[redacted]

Class: SP21-INTRO TO PROBABILITY AND STATS_01

Instructor: Nicholas Jacob

February 21, 2021

Odds of a Highschool Volleyball Player competing in College 2020

Part 1

The data provided for this project originated from Dr. Jacob's GitHub and it is called Southern Volleyball College Teams 2020. I will examine the data from odds of a High School volleyball player competing in College 2020. This data is appealing to me because I am currently competing in College and I would like to learn more about what are the chances of high school volleyball players to have the same opportunity. In the 2019-2020 year, there were 477,117 (f) and 67,243 (m) High School volleyball players, while 28,796 (f) and 3,929 (m) College volleyball players. There is a 5.8% and 17:1 ratio of a female High School volleyball player making any College roster; 5.6% and 18:1 ratio for men.

I have decided to compile a list of all the schools in the United States that sponsor a varsity Volleyball team during 2020. There are many variables that can influence to a High School volleyball player having a chance to play in College. Some of the quantitative data that will be presented in the table below (table 1) are the scholarships for both men and women, college admission rate and tuition. The categorical data includes the division of the school, and the state and city where the school is located. This data can be used to compare the chances of High School volleyball players.

Table 1

Southern Schools with Varsity Volleyball Teams 2020:	City	State	Division	Tea	ms	Avr Men's Scholarship	Avr Women's Scholarship	Avr SAT Math	Avr ACT Score	Admission Rate	In-State Tuition	Out of State Tuition
Cameron University	Lawton	OK	NCAA II		W	6,483	7,648				6,450	15,870
East Central University	Ada	OK	NCAA II		W	4,295	5,495	505	21	83%	6,810	16,020
Langston University	Langston	OK	NAIA		W	1,622	1,752				6,226	13,586
Mid-America Christian Univ	Oklahoma City	OK	NAIA		W	6,923	6,242				19,604	19,604
Northeastern Oklahoma A&M	Miami	OK	NJCAA		W	3,551	3,109				4,758	11,170
Northwestern Oklahoma State Univ	Alva	OK	NCAA II		W	2,700	2,959	570	20	62%	7,471	14,558
Oklahoma Baptist University	Shawnee	OK	NCAA II		W	12,011	11,046	535	23	64%	28,258	28,258
Oklahoma Christian University	Edmond	OK	NCAA II		W	8,487	11,152	575	24	65%	22,760	22,760
Oklahoma City University	Oklahoma City	OK	NAIA		W	17,092	16,575	580	26	76%	31,026	31,026
Oklahoma Panhandle State Univ	Goodwell	OK	NAIA		W	2,417	3,435				7,930	8,674

My conclusion from looking at this data is that women have a better chance than men when joining a College roster. It also shows that the lower division, the less athletes and scholarships there are. I want to further examine what are the chances of all these High School volleyball players to get into any program and get a scholarship, because it shows that not all athletes receive a scholarship. I want to see what the possibilities for these athletes are to be able to receive financial help.

Part 2

For the second part of the project, I used the school's division and the state where they are located to find their frequency and relative frequency. These tables (2.1 & 2.2) can be used to compare the data.

Table 2.1

Division	Frequency	Relative frequency
NCAA II	11	0.3667
NAIA	6	0.2
NJCAA	13	0.4333
TOTAL	30	1

Table 2.2

State	Frequency	Relative frequency
OK	10	0.333
FL	10	0.333
TX	10	0.333
TOTAL	30	1

In the tables above (2.1 & 2.2), it shows the number of schools that belong to each division. This data shows that there are mostly National Junior College Athletic Association (NJCAA) as well as National Collegiate Athletic Association Division II (NCAA II). National Association of Intercollegiate Athletics (NAIA) is a lower division with small athletics programs. Many High School volleyball players strive to go to a NCAA II or NJCAA, but NAIA could definitely be a great option for them.

Next, I will compare the number of schools on each division per state.

