[redacted]

Dr. Jacob

Math 1223

3/4/2021

College Football Player Passing Stats 2020

The data set I chose is the college football player passing stats for 2020. The data set includes all conferences that played this year, and the data can be found at https://www.espn.com/college-football/stats/player. 2020 was a crazy year in college football and many teams missed games because of coronavirus so think it will be interesting digging into these passing stats. Variables I have interest in and that I will be using in my analysis for the categorical variables will be what class they are in and the conference they play in. The quantitative variables that are going to be used will be the passer rating (RTG) and the passing touchdowns (TD). I am interested in this data because I really like college football and this data set caught my eye while I was looking. In conclusion something I hope to find using this data set would be if passer rating (RTG) has a correlation with the amount of passing touchdowns (TD).

RK	NAME	Class	TEAM	Conference	POS	CMP	ATT	CMP%	YDS	AVG	LNG	TD	INT	SACK	RTG
1	Mac Jones	JR	ALA	SEC	QB	311	402	77.4	4,500	11.2	90	41	4	13	203.1
2	Kyle Trask	SR	FLA	SEC	QB	301	437	68.9	4,283	9.8	71	43	8	20	180
3	Zach Wilson	JR	BYU	Ind	QB	247	336	73.5	3,692	11	78	33	3	11	196.4

4	Sam Howell	SO	UNC	ACC	QB	237	348	68.1	3,586	10.3	87	30	7	33	179.1
5	Dillon Gabriel	SO	UCF	American	QB	248	413	60	3,570	8.6	93	32	4	22	156.3
6	Brady White	SR	MEM	American	QB	254	420	60.5	3,380	8	92	31	10	20	147.7
7	Matt Corral	SO	MISS	SEC	QB	231	326	70.9	3,337	10.2	91	29	14	18	177.6
8	Trevor Lawrence	JR	CLEM	ACC	QB	231	334	69.2	3,153	9.4	83	24	5	15	169.2
9	Shane Buechele	SR	SMU	American	QB	242	370	65.4	3,095	8.4	85	23	6	20	152.9
10	Spencer Rattler	FR	OU	Big 12	QB	214	317	67.5	3,031	9.6	61	28	7	22	172.6
11	Carson Strong	SO	NEV	MWC	QB	249	355	70.1	2,858	8.1	85	27	4	20	160.6
12	Ian Book	SR	ND	ACC	QB	228	353	64.6	2,830	8	75	15	3	25	144.3
13	Brock Purdy	JR	ISU	Big 12	QB	243	365	66.6	2,750	7.5	65	19	9	13	142.1
14	D'Eriq King	SR	MIAMI	ACC	QB	211	329	64.1	2,686	8.2	89	23	5	26	152.7
15	Malik Cunningham	JR	LOU	ACC	QB	195	304	64.1	2,617	8.6	82	20	12	28	150.3
16	Jarret Doege	JR	WVU	Big 12	QB	239	374	63.9	2,587	6.9	70	14	4	20	132.2
17	Sam Ehlinger	SR	TEX	Big 12	QB	194	322	60.2	2,566	8	78	26	5	22	150.7
18	Phil Jurkove	SO	BC	ACC	QB	205	336	61	2,558	7.6	77	17	5	24	138.7
19	Grayson McCall	FR	CCU	Sun Belt	QB	172	250	68.8	2,488	10	75	26	3	11	184.3
20	Bo Nix	SO	AUB	SEC	QB	214	357	59.9	2,415	6.8	91	12	7	21	123.9
21	Kenny Pickett	SR	PITT	ACC	QB	203	332	61.1	2,408	7.3	75	13	9	18	129.6
22	Connor Bazelak	FR	MIZ	SEC	QB	218	324	67.3	2,366	7.3	69	7	6	12	132
23	Desmond Ridder	JR	CIN	American	QB	186	281	66.2	2,296	8.2	60	19	6	18	152.9
24	Kellen Mond	SR	TA&M	SEC	QB	188	297	63.3	2,282	7.7	54	19	3	7	146.9
25	Cornelious Brown IV	FR	GAST	Sun Belt	QB	178	304	58.6	2,278	7.5	65	17	10	12	133.4

PART 2

For the second part of the project, I am looking at the frequency and relative frequency of the player's conferences. The table below shows the data I collected. There are a total of 7 different conferences in the data I collected. The 7 conferences are the ACC, American, Big 12, Ind, MWC, SEC, Sun Belt conferences. The ACC had the highest percentage of quarterbacks at 28%.

