[redacted]

FIFA Career Mode English Premier League Players

Introduction to Probability and Stats

East Central University

03/11/2023

PROJECT PT 1 – REPORT

Below are 20/264 English Premier League players from England. For each player, their club, position, potential, rating, value, and a few other descriptions that would be valuable to someone who may want to select these players on FIFA in career mode. I find this data set interesting to me because it shows how some of my favorite players are ranked and helps me decide which one, I would like to do career mode with that would help me win and invest. This data is based on facts and not my own opinion that I formed watching real life games. I found this data on *Sports Statistics Sports Data*, where there are other sports and different data sets for each sport.

This data set has both quantitative and categorical variables. Some of the quantitative, ratio, and ordinal variables are the overall rating, potential, value, height, and weight. Some of the categorical variables are the different clubs and positions which is nominal. The FIFA ID and nationality ID are used when identifying players on FIFA. Height and weight are used to compare different players along with their rating and value.

REFERENCE

"FIFA 2022 Dataset CSVS (19k+ Players, 100+ Attributes)." *Sports Statistics Sports Data SportsStatisticscom*, https://sports-statistics.com/soccer/fifa-2022-dataset-csvs/.

FIFA ID	long_name	player_positions	overall	potential	value_eur	age	height_cm	weight_kg	club_team_id	club_name	league_name	club_position
220407	Martin Dúbravka	GK	81	81	13000000	32	190	80	13	Newcastle United	English Premier League	SUB
221363	Donny van de Beek	CM, CAM, CDM	81	86	38500000	24	184	74	11	Manchester United	English Premier League	SUB
221479	Dominic Calvert-Lewin	ST	81	85	37500000	24	187	71	7	Everton	English Premier League	ST
224081	Kalvin Phillips	CDM, CM, CB	81	85	32500000	25	178	72	8	Leeds United	English Premier League	CDM
226110	Nicolas Pépé	RM, RW	81	82	30500000	26	183	73	1	Arsenal	English Premier League	RM
226226	Giovani Lo Celso	CM, CAM	81	85	35500000	25	177	68	18	Tottenham Hotspur	English Premier League	SUB
234742	Harvey Lewis Barnes	LM, LW	81	84	35000000	23	180	66	95	Leicester City	English Premier League	LM
238074	Reece James	RWB, RB	81	86	37000000	21	178	87	5	Chelsea	English Premier League	RCB
239231	Marc Cucurella Saseta	LM, LB	81	87	41500000	22	172	66	1808	Brighton & Hove Albion	English Premier League	SUB
245279	Sergio Reguilón Rodríguez	LB	81	84	32500000	24	178	68	18	Tottenham Hotspur	English Premier League	LB
162347	João Filipe Iria Santos Moutinho	CM	80	80	9500000	34	170	61	110	Wolverhampton Wanderers	English Premier League	LCM
169588	Jonathan Grant Evans	CB	80	80	10000000	33	188	77	95	Leicester City	English Premier League	SUB
183855	Angelo Obinze Ogbonna	CB	80	80	10000000	33	191	86	19	West Ham United	English Premier League	LCB
189690	Vicente Guaita Panadero	GK	80	80	3600000	34	190	80	1799	Crystal Palace	English Premier League	GK
207599	Michael Keane	CB	80	81	20000000	28	191	82	7	Everton	English Premier League	LCB
210455	Jonathan Castro Otto	LWB, RWB	80	81	22000000	27	170	70	110	Wolverhampton Wanderers	English Premier League	SUB
211117	Bamidele Alli	CAM, CM	80	82	26500000	25	188	80	18	Tottenham Hotspur	English Premier League	LCM
213661	Andreas Bødtker Christensen	CB	80	85	28000000	25	187	81	5	Chelsea	English Premier League	CB
220971	Naby Keïta	CM	80	83	27500000	26	172	64	9	Liverpool	English Premier League	SUB

PROJECT PT 2

Below is a frequency table of the positions left back, right back, center mid, center back, striker, and goalkeeper. These are the most common positions, and this table will show which positions are most popular among English players in the EPL. Players or managers could use this information if they are wanting to see which players are most likely to make it to the EPL based on their positions. According to the data, the top 3 positions for English players in the EPL is center back, striker, and goalkeeper.