States	NAIA	NJCAA	NCAA II
Oklahoma	4	1	5
Florida	2	4	4
Texas	0	8	2

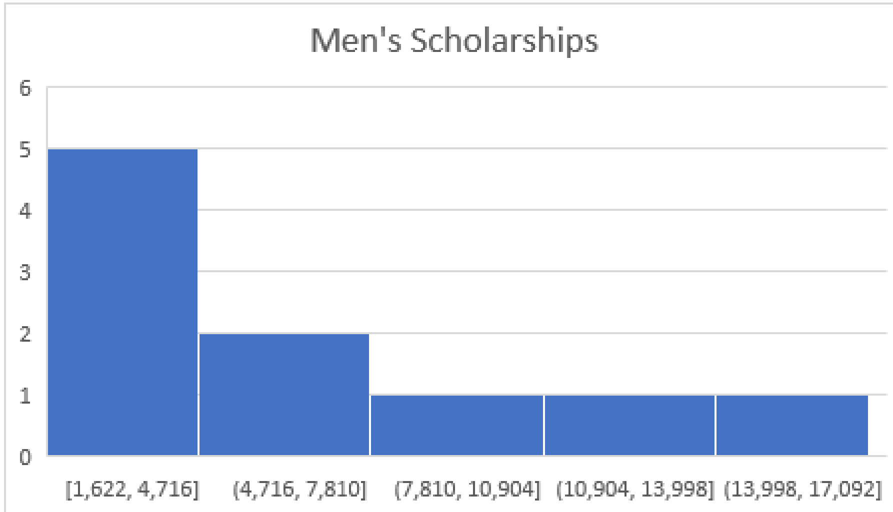
This table was created with the purpose of comparing the numbers of schools for each division and determine what are the chances these High School volleyball players will have of getting into a collegiate volleyball program. From this data you can see that most states are well balanced, except for Texas which had mostly NJCAA schools and no NAIAs whatsoever. This data shows that High School volleyball players that live in the area of Texas unfortunate have less chance to get into an accessible division because there are not many.

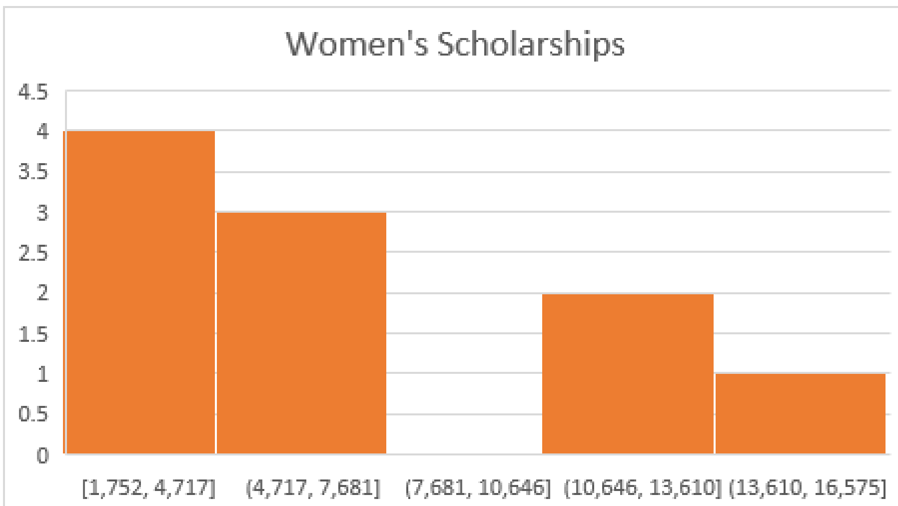
Part 3

For the third part of the project, I decided to create a histogram and a box chart to compare my quantitative data, which in this case are the scholarships given to both men and women. In table 3.1 it shows the number of scholarships given to men and women in the Colleges that I have previously chosen for Oklahoma. Table 3.2 is a box chart that shows the number of scholarships given to both men and women in mentioned schools.

Table 3.1

Table 3.2





The five-number summary for the men's scholarships (blue) is: minimum = 1622, Q1 = 2700, median = 5389, Q3 = 8487 and maximum = 17092. The mean would be 6558.1 and the standard deviation is 4888.99. This chart has no outliers and is skewed right. The five-number summary for the women's scholarships (orange) is: minimum = 1752, Q1 = 3109, median = 5868.6, Q3 = 11046, and maximum = 16575. The mean is 6941.3 and standard deviation is 4716.82. In this table there are not any outliers, and this table is also skewed right.

From these charts we can see how men and women have the same opportunity to receive some kind of aid (scholarship). I thought that since there are not many men's teams, compared to women, they would have a lower chance for a scholarship, but chances are almost equal. There is not a great difference between the minimum and maximum amount of these scholarships. Even the mean of these scholarships is close in range. Using this data to find this information gives me more hope because chances are equal for both men and women to be part of a Collegiate volleyball team.

Part 4

For the fourth part of the project, hypothesis testing, I will continue to use the same data and topic from the previous part (Men and Women's Scholarships). For my quantitative variable, I decided to use the women's average scholarships because it is more common to have Women's teams around this area of Oklahoma.

$$H_{o:}\ \mu = 6{,}941$$

$$H_{a:}\ \mu \neq 6{,}941$$

For the categorical variable, I am going to look at the school divisions. There are three divisions: NAIA, NJCAA, and NCAA II. I have decided to use the number of schools that are NCAA II and compare it to the rest of the schools.

