[redacted]

Intro to Probability and Statistics 1223

Dr. Jacob

23 August 2020

2008-2009 Bedlam Men's Basketball Statistics

I was setting trying to think of what I should do for my report. I went through my interest to try and find something that I would find interesting. So, I thought of the OU and OSU men's basketball team when Blake Griffin was playing. I love basketball and grew up watching it with family. My family is a very big Oklahoma Sooners fan, so I thought what better to do than the bedlam teams. I used to watch the sooners and developed the passion by watching Blake Griffin. So, I chose to do the statistics over Oklahoma University and Oklahoma State University men's basketball teams from 2008-2009.

I found these articles by looking the statistics for both teams and looking at the rosters for that year. Here are both websites for the two basketball teams.

Oklahoma University:https://www.espn.com/mens-college-basketball/team/stats/_/id/201/season/2009/table/game/sort/gamesPlayed/dir/desc, and the Oklahoma State University: https://www.espn.com/mens-college-basketball/team/stats/_/id/197/season/2009.

There are many variables that can be used to test different ideas. Some of the quantitative data that is going to be used is the field goal percentage (FG%), the free throw percentage (FT%), average points per game (PPG) and the three-point field goal percentage (3P%). These will be used to compare the team averages against each other and then compare the individual player. Other data that can be used it categorical data. Some of the categorical data would be which class they are in, the positions they play, and the team they play for. It can be used to look at comparison between positions and their stats.

Player	Basketball Number	Class	Position	Height	School	Games Played	Average Points Per Game	Field Goal Percentage (Per Game)	Free Throw Percentage (Per Game)	Three-point Field Goal Percentage (Per Game)
Anthony Brown	4	SR	F	6'6"	OSU	33	2.4	0.508	0.594	0.000
Byron Eaton	0	SR	G	5'11"	OSU	35	14.3	0.421	0.773	0.299
Terrel Harris	1	SR	G	6'5"	OSU	35	13.9	0.488	0.785	0.377
Obi Muonelo	2	JR	G	6'5"	OSU	35	12.7	0.416	0.678	0.398
Ibrahima Thomas	32	SO	F	6'11"	OSU	7	8.3	0.490	0.400	0.286
James Anderson	23	SO	G	6'6"	OSU	35	18.2	0.482	0.829	0.408
Malcon Kirkland	33	SO	F	6'7"	OSU	23	2.3	0.415	0.655	0.000
Marshall Moses	33	SO	F	6'7"	OSU	32	7.0	0.562	0.630	0.000
Nick Sidorakis	15	SO	G	6'4"	OSU	33	1.1	0.394	0.500	0.345
Scott Warner	55	SO	C	6'11"	OSU	4	0.0	0.000	0.500	0.000
Blaine Booher	31	FR	G	6'4"	OSU	3	0.7	1.000	0.000	0.000
Garrett Thomas	25	FR	G	6'2"	OSU	3	0.0	0.000	0.000	0.000
Greg Hughes	45	FR	F	6'8"	OSU	1	2.0	1.000	0.000	0.000
Keiton Page	12	FR	G	5'9"	OSU	35	8.6	0.432	0.767	0.399
Nolan Cox	53	FR	G	6'6"	OSU	1	0.0	0.000	0.000	0.000
Teeng Akol	22	FR	F	6'11"	OSU	6	0.3	0.250	0.000	0.000
Austin Johnson	20	SR	G	6'3"	OU	36	8.6	0.423	0.781	0.346
Omar Leary	11	SR	G	5'10"	OU	34	1.9	0.352	0.833	0.349
Taylor Griffin	32	SR	F	6'7"	OU	36	9.6	0.536	0.700	0.357
Beau Gerber	45	JR	F	6'8"	OU	13	0.6	0.600	1.000	0.000
Juan Pattillo	12	JR	F	6'6"	OU	19	5.9	0.551	0.743	0.000
Ryan Wright	1	JR	F	6'9"	OU	32	1.8	0.341	0.577	0.000
Tony Crocker	5	JR	G	6'6"	OU	36	9.6	0.394	0.772	0.349
Blake Griffin	23	SO	F	6'10"	OU	35	22.7	0.654	0.590	0.375
Cade Davis	34	SO	G	6'5"	OU	35	4.7	0.383	0.656	0.345
Kyle Cannon	31	SO	F	6'8"	OU	9	1.9	0.286	0.750	0.222
Orlando Allen	21	SO	C	6'11"	OU	19	1.8	0.577	0.400	0.000
Ray Willis	41	FR	G	6'6"	OU	16	3.3	0.372	0.778	0.333
T.J. Franklin	3	FR	G	5'11"	OU	12	0.3	0.000	0.250	0.000
Willie Warren	13	FR	G	6'4"	OU	36	14.6	0.473	0.781	0.372

