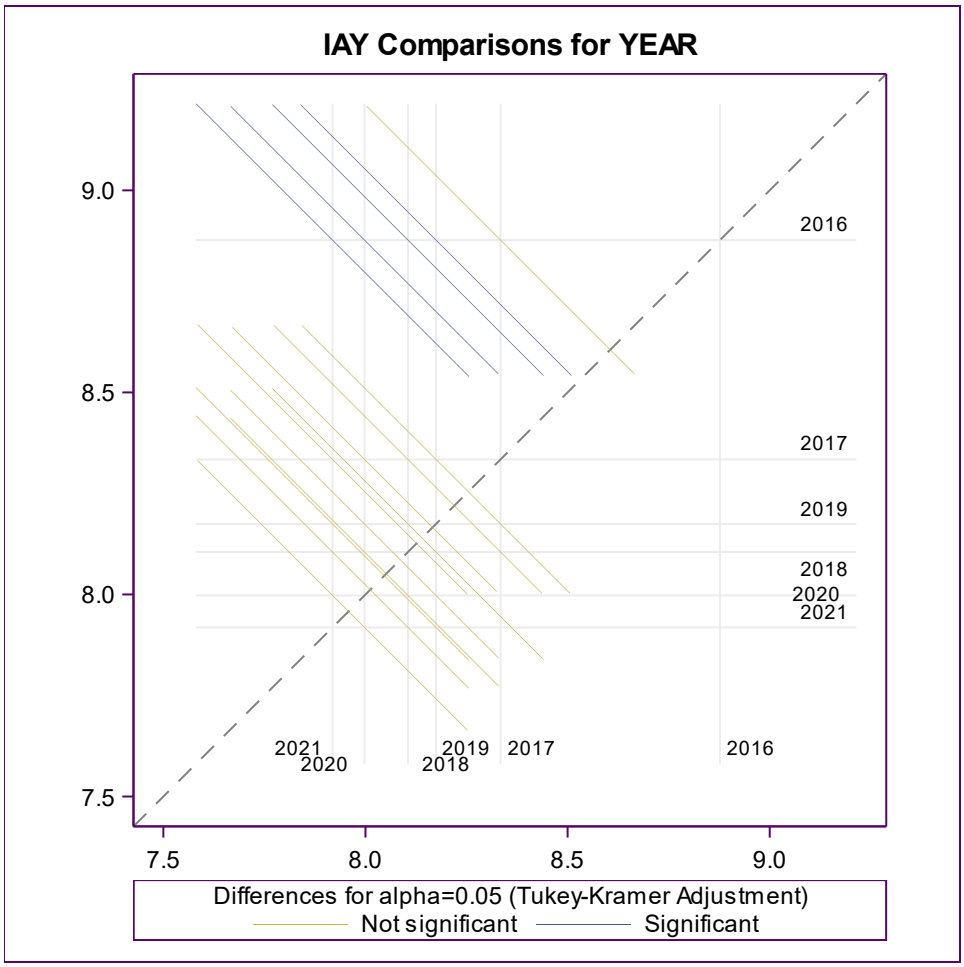
Dependent Variable: IAY

i/j	1	2	3	4	5	6
1		0.1767	0.0137	0.0340	0.0024	0.0009
2	0.1767		0.9196	0.9826	0.6779	0.4730
3	0.0137	0.9196		0.9997	0.9972	0.9682
4	0.0340	0.9826	0.9997		0.9727	0.8851
5	0.0024	0.6779	0.9972	0.9727		0.9994
6	0.0009	0.4730	0.9682	0.8851	0.9994	

The 2016 season mean IAY is significantly different than the 2018, 2019, 2020 and 2021 season. The 2017 to 2021 seasons don’t have any significant differences.