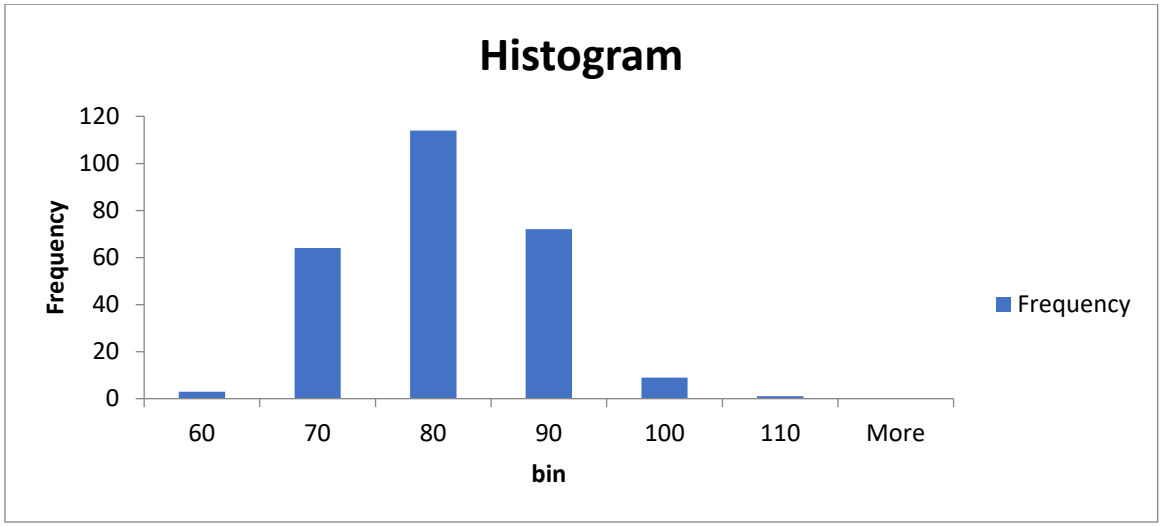
FREQUENCY	**FREQUENCY**	**RELATIVE FREQUENCY**
LB	7	7%
RB	8	8%
CM	6	6%
CB	42	41%
ST	18	17%
GK	22	21%
TOTAL	**103**	**1**

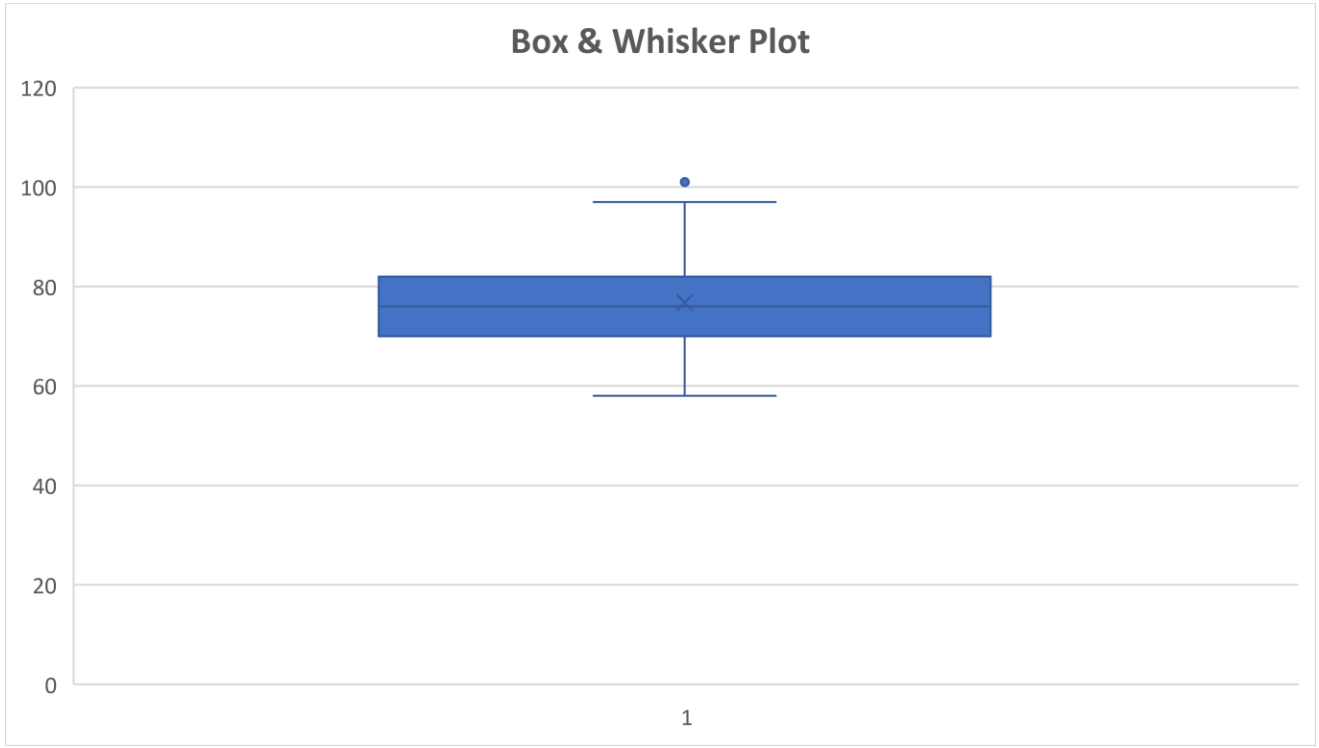
Below is a two-way table for the teams Newcastle United, Manchester United, Everton, Leeds United, Arsenal, and Tottenham Hotspur and the positions goalkeeper, striker, center back, center-mid, left back, and right back. I chose these teams because they have the most right and left backs while some teams do not have any. This table shows how many of each position a team has on their roster. Again, this could be used by players and coaches that wanted to see what the chances are of their players making in the EPL by position. Opponents could also use this to see how many they have in each position. The three positions that show up on a roster the most is center mid, center back, and goalkeeper. I think center-mid shows up the most because a starting lineup usually consists of 2-3 players plus subs. Same for the position center back, usually having 2-3 on a starting line, depending on formation.