My conclusion just looking at the data, it shows a clear difference between the stats of different players. It shows that if a player has a good field goal percentage and a good free throw percentage, they may have a bad three-point field goal percentage or vice versa. I want to look at which bedlam team is better with these stats? Also, look at the best players for both teams and look at the correlation and difference between them.

For the second part of the project, I looked at the frequency and relative frequency of the team's positions, class, and height. Each table is comparing between OU and OSU. They are labeled by the color of the team. Below are the tables that contain that data.

Position	Frequency	Relative Frequency
Guard	7	0.5
Forward	6	0.42857143
Center	1	0.07142857
Total	14	1

Position	Frequency	Relative Frequency
Guard	9	0.5625
Forward	6	0.0625
Center	1	0.375
Total	16	1

Class	Frequency	Relative Frequency
Senior	3	0.21428571
Junior	4	0.28571429
Sophomore	4	0.28571429
Junior	3	0.21428571
Total	14	1

Class	Frequency	Relative Frequency
Senior	3	0.1875
Junior	1	0.0625
Sophomore	6	0.375
Junior	6	0.375
Total	16	1

Height	Frequency	Relative Frequency
Unrecorded	0	0
5'8"-5'9"	0	0
5'10"-5'11"	2	0.142857143
6'0"-6'1"	0	0
6'2"-6'3"	1	0.071428571
6'4"-6'5"	2	0.142857143
6'6"-6'7"	4	0.285714286
6'8"-6'9"	3	0.214285714
6'10"-6'11"	2	0.142857143
Total	14	1

Height	Frequency	Relative Frequency
Unrecorded	4	0.25
5'8"-5'9"	1	0.0625
5'10"-5'11"	1	0.0625
6'0"-6'1"	0	0
6'2"-6'3"	0	0
6'4"-6'5"	3	0.1875
6'6"-6'7"	4	0.25
6'8"-6'9"	0	0
6'10"-6'11"	3	0.1875
Total	16	1

From this data it shows that both teams are made up mainly forwards and guards, both only having one center. It also shows that OU's team was spread out when it came to the class data. They had a team with the same number of players for every class. However, for OSU's team, it shows that they had more underclassmen than upperclassmen. This could be a factor in why OU had a better record than OSU. Could it have been because OSU mainly had underclassmen? Lastly for the height, you can see that there are many players with different heights but the majority for both teams were 6'4" or taller. 78.57% of OU's players were 6'4" or taller, while 62.5% of OSU's players were 6'4" or taller.

Next, I compared the players height against the players position, which is shown below.

OU	Forward	Guard	Center
5'8"-5'9"	0	0	0
5'10"-5'11"	0	2	0
6'0"-6'1"	0	0	0
6'2"-6'3"	0	1	0
6'4"-6'5"			
6'6"-6'7"	2	2	0
6'8"-6'9"	3	0	0
6'10"-6'11"	1	0	1