ANOVA Test Average Intended Air Yards Differences By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer



ANOVA Test For Aggressiveness Average Differences By Year

Class Level Information		
Class	**Levels**	**Values**
YEAR	6	2016 2017 2018 2019 2020 2021

Number of Observations Read	237
Number of Observations Used	237

H0: M1=M2=M3=M4=M5=M6
Ha: at least one mu is not equal

Significance level: 0.05

ANOVA Test For Aggressiveness Average Differences By Year

Dependent Variable: AGG
AGG

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	447.697289	89.539458	10.30	<.0001
Error	231	2007.829884	8.691904		
Corrected Total	236	2455.527173			

R-Square	Coeff Var	Root MSE	AGG Mean
0.182322	17.57708	2.948204	16.77300

Source	DF	Type I SS	Mean Square	F Value	Pr > F
YEAR	5	447.6972887	89.5394577	10.30	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
YEAR	5	447.6972887	89.5394577	10.30	<.0001

The F Value is 10.30 and the p-value is <0.0001, less than the significance level of 0.05. Therefore, there is strong evidence to conclude that at leaset one of the season’s mean aggressiveness is significantly different from at least one season.

ANOVA Test For Aggressiveness Average Differences By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

YEAR	AGG LSMEAN	LSMEAN Number
2016	19.3230769	1
2017	17.7073171	2
2018	15.8384615	3
2019	16.5205128	4
2020	16.0463415	5
2021	15.1500000	6

Least Squares Means for effect YEAR
Pr > |t| for H0: LSMean(i)=LSMean(j)

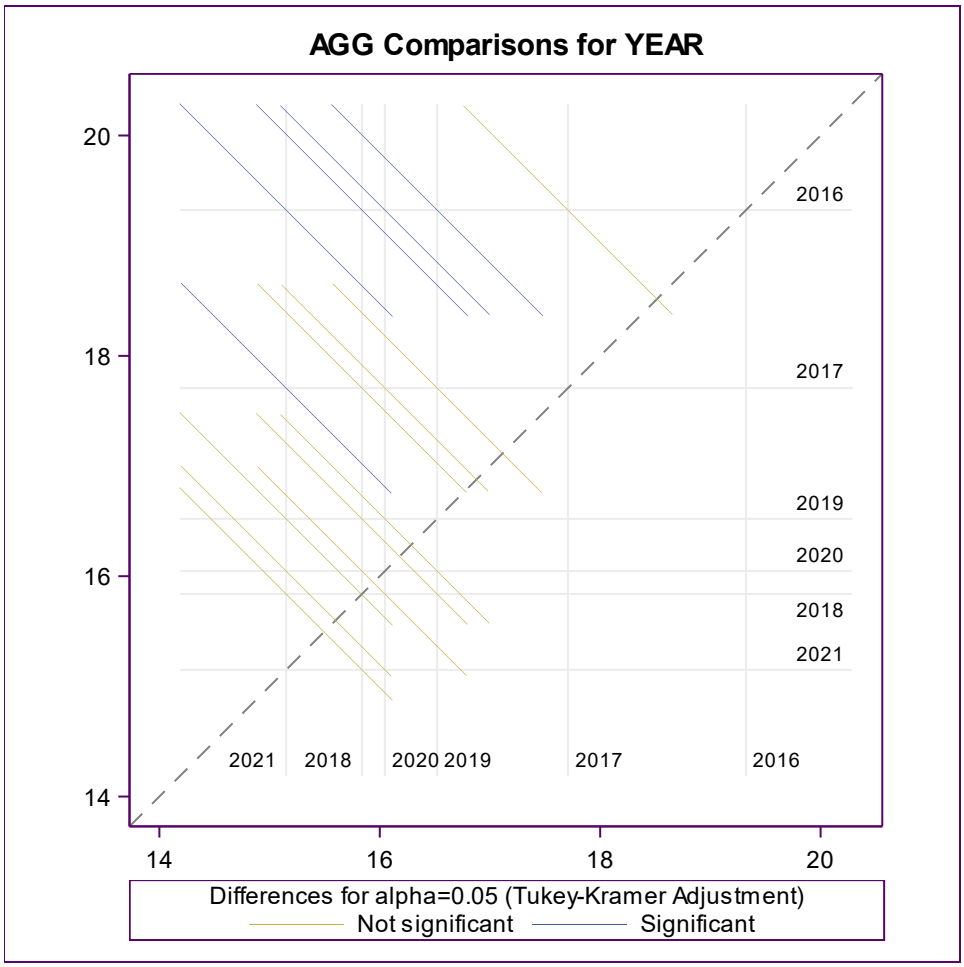
Dependent Variable: AGG

i/j	1	2	3	4	5	6
1		0.1437	<.0001	0.0005	<.0001	<.0001
2	0.1437		0.0557	0.4680	0.1140	0.0021
3	<.0001	0.0557		0.9104	0.9996	0.9094
4	0.0005	0.4680	0.9104		0.9795	0.3234
5	<.0001	0.1140	0.9996	0.9795		0.7565
6	<.0001	0.0021	0.9094	0.3234	0.7565	

The 2016 season’s mean AGG is significantly different than the 2018-2021 seasons’. The 2017 season’s mean AGG is significantly different than the 2021 season.