two-way	GK	ST	CB	CM	LB	RB	TOTAL
Newcastle United	3	1	4	0	0	0	8
Manchester United	2	0	3	0	0	0	5
Everton	1	3	2	0	0	1	7
Leeds United	1	1	2	1	0	0	5
Arsenal	0	0	3	0	0	0	3
Tottenham Hotspur	0	0	1	0	1	1	3
TOTAL	7	5	15	1	1	2	31

PROJECT PT 3

Below is a Histogram and Box and Whisker Plot for the weight (Kilograms) of the players in the English Premier league. The mean or average weight of the players is 76.7338403 kg, and the standard deviation is 7.86755072. The Histogram shows that there are more players that weigh around 70-90 kg. The Histogram is almost symmetrical. The Box & Whisker Plot shows the minimum weight is 58 kg, the 1st Quartile is 70 kg, the median is 76 kg, the 3rd Quartile is 82 kg, and the maximum weight is 101 kg. There is one outlier which is the maximum weight.




Box & Whisker Plot
120
100
80
60
40
20
0
1

PROJECT PT 4

Quantitative Variable Hypothesis

H_0: μ = 27 years old

H_a : $\mu \neq 27$ years old

For my quantitative variable, my null hypothesis is that the average age of FIFA English players in the English Premier League is 27 years old. Around 27 years old is when players supposedly hit their "prime" in their soccer careers. Of course, there are the young stars and older legends, but I predict the average age to be 27. My alternative hypothesis is that the average age is not equal to 27. After calculating the mean, the average age of FIFA EPL players is 26.8 years of age. My hypothesis is close but is still incorrect.

Categorical Variable Hypothesis

H_0: p = .20

H_a: p ≠ .20

For my categorical variable, my null hypothesis is that 20% of players in the English Premier League FIFA career mode players play for the club Arsenal. I picked the team Arsenal because it is currently the best team in the EPL. My alternative hypothesis is that 20% of players do not play for the club Arsenal in FIFA career mode. After doing some math, only 6% of FIFA EPL players play for the club Arsenal. My hypothesis is way off. Only 16 players out of 264 play for Arsenal.