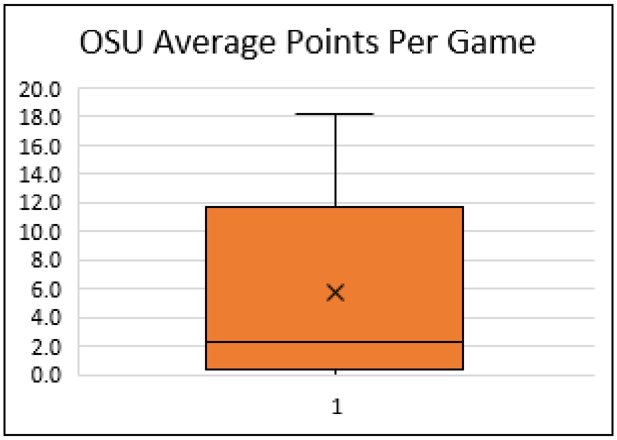
OSU	Forward	Guard	Center
5'8"-5'9"	0	1	0
5'10"-5'11"	0	1	0
6'0"-6'1"	0	0	0
6'2"-6'3"	0	1	0
			0
6'6"-6'7"	3	2	0
6'8"-6'9"	1	0	0
6'10"-6'11"	2	0	1

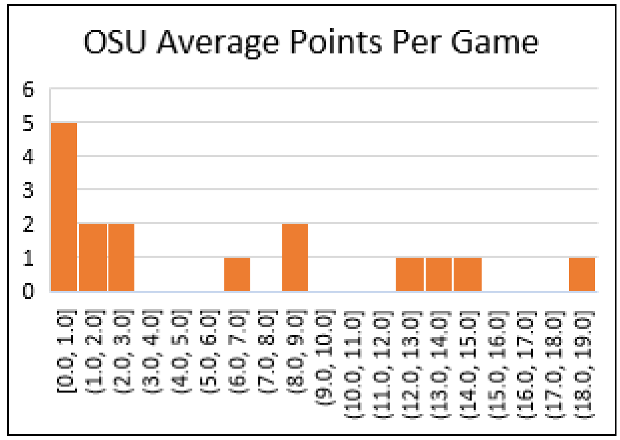
This table was created to compare the height of players to their positions. As you can tell from the graphs the height of the center is very tall, between 6'10" and 6'11". This is due to the player having to be ready on the inside to block up and score and gives the player a more built-in advantage. Also, from the data, you can see that most of the guards are shorter. This is normal. They stay out towards the three-point line and take the advantage of threes and the inside jump shots. Lastly, you notice that the forwards are mainly as tall as the center. This is an advantage for the shooters to take block shots and shorter jump shots.

For the third part of the project I have created a box graph and a histogram for each team for two of my quantitative datasets. In figure 3.1, it shows the histogram and the bar graph of OSU's Average Points Per Game. These graphs show lots of data. From them you can find the five-number summary, the mean, and from the data points, the standard deviation. The five-number summary for figure 3.1 is: Min=0, Q_1=0.4, Median=2.4, Q_3=11.7, and the Max=18.2.

The mean for this is 5.7 and the standard deviation is 6.1763932. With no outliers. The distribution of this data looks skewed to the left.

