IB Maths Essay: What is the correlation between the height of football players and their save percentages?

(Name)

(University)

Table of Contents

Raw Data for Analysis

This is the raw data for analysis:

Table 1: Raw Data

2015 (FOXSports 2015)		**2013** (FOXSports 2013)		**Average**	
Save %	**Height (cm)**	**Save %**	**Height (cm)**	**Save %**	**Height (cm)**
0.295	188	0.417	183	0.356	185
0.261	188	0.333	185	0.297	187
0.333	188	0.377	193	0.355	191
0.294	183	0.431	193	0.363	188
0.322	173	0.367	185	0.344	179
0.254	170	0.351	191	0.302	180
0.183	184	0.294	183	0.238	183
0.198	191	0.277	185	0.237	188
0.262	191	0.436	188	0.349	189
0.308	180	0.308	183	0.308	182
0.364	180	0.405	191	0.385	185
0.333	183	0.469	193	0.401	188
0.177	193	0.341	191	0.259	192
0.279	185	0.378	188	0.329	187
0.250	185	0.341	183	0.296	184
0.200	196	0.302	196	0.251	196
0.213	193	0.228	185	0.220	189
0.174	196	0.207	185	0.190	190
0.174	188	0.314	191	0.244	189
0.216	183	0.393	175	0.305	179
0.269	185	0.400	188	0.335	187
0.233	188	0.313	178	0.273	183
0.208	193	0.364	191	0.286	192
0.147	191	0.364	160	0.255	175
0.152	196	0.167	183	0.159	189

Rationale and Introduction

One of the most common sports is football within the United Kingdom. As such, one of the most well-known game series is the FIFA World Cup. However, it is noted that all players have different heights. Moreover, each of these players have different save percentages. These save percentages refer to the percentage of saves that a goalie has made in comparison to shots faced (FOXSports 2015; FOXSports 2013). The rationale for this study came from noting the height differences and save percentages differences, leading to the question: What is the correlation between football players' height and save percentage?

Statement of Task

The purpose of this study is to determine if there is a correlation and/or relationship between the height of football players and the save percentage. This was done for the years 2013 and 2015 for the FIFA World Cup. Therefore, through all instances, the x variable is height, whereas the y variable is save percentage.

Plan of Investigation

In order to determine the correlation and/or relationship between the height of football players and the save percentage, descriptive statistics (mean, median, and range) will be used to describe the data for both variables for 2013, 2015, and the average of the years. Following this, the standard deviation will be calculated for 2013, 2015, and the average of the years, for both variables. This will be followed by the calculation of the Pearson's Product-Moment Correlation Coefficient for both years and the average of the years.

Mathematical Investigation

The data comes from the FIFA World Cup website for 2013 and 2015. As the data is from the official FIFA site, it is expected that the data is reliable. The mathematical investigation includes descriptive statistics (mean, median, range, and standard deviation), and Pearson's Product-Moment Correlation Coefficient (height versus save percentage for both years).

Descriptive Statistics

The descriptive statistics used within this study are mean, median, range, and standard deviation. The mean is the average of the data. The formula for mean is:

Equation 1: Formula for Mean

In this formula, refers to the sum of the data points and *n* refers to the total number of data points. Based on the data in Appendix A, the mean for the 2013 height is 185.81 cm and the mean save percentage for the same year is 0.343. The mean for the 2015 height is 185.79 and the mean save percentage for the same year is 0.244.

The median is determined by first ordering the values from smallest to largest and then determining what number lies in the middle. If the data count is even, then the median two numbers are averaged. The median for 2013 height is 185 cm and the median 2015 save percentage is 0.351. The median for 2015 height is 188 cm and the median 2015 save percentage is 0.250.