PROJECT PT 5

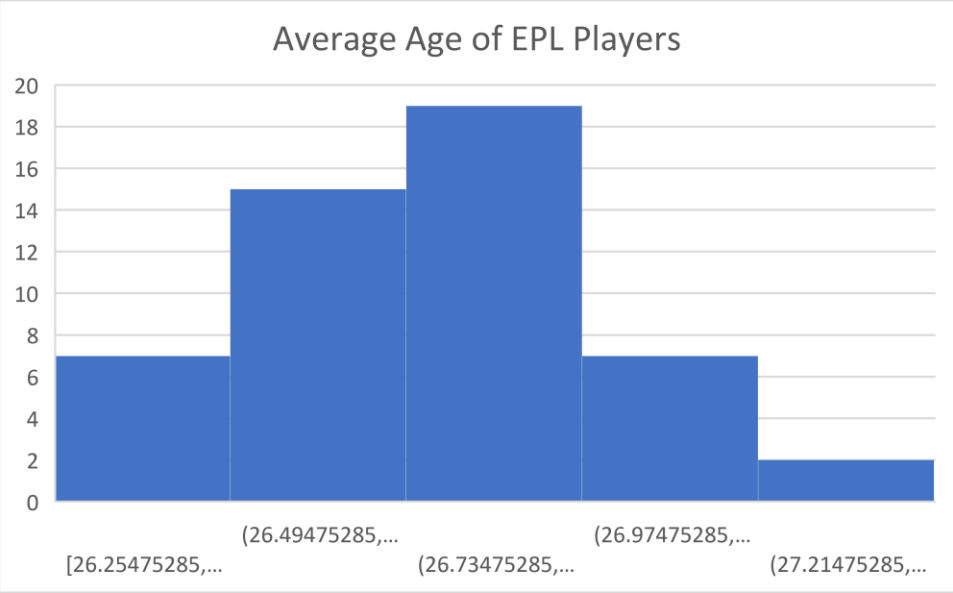
Quantitative Variable Bootstrap

Standard Error: *0.24368786*

95% Confidence Interval: *26.2715255 – 27.1485072*

Reject or Fail to Reject? *I fail to reject my hypothesis which was H_0: μ = 27 years old, the average age of 27 does fall in my Confidence Interval.*

Histogram:



Categorical Variable Bootstrap

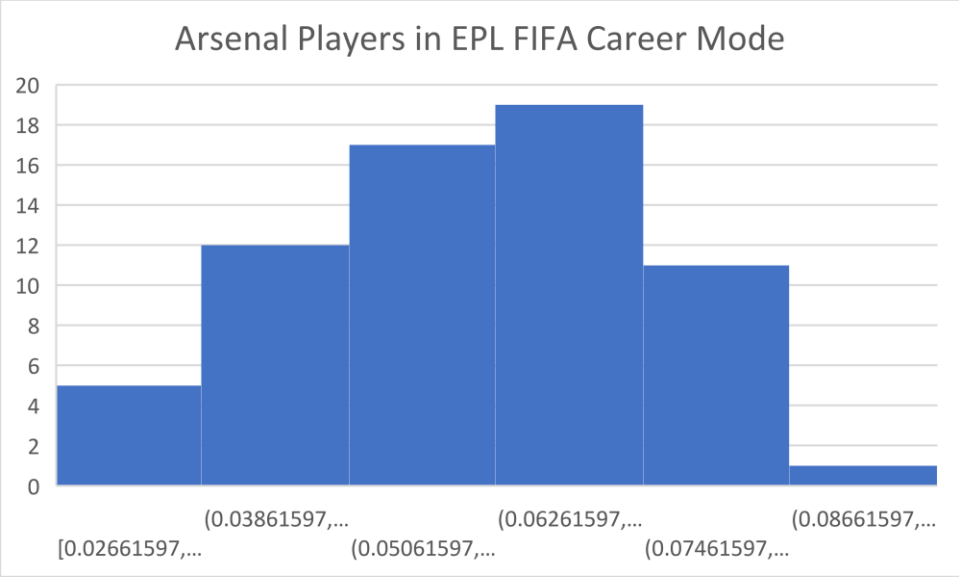
*****For the categorical variable I used the number "1" for number of players who play for Arsenal and "0" for those who play for another club

Standard Error: *0.01407407*

95% Confidence Interval: *0.0325415 – 0.08882324*

Reject or Fail to Reject? *Reject my hypothesis that 20% of EPL players in FIFA play for Arsenal. Less than 20% of EPL players play for Arsenal.*

Histogram:



PROJECT PT 6

Categorical Inference with Formulas

In part 6, I retested my categorical variable using the formulas I learned in Module 6. When I did a bootstrap for my categorical variable, I rejected my null hypothesis.