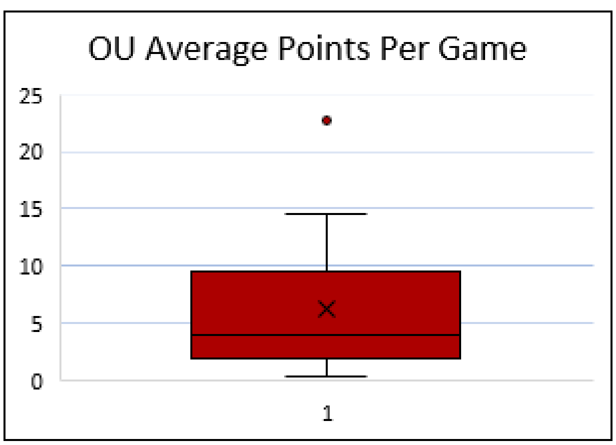
Figure 3.1

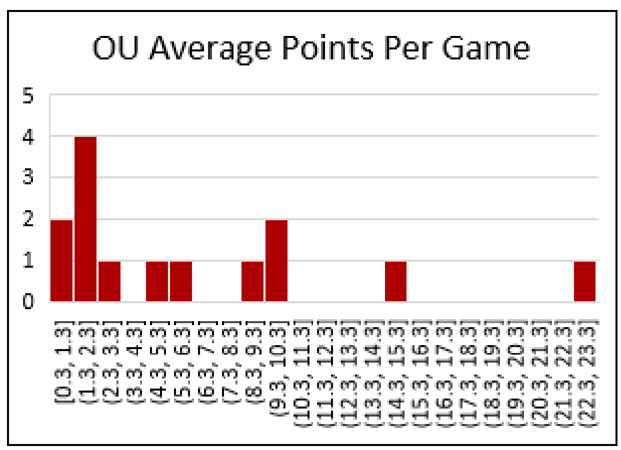




In figure 3.2, it shows the histogram and the bar graph of OU's Average Points Per Game. The five-number summary for figure 3.2 is: Min=0.3, Q_1=1.8, Median=4, Q_3=9.6, and the Max=14.6. The mean for this is 6.235 and the standard deviation is 6.3506035. With only one outlier at the point 22.7. The distribution of this data looks skewed to the left.

Figure 3.2

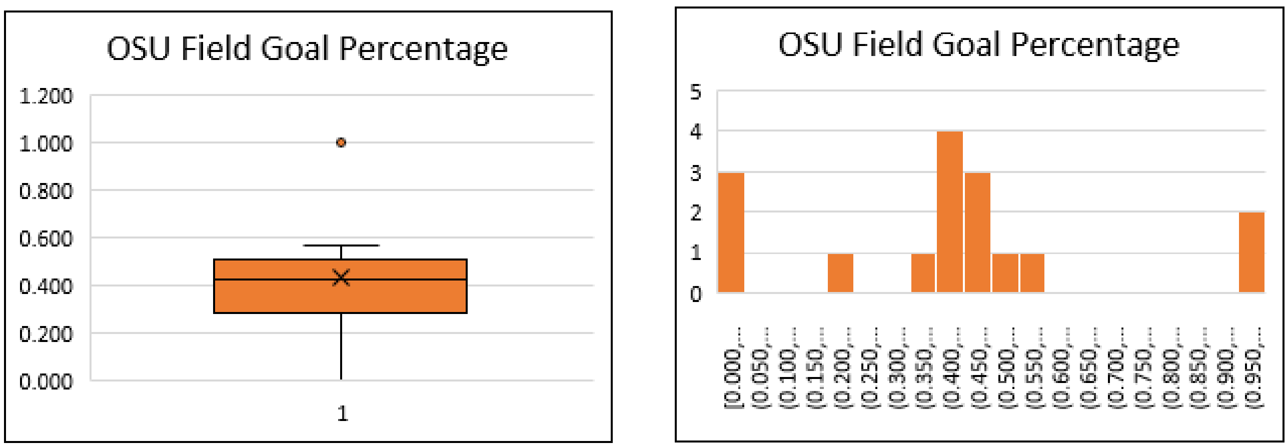




In figure 3.3, it shows the histogram and the bar graph of OSU's Average Field Goal Percentage. The five-number summary for figure 3.3 is: Min=0.000, Q_1=0.286, Median=0.427,

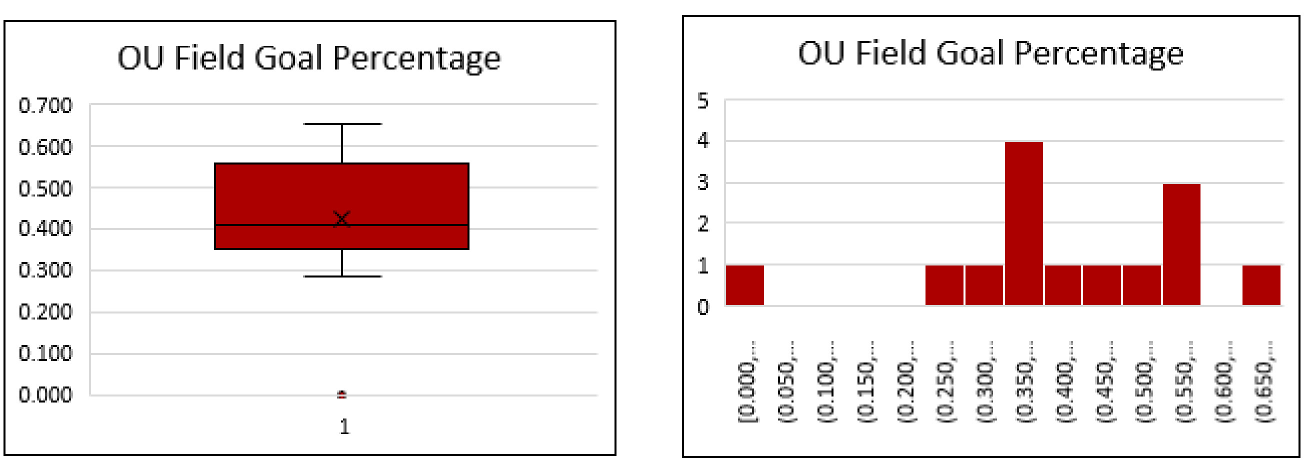
Q_3=0.504, and the Max=0.562. The mean for this is 0.429 and the standard deviation is 0.2911870. With only one outlier at the point 1.000. The distribution of this data looks somewhat symmetric.