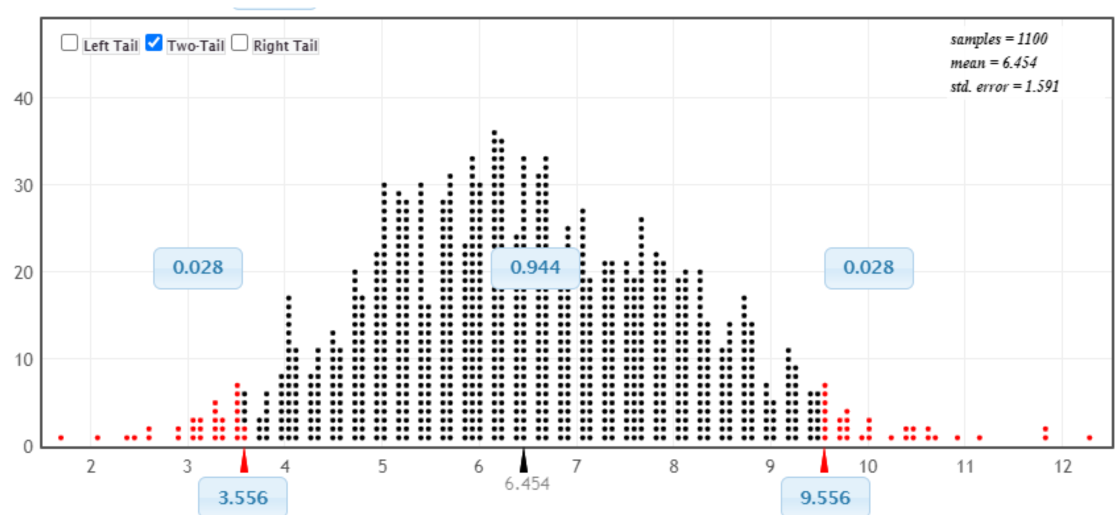
$$H_{o:}\ \mathrm{p} = 1/3$$

$$H_{a:}\ p \neq 1/3$$

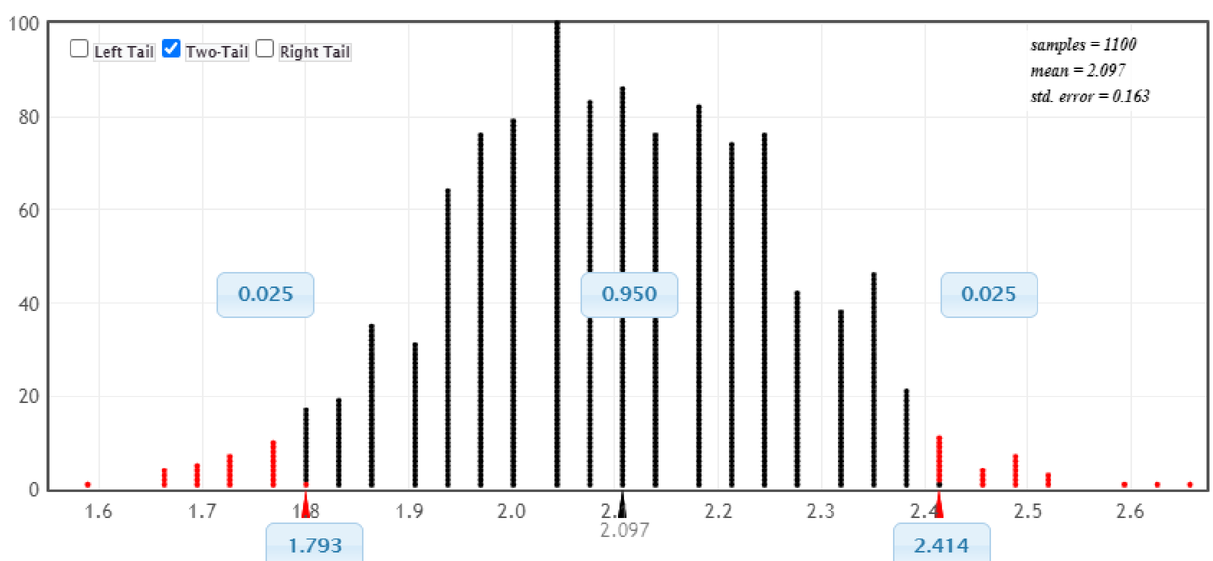
Part 5

For part five of the project, I will be using the bootstrapping method to test my hypothesis. For my quantitative variable, I used the women's average scholarships. I decided to use a sample size of 1,100. As shown in the graph, the standard error is 1,591. The 95%

confidence interval makes the low 3,556 and the high 9,556, which is also presented in the graph. The mean is stated as 6,454, and this would mean that I would have to fail to reject my hypothesis because it is not equal.