Table 2.1

Frequency	Frequency	Relative Frequency
ACC	7	28%
American	4	16%
Big 12	4	16%
Ind	1	4%
MWC	1	4%
SEC	6	24%
Sun Belt	2	8%
	25	

I also created a two-way table to show what conference and class players are in. The two-way table is listed below. This is a great table to see if any of the conferences are playing a certain class of quarterback. The ACC has 3 senior quarterbacks which could have a relationship with the relative frequency. The Sun Belt has 2 Freshman quarterbacks in the top 25 for passing stats for 2020 which I found interesting because that is a smaller conference compared to the SEC, ACC, and Big 12 conferences.

Table 2.2

Two Way Table	FR	SO	JR	SR
SEC	1	2	1	2
Ind	0	0	1	0
ACC	0	2	2	3
American	0	1	1	2
Big 12	1	0	2	1
MWC	0	1	0	0
Sun Belt	2	0	0	0

PART 3

For the third part of the project, I am looking at passing touchdowns (TD). There are 25 different quarterbacks I will be looking at that have thrown passing touchdowns in this data set. I

will find the mean, standard deviation, five-number summary and display a histogram and box plot that will be listed below. There are no outliers in this data set you can tell by looking at (Figure 3.2). The distribution of the data is skewed to the right because the mean is larger than the median you can see what I am talking about by looking at (Figure 3.3). I also have the five-number summary listed in (Figure 3.3). The tables below provide great insights into the data I have collected by being very readable.

Table 3.1

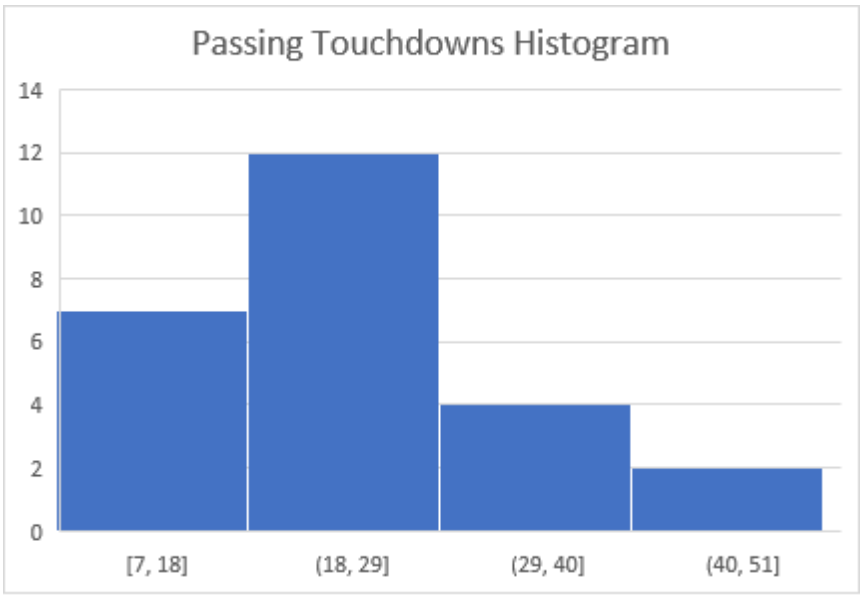
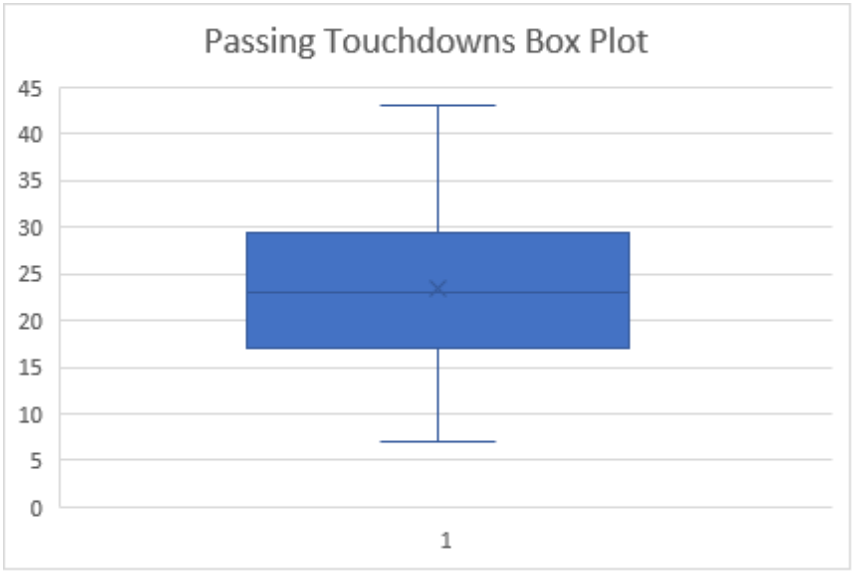