Figure 3.3



In figure 3.4, it shows the histogram and the bar graph of OU's Average Field Goal Percentage. The five-number summary for figure 3.4 is: Min=0.286, Q_1=0.349, Median=0.409, Q_3=0.558, and the Max=0.654. The mean for this is 0.424 and the standard deviation is 0.1651422. With only one outlier at the point 0.000. The distribution of this data looks a little skewed to the right.

Figure 3.4



In the fourth part of this project I looked to test a hypothesis about average point per game.

For my quantitative variable, I used the average points per game. This is to see how the average points per game can affect the total games played. You would like to have players who score more, play more from a fan’s perspective.

$$H_{o:}\ \mu = 6.0$$

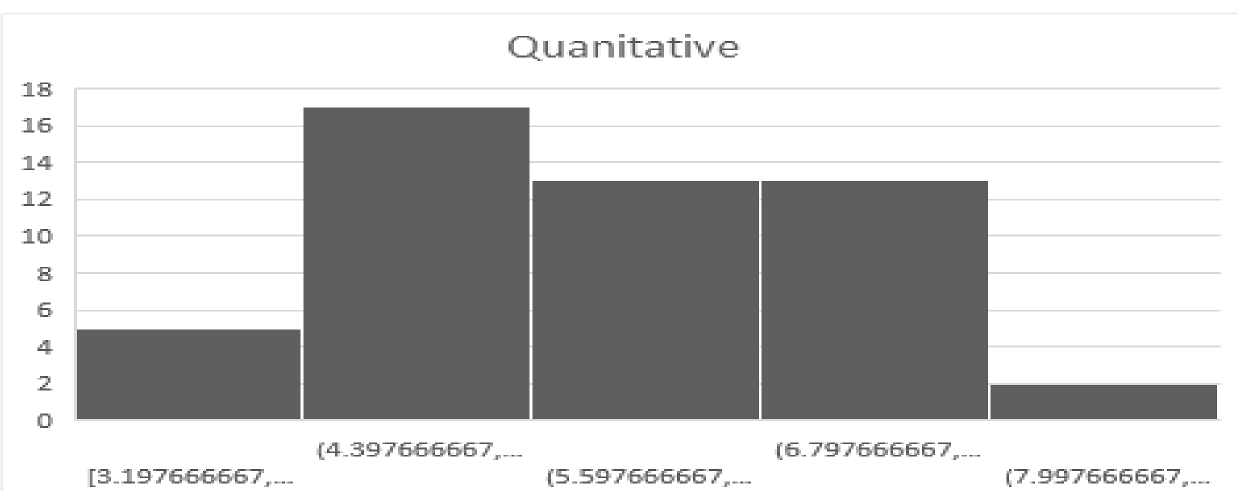
$$H_{a:}\ \mu \neq 6.0$$

For my categorical variable, I am going to look at the positions of players. There are three different positions: forward, guard, and center. If you look through the data, the forwards average more points per game. So, I am going to test to see if forwards are 1/3 of the players on each team.