The range is used to determine how wide the data variability is. This is done by ordering the numbers from smallest to largest. The range for 2013 height is 160 cm to 196

cm, whereas the range for 2013 save percentage is 0.167 to 0.469. On the other hand, the range for 2015 height is 170 cm to 196 cm, whereas the range for 2015 save percentage is 0.147 to 0.364. This is shown through the following two figures with the raw data:

Figure 1: Height Bar Graph

Figure 2: Save Percentage Bar Graph

The standard deviation shows the normal distribution of data, evident through the distance of the data points from the mean. For this study, the standard deviation by:

Equation 2: Formula for Standard Deviation

In this formula, N is the number of data points, is the individual data point, and is the mean of the data range. For this study, the standard deviation is calculated for both years, as well as for both variables. This is shown in the following table:

Table 2: Standard Deviation

	2015		**2013**		**Average**	
	Save %	**Height**	**Save %**	**Height**	**Save %**	**Height**
$N-1$	24	24	24	24	24	24
	0.042	0.042	0.042	0.042	0.042	0.042
	0.244	187	0.343	186	0.293	187
	0.051	1.160	0.074	-2.880	0.063	-1.280
	0.017	1.160	-0.010	-0.880	0.004	0.720
	0.089	1.160	0.034	7.120	0.062	4.720

	0.050	-3.840	0.088	7.120	0.070	1.720
	0.078	-13.840	0.024	-0.880	0.051	-7.280
	0.010	-16.840	0.008	5.120	0.009	-6.280
	-0.061	-2.840	-0.049	-2.880	-0.055	-3.280
	-0.046	4.160	-0.066	-0.880	-0.056	1.720
	0.018	4.160	0.093	2.120	0.056	2.720
	0.064	-6.840	-0.035	-2.880	0.015	-4.280
	0.120	-6.840	0.062	5.120	0.092	-1.280
	0.089	-3.840	0.126	7.120	0.108	1.720
	-0.067	6.160	-0.002	5.120	-0.034	5.720
	0.035	-1.840	0.035	2.120	0.036	0.720
	0.006	-1.840	-0.002	-2.880	0.003	-2.280
	-0.044	9.160	-0.041	10.120	-0.042	9.720
	-0.031	6.160	-0.115	-0.880	-0.073	2.720
	-0.070	9.160	-0.136	-0.880	-0.103	3.720
	-0.070	1.160	-0.029	5.120	-0.049	2.720
	-0.028	-3.840	0.050	-10.880	0.012	-7.280
	0.025	-1.840	0.057	2.120	0.042	0.720
	-0.011	1.160	-0.030	-7.880	-0.020	-3.280
	-0.036	6.160	0.021	5.120	-0.007	5.720
	-0.097	4.160	0.021	-25.880	-0.038	-11.280
	-0.092	9.160	-0.176	-2.880	-0.134	2.720
	0.003	1.346	0.005	8.294	0.004	1.638
	0.000	1.346	0.000	0.774	0.000	0.518
	0.008	1.346	0.001	50.694	0.004	22.278
	0.003	14.746	0.008	50.694	0.005	2.958
	0.006	191.546	0.001	0.774	0.003	52.998
	0.000	283.586	0.000	26.214	0.000	39.438
	0.004	8.066	0.002	8.294	0.003	10.758
	0.002	17.306	0.004	0.774	0.003	2.958
	0.000	17.306	0.009	4.494	0.003	7.398
	0.004	46.786	0.001	8.294	0.000	18.318
	0.014	46.786	0.004	26.214	0.008	1.638
	0.008	14.746	0.016	50.694	0.012	2.958
	0.004	37.946	0.000	26.214	0.001	32.718
	0.001	3.386	0.001	4.494	0.001	0.518
	0.000	3.386	0.000	8.294	0.000	5.198
	0.002	83.906	0.002	102.414	0.002	94.478
	0.001	37.946	0.013	0.774	0.005	7.398
	0.005	83.906	0.019	0.774	0.011	13.838
	0.005	1.346	0.001	26.214	0.002	7.398
	0.001	14.746	0.002	118.374	0.000	52.998
	0.001	3.386	0.003	4.494	0.002	0.518
	0.000	1.346	0.001	62.094	0.000	10.758
	0.001	37.946	0.000	26.214	0.000	32.718