For my categorical variable, I would have to see if the number of NCAA II schools is equal to the number of other divisions (schools) which equals would be 1 out of 3, shown in the previous parts of the project. I decided to use a bootstrap sample of 1,100.



The graph shows that the mean is 1.671 and the standard error is 0.0.163. The 95% confidence interval from the graph shows that the low would be 1.793 and the high 2.414. I previously stated in my hypothesis that my number would be .333 (1/3) and I would have to fail to reject my hypothesis because it is close but not quite there.

Part 6

For this part of the project, I will be retesting the hypothesis for my categorical variable. In my null hypothesis, I stated that pC = 1/3, meaning that the number of NCAA II schools would be 1/3 of all the schools in the area, and the alternative being pC ≠ 1/3. The formulas that I used are the ones provided in class. My sample size (n) was 30 schools. Even though there are 30 schools, I tested to see if 1/3 were NCAA Division II. Therefore, my proportion (p) would be .333. In order to find my p hat value, I took the 11 schools that were NCAA II and divided it by the total number of schools. The p hat value was 0.366. now, I would be able to create my test hypothesis. Ho: p = .333 and Ha: p ≠ .333.

I will then find my standard error, z score, p-value and 95CI. The formula to find my z value is **Phat-P/SE**; **SQRT(P*(1-P)/n)** will help me find the SE, and **Phat-z critical*SE** to find the confidence intervals. Then, to change the minus to a plus and to find the z critical I would use **norm.s.inv(1-0.05/2)**. Through excel I found my p-value to be 0.839604. After using this method to retest my hypothesis, I still fail to reject my hypothesis. The proportion fell within the value of the 95% confidence interval, which is not a bad thing.

Sample size (n)	Proportion (p)	P hat	Low 95%	High 95%
30	0.333	0.366		
N		P hat		
z*	p-value	SE		
0.4925389168	0.839604	0.06699978189	0.1990004362	0.4669995638

Part 7

For the next part I will test my quantitative variable. I will be using formulas to find the standard error, x bar - mu, z score, p value, critical z score and the 95% confidence interval. My sample size (n) was ten schools. I wanted to see what was the average of scholaship amount granted by schools. My mean (u) would be 6,941 and my hypothesis would be $H_{o:}$ μ = 6,941 and $H_{a:}$ μ ≠ 6,941.

Next, I will find the standard error, z score, p-value and 95% confidence interval. The formula to find the standard error is SD/SQRT (n). The calculated standard error was 1441.8816218. The x bar – mu calculated was 1.069. The formula to find the z-score is (x bar-mu)/SE which was 0.005510309. Through excel I found the p-value to be 0.995612. Then the t critical score (T*) was 1.281551566. Now, for the 95% confidence interval I used P+/-((T*)*SE)). The lower 6941.3 ± 1.96*1414.8 = 4168.337 and the upper 6941.3 ± 1.96*1414.8 = 9714.263. I fail to reject my hypothesis. Even though it is between the values it was a good guess for the overall population.

Sample size (n)	Standard deviation	T score	Low 95%	High 95%
10	4474.764	0.005510309		
SE	p-value	T*		
1441.8816218	0.995612	1.281551566	4198.337	9714.263

Part 8

For the last part of the project I will work with conditional probability and the two way table from part two. This table contains the school's divisions by State.

States	NAIA	NJCAA	NCAA II	Total
Oklahoma	4	1	5	10
Florida	2	4	4	10
Texas	0	8	2	10
Total	**6**	**13**	**11**	**30**

With this 2-way table, I will create a conditional probability. For this probability I will test if one school was chosen, what is the probability that the school is NCAA II in Oklahoma. First, I will be looking at the probability of getting a school that is NCAA Division II **(P(B)) = 11/30 = 0.3666.** Next, I will look at how many NCAA II are available in Oklahoma **P(A and B) 5/30 = 0.1666.** Now, to find **P(A|B) = A cap B/P(B) = 0.1666/0.3666 = 0.4544.** Meaning there is a 45% chance of choosing a school that is NCAA Division II.