$$H_{o:}\ p = .33$$

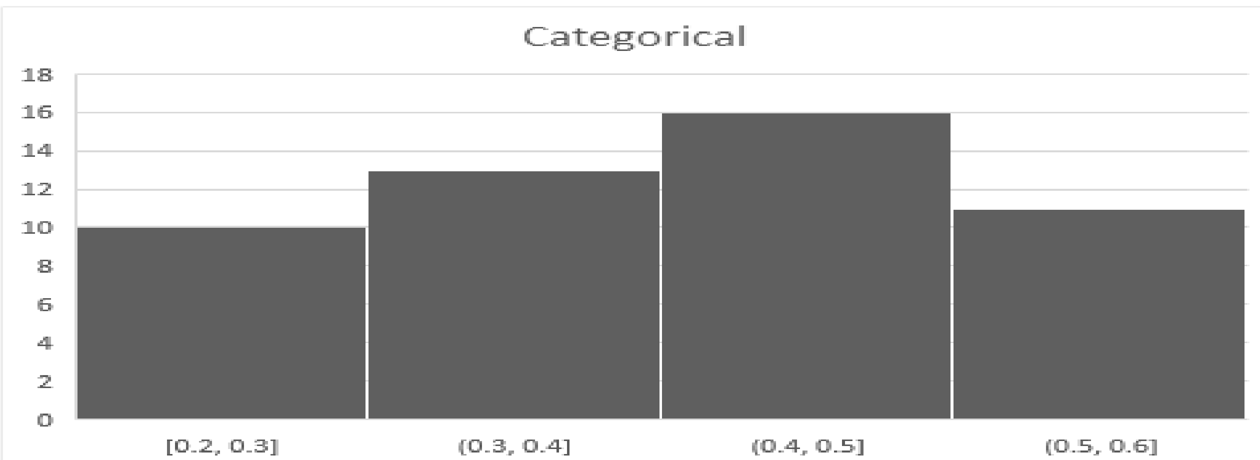
$$H_{a:}\ p \neq .33$$

Part five of this project has you look at the bootstrapping method to test your hypothesis. For the quantitative variable, I took the average points per game for every player for both OU and OSU. With this data I took a bootstrap sample and repeated it 50 times. The bootstrap distribution was 5.93688 and the standard error was 1.24617. With this information I found the 95% confident interval. The lower= 3.44454 and the upper= 8.42922.



Looking at the information, I conclude and fail to reject my hypothesis. My hypothesis stated that it equals 6.0. This value fell within the 95% confident interval.

For the categorical variable, I took the player positions for both OU and OSU. I took each player for OU and OSU and saw that forwards tended to score more. I changed the forwards to 1 and held the guards and center as 0. With this data I took a bootstrap sample and repeated it 50 times. The bootstrap distribution was 0.4113 and the standard error was 0.1058. With this information I found the 95% confident interval. The lower= 0.1998 and the upper= 0.6228.



Looking at the information, I conclude, and I fail to reject my hypothesis. My hypothesis stated that it equals .33. This value fell within the 95% confident interval.

In part five of this project, I have recreated my hypothesis and reran my test. The data that I got from this test had an original mean of 0.4. The bootstrap mean was 0.4102, had a standard error of 0.0853. The 95% confidence interval had a low of 0.2401 and a high of 0.5813. It also had the histogram below. The histogram had normal distribution with a bell-shaped curve.

On part six of my project I have used equations to test my hypothesis instead of a bootstrap distribution. In this equation I found the standard error, z-score, p, phat, p-value, and computed an 95% confidence interval. My sample size(n) was thirty players. Out of these thirty players I tested to see if 1/3 of their players were forwards. My proportion(p) therefore would be .33. To find my p hat value I took the twelve players that were forwards and divided it by the total amount of players. The p hat value was 0.4. With this information set up, I then created my test hypothesis. Ho: $p=.33$ and Ha: $p\neq.33$.

Next, I am going to find the standard error, z score, p-value, and 95CI. To find the standard error, you take the SQRT(p*(1-p)/n). The standard error computed was 0.085848704. Then you find the z score, which is (p hat-p)/SE. The z score came out to 0.815387963. Lastly, you take the z score and find the p value. Through excel I found that the p-value was 0.207425. Lastly, I computed a 95% confidence interval. This was done by following this equation. P+/-(2*SE). The lower is .33-(2*0.085848704) = 0.158302592. The upper is .33+(2*0.085848704) = 0.51697408. With this information, I still fail to reject my hypothesis. The proportion still fell within the value that were computed with the 95% confidence interval. Making the guess that forwards make up .33 would be a good guess for an overall population.