ANOVA Test For Aggressiveness Average Differences By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer



ANOVA Test For Air Yards to Sticks Differences By Year

Class Level Information		
Class	**Levels**	**Values**
YEAR	6	2016 2017 2018 2019 2020 2021

Number of Observations Read	237
Number of Observations Used	237

H0: M1=M2=M3=M4=M5=M6
Ha: at least one mu is not equal

Significance level: 0.05

ANOVA Test For Air Yards to Sticks Differences By Year

Dependent Variable: AYTS
AYTS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	19.5794703	3.9158941	3.74	0.0028
Error	231	241.9810360	1.0475370		
Corrected Total	236	261.5605063			

R-Square	Coeff Var	Root MSE	AYTS Mean
0.074856	-147.5473	1.023493	-0.693671

Source	DF	Type I SS	Mean Square	F Value	Pr > F
YEAR	5	19.57947032	3.91589406	3.74	0.0028

Source	DF	Type III SS	Mean Square	F Value	Pr > F
YEAR	5	19.57947032	3.91589406	3.74	0.0028

The F Value is 3.74 and the p-value is 0.0028, less than the significance level of 0.05. Therefore, there is strong evidence to conclude that there's at least one season's mean AYTS is different from at least one other season.

ANOVA Test For Air Yards to Sticks Differences By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

YEAR	AYTS LSMEAN	LSMEAN Number
2016	-0.07435897	1
2017	-0.66829268	2
2018	-0.86666667	3
2019	-0.85384615	4
2020	-0.76829268	5
2021	-0.93421053	6

Least Squares Means for effect YEAR
Pr > |t| for H0: LSMean(i)=LSMean(j)