H_0: p = .20

H_a: p ≠ .20

After performing the second test, I still reject my hypothesis because it did not fall within the confidence intervals. 20% of EPL players in FIFA Career mode do not play for the club team Arsenal. As shown in the screenshot below my lower bound 95% confidence interval is 0.02032602 and my upper bound 95% confidence interval is .11701304.

*I went back and added how I computed my formulas & I redid them. The confidence intervals are quite a bit different than the bootstrap, not sure where I am going wrong.

	A	B	C	D	E	F	G	H
1	**sample**	**p**	**stat/phat**	**alpha**				
2	263	0.2	0.06866953	0.05				
3								
4	z*		se		z		95% CI	
5	1.96		0.02466506		-0.0004994		0.02032602	
6							0.11701304	
7			se		z			
8			sqrt(p*(1-p)/sample		(phat-p)/sample			
9							95% ci	
10							phat+or-1.96*SE	
11								
12								

PROJECT PT 7

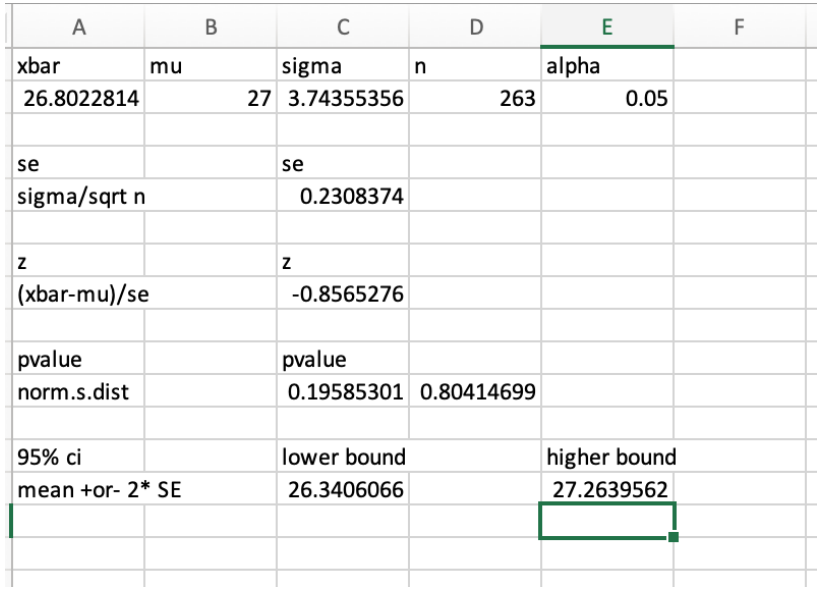
Quantitative Inference with Formulas

In part 7, I retested my quantitative variable using the formulas I learned in Module 7. When I did a bootstrap for my categorical variable, I failed to reject my null hypothesis.

H_0: μ = 27 years old

H_a : $\mu \neq 27$ years old

After performing this test with formulas, I still fail to reject my null hypothesis that the average age of FIFA EPL Career Mode players is 27 years old. As shown below, my 95% confidence intervals are [26.3406066, 27.2639562]. 27 falls within these intervals. I provided a screenshot below of how I did my formulas.



A	B	C	D	E	F
xbar	mu	sigma	n	alpha	
26.8022814	27	3.74355356	263	0.05	
se		se			
sigma/sqrt n		0.2308374			
z		z			
(xbar-mu)/se		-0.8565276			
pvalue		pvalue			
norm.s.dist		0.19585301	0.80414699		
95% ci		lower bound		higher bound	
mean +or- 2* SE		26.3406066		27.2639562	

PROJECT PT 8

CONDITIONAL PROBABILITY FROM A TWO-WAY TABLE

two-way	GK	ST	CB	CM	LB	RB	TOTAL
Newcastle United	3	1	4	0	0	0	8
Manchester United	2	0	3	0	0	0	5
Everton	1	3	2	0	0	1	7
Leeds United	1	1	2	1	0	0	5
Arsenal	0	0	3	0	0	0	3
Tottenham Hotspur	0	0	1	0	1	1	3
TOTAL	7	5	15	1	1	2	31

$$P(A|B)= P(A\cap B)/ P(B)$$

Using the formula above I will be computing 2 conditional probabilities from the two-way table I created in Module 2.

- P(Newcastle United | GK) = 3/7 or a 43% chance that Newcastle is chosen given that it came from GK or goalkeeper.
- P(CB | Newcastle United) = 4/8 or a 50% chance that CB or center back is chosen given that it come from Newcastle United.