Sample Size n	p	p hat					
30	0.33	0.4					
H_O: p=0.33							
H_A:p≠0.33						0.1583	(LOWER)
	SQRT(p*(1-p)/n)	(p hat-p)/SE				0.5017	(UPPER)
	SE	Z		p-value			
	0.085848704	0.815387963		0.792574864			
				0.207425136	(1-p)		

On part seven of my project I have used equations to test my hypothesis instead of a bootstrap distribution. In this equation I found the standard error, x bar- mu, z-score, p-value, critical z score, and computed an 95% confidence interval. My sample size(n) was thirty players. Out of these thirty players I tested to see if it was practical to guess that if the average points per game of the players was 6. My mean(u) therefore would be 6.0. With this information set up, I then created my test hypothesis. Ho: u=6.0 and Ha: u≠6.0.

Next, I am going to find the standard error, z score, p-value, and 95CI. To find the standard error, you take the SD/SQRT (n). The standard error computed was 1.123604115. Then find x bar- mu. It computed to be 0.03. Then you find the z score, which is (x bar- mu)/SE. The z score came out to 0.026699795. Next, you take the z score and find the p value. Through excel I found that the p-value was 0.489349588. Then, take the NORM.S.INV(0.995) to find the z critical score(Z*), which was 2.575829. Lastly, I computed a 95% confidence interval. This was done by following this equation by taking P+/-((Z*) *(SE)). The lower is 6.0- (2.575829 * 1.1236 04115) = 3.105787595. The upper is 6.0- (2.575829*1.123604115) = 8.894212405. With this information, I still fail to reject my hypothesis. The mean still fell within the value that were computed with the 95% confidence interval. Making the guess that forwards make up 6.0 would be a good guess for an overall population.

old	SD	n	new	
5.97	6.15423	30	6	
SE	x bar-mu	z	p	Z*
1.123604115	0.03	0.026699795	0.510650412	2.57583
			0.489349588	
	CI			
	lower	3.105787595		
	upper	8.894212405		

For the last and final part of this project we are going to look at conditional probabilities for the two-way table in part two. For the two-way table, I added both OU and OSU into one table (shown below). The table looks the at position of the players and the heights of players at that position.

OU and OSU	Forward(F)	Guard(G)	Center(C)	Total	Probability
5'8"-5'9" (1)	0	1	0	1	1/30 or 0.0333
5'10"-5'11" (2)	0	3	0	3	3/30 or0.1000
6'0"-6'1" (3)	0	0	0	0	0/30 or 0.0000
6'2"-6'3" (4)	0	2	0	2	2/30 or 0.0667
6'4"-6'5" (5)	0	6	0	6	6/30 or 0.2000
6'6"-6'7" (6)	5	4	0	9	9/30 or 0.3000
6'8"-6'9" (7)	4	0	0	4	4/30 or 0.1333
6'10"-6'11" (8)	3	0	2	5	5/30 or 0.1667
Total	12	16	2	**30**	
Probability	12/30 or 0.4	16/30 or 0.5333	2/30 or 0.0667		

A player is randomly drawn. What is the probability that they are?

1. a guard = 16/30
2. a center = 2/30
3. a forward = 12/30
4. 6’4”-6’5” = 6/30
5. 6’10”-6’11” = 5/30
6. 6’6” or taller = 18/30
7. a guard, given that they are 6’6”-6’7” = 4/9
8. a forward, given that they are 6’8”-6’9” = 4/4
9. are 6’4”-6’5”, given they are a guard = 6/16
10. are 6’10”-6’11”, given they are a forward = 3/12

I found 7-10 by taking the total amount of the given, so in number 7 the total amount was 9 6’6”-6’7” players. Then I found the total number of guards that were that height, which was 4. So, I took the 4 found and divided it by the 9 total players that were that tall. So the probability that a guard is chosen, given they are 6’6”-6’7” tall is 4/9.