Table 3.3

Minimum	7
Q1	17
Q2 Median	23
Q3	29.5
Maximum	43
Mean	23.52
Range	36
Standard Deviation	8.799242392

Table 3.2



PART 4

For the fourth part of the project, I am creating a hypothesis test for a quantitative and categorical variable. The hypothesis will be listed below. I will be examining passer rating (RTG) and the class of the top 25 quarterbacks.

Quantitative Hypothesis:

My hypothesis is that the average passer rating (RTG) for the top 25 quarterbacks for 2020 is 150. My alternative hypothesis is that it does not equal 150. Passer rating is calculated using a player's passing attempts, completions, yards, touchdowns, and interceptions so I thought it was the best piece of data to create a hypothesis out of.

H_0: $\mu = 150$

Ha: $\mu \neq 150$

Categorical Hypothesis:

My hypothesis is that out of the top 25 quarterbacks for 2020 30% are Seniors (SR). My alternative hypothesis is that it is greater than or equal to 30%. The longer you have played the better your stats usually are so that is why I think seniors make up 30% of the top 25 quarterbacks for 2020.

H_0: $p = 0.30$

Ha: $p \geq 0.30$

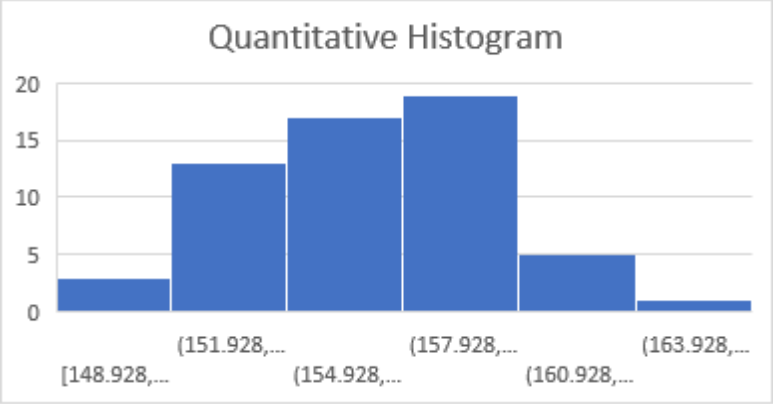
PART 5

For the fifth part of the project, I am computing the standard error for my quantitative and categorical hypothesis using bootstrapping. I will also reject or fail my null hypothesis depending on what my results are.

Quantitative Variable:

The standard error for my quantitative variable is 3.311702. The 95% confidence interval had a lower number of 149.7288 and an upper number of 162.9756. After reviewing the data, I fail to reject the null hypothesis. (Figure 5.1) is a histogram for the quantitative variable.

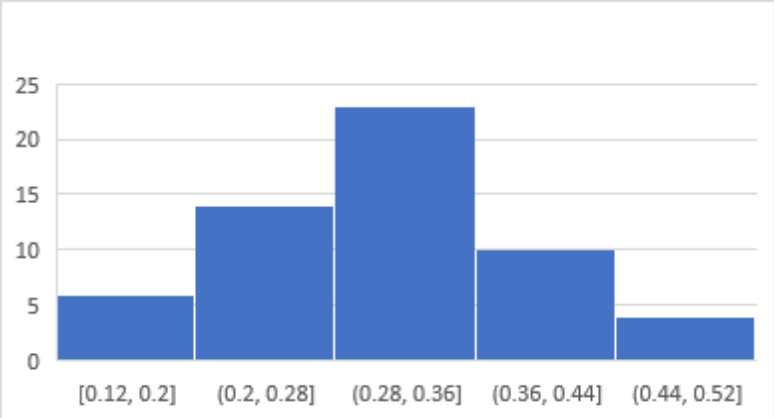
Table 5.1



Categorical Variable:

The standard error for my categorical variable is 0.088247. The 95% confidence interval had a lower number of 0.153331 and an upper number of 0.506318. After reviewing the data, I fail to reject the null hypothesis. (Figure 5.2) is a histogram for the categorical variable.

Table 5.2



PART 6

For the sixth part of the project, I am repeating the hypothesis test on the categorical variable utilizing the appropriate formulas. I will also compare my 95% confidence interval from my bootstrapping in part 5 to this one in part 6 using formulas.

Table 6.1

Statistic	Formula	Result
N	Sample Size	25
X	Number of times (SR) appears	8
Proportion	(in hypothesis)	0.30
P Value	=NORM.S. DIST (Z,True)	0.58637
P Hat	=X/N	0.32
Standard Error	=SQRT (p*(1-p)/n	0.091652
Z Statistic	=(P-hat-p)/SE	0.218218
Z*	=NORM.S.INV (0.975)	1.959963985
CI low	=p-hat-(2*SE)	0.136697
CI high	=p-hat+(2*SE)	0.503303

The 95% confidence interval had a lower number of 0.153331 and an upper number of 0.506318 using bootstrapping and using the formulas the lower number was 0.136697 and the upper number was 0.503303. (Figure 6.1) shows what I did to find my results. I fail to reject the null hypothesis because 30% falls within the 95% confidence interval.

PART 7

For the seventh part of the project, I am repeating the hypothesis test on the quantitative variable utilizing the appropriate formulas. I will also compare my 95% confidence interval from my bootstrapping in part 5 to this one in part 7 using formulas. Part 7 will be very similar to part 6 as in how I set them up.

Table 7.1

Statistic	Formula	Result
MU	(in hypothesis)	150
X-bar	(sum of the numbers/n)	156.38
Sigma	(standard deviation)	21.35749283
N	Sample size	25
Standard Error	Sigma/SQRT(n)	4.271498566
T Statistic	X-bar-mu/SE	121.2635186
Alpha	For 95% CI	0.05
T*	T.INV (alpha, n-1)	1.71088208
CI Low	X-bar-T-Critical Value*SE	149.0719696
CI High	X-bar+T-Critical Value*SE	163.6880304

The 95% confidence interval had a lower number of 149.7288 and an upper number of 162.9756 using bootstrapping and using the formulas the lower number was 149.0719696 and the upper

number was 163.6880304. (Figure 7.1) shows what I did to find my results. I fail to reject the null hypothesis because 150 falls within the 95% confidence interval.

PART 8

For eighth and final part of this project, I am going to use the two-way table from part 2 to create 2 conditional probabilities. I will interpret their meaning and explain how they were computed.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Table 8.1

Two Way Table	FR	SO	JR	SR	**Total**
SEC	1	2	1	2	**6**
Ind	0	0	1	0	**1**
ACC	0	2	2	3	**7**
American	0	1	1	2	**4**
Big 12	1	0	2	1	**4**
MWC	0	1	0	0	**1**
Sun Belt	2	0	0	0	**2**
Total	**4**	**6**	**7**	**8**	

I am going to be determining the probability of SEC players given they are a Senior.

P=2/8 ≈ 25%

I am going to be determining the probability of ACC players given they are a Sophomore.
P=2/6 ≈ 33%

My first conditional probability looks at the probability of SEC players that are a senior from the data set I used throughout the last 8 weeks. The percentage was approximately 25% meaning 2 out of 8 SEC players were seniors. This statistic shows the disadvantage that SEC seniors are at being in the top 25 of college passing stats for 2020.

My second conditional probability looks at the probability of ACC players that are a sophomore from the data set I used throughout the last 8 weeks. The percentage was approximately 33% meaning 2 out of 6 ACC players were sophomores. This statistic shows the advantage that ACC sophomores are at being in the top 25 of college passing stats for 2020.