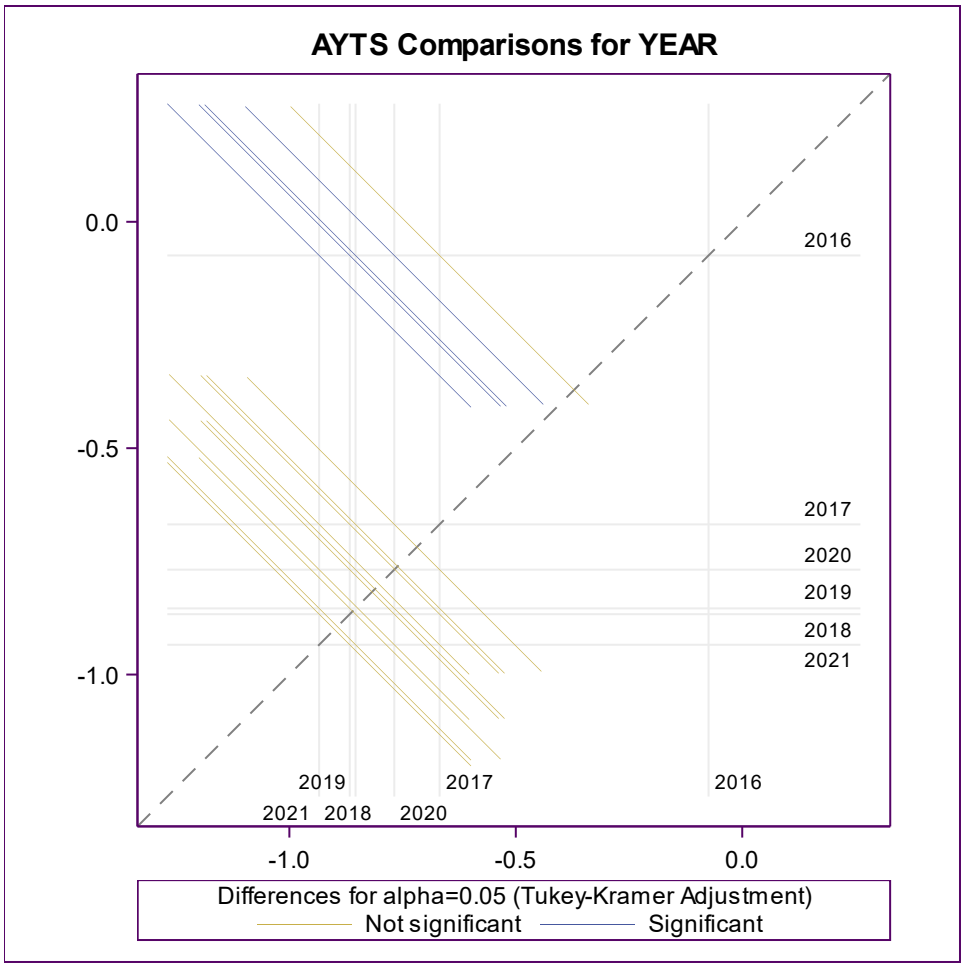
Dependent Variable: AYTS

i/j	1	2	3	4	5	6
1		0.1028	0.0096	0.0115	0.0320	0.0038
2	0.1028		0.9540	0.9654	0.9978	0.8580
3	0.0096	0.9540		1.0000	0.9981	0.9997
4	0.0115	0.9654	1.0000		0.9990	0.9994
5	0.0320	0.9978	0.9981	0.9990		0.9794
6	0.0038	0.8580	0.9997	0.9994	0.9794	

The 2016 season AYTS mean is significantly different than the 2018-2021 seasons.

ANOVA Test For Air Yards to Sticks Differences By Year

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer



ANOVA Test For Completion Percentage Above Expectation Differences By Year

Class Level Information		
Class	**Levels**	**Values**
YEAR	6	2016 2017 2018 2019 2020 2021

Number of Observations Read	237
Number of Observations Used	237

H0: M1=M2=M3=M4=M5=M6
Ha: at least one mu is not equal

Significance level: 0.05

ANOVA Test For Completion Percentage Above Expectation Differences By Year

Dependent Variable: CPAE
CPAE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	321.498168	64.299634	5.61	<.0001
Error	231	2646.098288	11.454971		
Corrected Total	236	2967.596456			

R-Square	Coeff Var	Root MSE	CPAE Mean
0.108336	-738.6106	3.384519	-0.458228

Source	DF	Type I SS	Mean Square	F Value	Pr > F
YEAR	5	321.4981681	64.2996336	5.61	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
YEAR	5	321.4981681	64.2996336	5.61	<.0001

The F Value is 5.61 and the p-value is <0.0001, less than the significance level of 0.05. Therefore, there's strong evidence to conclude that at least one season's mean CPAE is significantly different from at least one other season's mean CPAE.

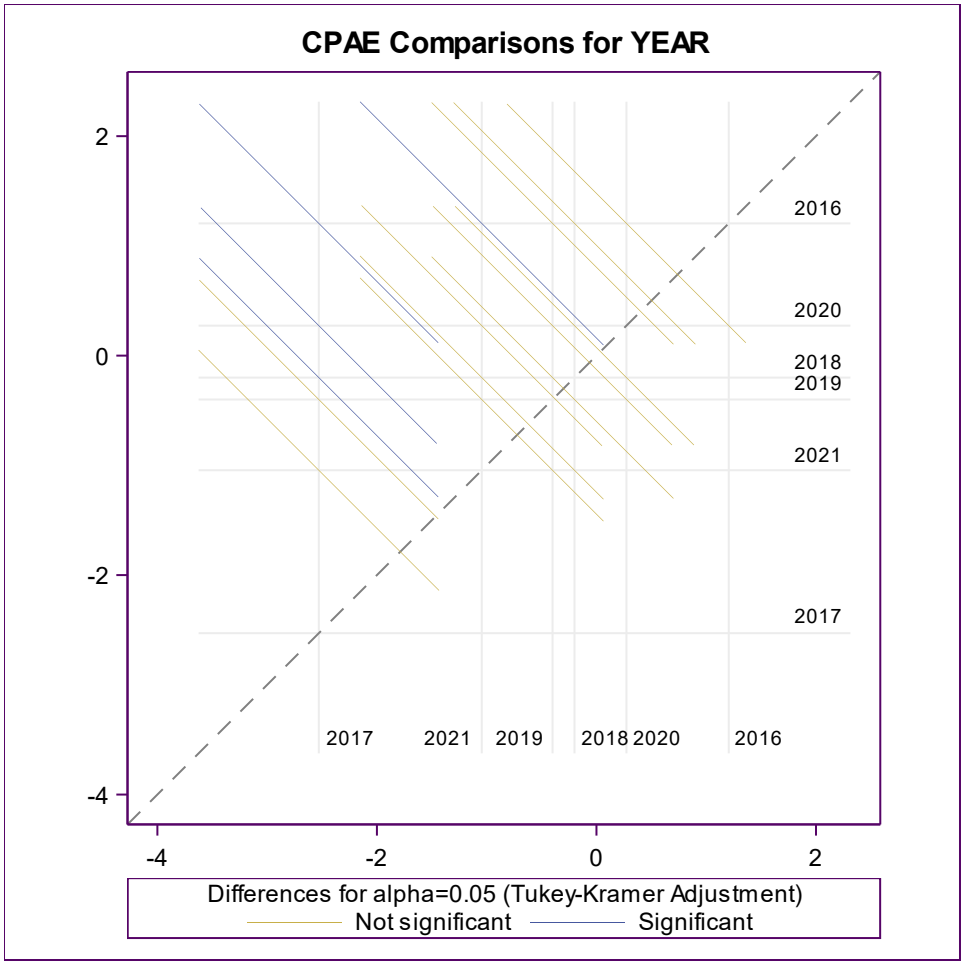
YEAR	CPAE LSMEAN	LSMEAN Number
2016	1.20512821	1
2017	-2.52926829	2
2018	-0.20000000	3
2019	-0.40000000	4
2020	0.27317073	5
2021	-1.04473684	6

Least Squares Means for effect YEAR
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: CPAE

i/j	1	2	3	4	5	6
1		<.0001	0.4464	0.2939	0.8212	0.0445
2	<.0001		0.0280	0.0590	0.0030	0.3758
3	0.4464	0.0280		0.9998	0.9891	0.8830
4	0.2939	0.0590	0.9998		0.9488	0.9606
5	0.8212	0.0030	0.9891	0.9488		0.5139
6	0.0445	0.3758	0.8830	0.9606	0.5139	

The 2016 season mean CPAE is significantly different than the 2017 and 2021 seasons. The 2017 season mean CPAE is significantly different than the 2016, 2018 and 2020 seasons.


CPAE Comparisons for YEAR
2
0
-2
-4
2016
2020
2018
2019
2021
2017
2017
2021
2019
2018
2020
2016
-4
-2
0
2
Differences for alpha=0.05 (Tukey-Kramer Adjustment)
Not significant
Significant

Summary

For the question of interest, several ANOVA tests showed there is strong evidence to conclude that the QB style and tendencies changed a significant amount through the past several years. However, the 2016 season was generally the only year that was significantly than several other seasons. The passer's in the 2016 season were generally far more aggressive passing the ball, especially down the field and at the first down marker the other seasons where the data is tracked. It's a peculiar result, but perhaps there were changes among teams in recent years to de-emphasis the aggressiveness nature of the passers in the 2016 season. Also, defenses could have started to employ more coverages that were to protect against deep plays, forcing teams to adjust and not getting back to the highs the 2016 season generally saw.

Of course, football is a game with a plethora of variables, some quantifiable and some not, that can have an effect on the game unbeknownst to even the most observant view. With that said, the analysis could be improved by taking a look at the quality of individual QB play as maybe there's been a decline is talent level of passers for the past several years. There also could be an emphasis to seek data for offensive and defensive plays, such as are deeper routes being called at a lower rate in recent years or have defenses relaxed off blitzing and allow passers more time in the pocket but gain less yards. The next generation stats are still relatively new and hopefully with more data there could be better, in-depth analysis of how the game is being played.

Biography

Hello, my name is Eric Polverari and I am a graduate student at Clemson University, currently pursuing a Master's Degree in Data Science and Analytics. I am hoping to break into the Data Science and Analytics field as my passion for statistics and analysis drives me to pursue a rewarding career.