	0.009	17.306	0.000	669.774	0.001	127.238
	0.008	83.906	0.031	8.294	0.018	7.398
∑	0.091	1055.360	0.125	1294.640	0.089	557.040
	0.004	43.973	0.005	53.943	0.004	23.210
	0.062	6.631	0.072	7.345	0.061	4.818

The changes in mean based on the standard deviation are shown in the following graphs for height and save percentages:

Figure 3: Mean and Standard Deviation for Height Graph

Figure 4: Mean and Standard Deviation for Save Percentage Graph

Pearson's Product-Moment Correlation Coefficient

Pearson's Product-Moment Correlation Coefficient is used to determine the relationship between two variables. In this context, each data point is plotted on the line of best fit in order to see the disbursement from this line. An upward slope indicates a positive correlation. A downward slope indicates a negative correlation. Correlations are measured as low (±0.1 to ±0.3), moderate (±0.3 to ±0.5), or high (±0.5 to ±1.0). To begin with, it is necessary to prepare the data for analysis. As noted previously, the *x* variable is height, whereas the *y* variable is save percentage. The following table is used to prepare the data for all three tests based on the raw data:

Table 3: Pearson's Product-Moment Correlation Coefficient

	$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$
2013	4647	8.577	1596.3	865079	3.068
2015	4671	6.099	1133.512	873735	1.579
Average	4657	7.337	1365.618	868063	2.246

The formula for Pearson's Product-Moment Correlation Coefficient involves a primary equation for *r* and three secondary equations, as shown below:

Equation 3: Formulas for Pearson's Product-Moment Correlation Coefficient

These can be condensed to result in:

Equation 4: Condensed Formula for Pearson's Product-Moment Correlation Coefficient

The formulas for all three tests are calculated below:

For the first test, there is a negative high correlation. However, the second and third tests show a low positive and negative correlation, respectively. However, in 2013, it is suggested that there was a low positive correlation between height and save percentage.

Discussion and Conclusion

The mean for the 2013 height is 185.81 cm and the mean save percentage for the same year is 0.343. The mean for the 2015 height is 185.79 and the mean save percentage for the same year is 0.244. This information shows that with a higher average height, there was an average higher save percentage and vice versa. The median for 2013 height is 185 cm and the median 2015 save percentage is 0.351. The median for 2015 height is 188 cm and the

median 2015 save percentage is 0.250. Based on this information, it is suggested that height has no bearing on save percentage as the median save percentage is lower, even with the higher median height. The range for 2013 height is 160 cm to 196 cm, whereas the range for 2013 save percentage is 0.167 to 0.469. On the other hand, the range for 2015 height is 170 cm to 196 cm, whereas the range for 2015 save percentage is 0.147 to 0.364. The difference in height range for 2013 is 25, whereas the difference in save percentage range for 2013 is 0.217. On the other hand, the difference in 2015 height range is 36, whereas the difference in 2015 save percentage range is 0.302. Similar to the mean, it is suggested that the range in heights may have an impact on the save percentage. For 2013 height, the standard deviation is 7.25, whereas the standard deviation for 2013 save percentage is 0.072. For 2015 height, the standard deviation is 6.50, whereas the standard deviation for 2015 save percentage is 0.062. This information suggests that there is a lower difference for 2015 height, as well as a lower difference in save percentages.

Based on the results of Pearson's Product-Moment Correlation Coefficient it is suggested that in 2015, there was a high correlation between height and save percentage. Moreover, this correlation was negative. This means that one variable is increasing as the other decreases. It can be theorized that save percentage increases as height decreases. However, in 2013, it is suggested that there was a low positive correlation between height and save percentage. This suggests that both variables are increasing, suggesting that there is a slight relationship between the variables. When analysing the average, there is a low negative correlation, again suggesting that one variable is increasing while the other decreases. Based on the results of Pearson's Product-Moment Correlation Coefficient, it is

evident that there is an overall negative relationship between height and save percentages of football players. This is confirmed through the trend lines shown in Figures 3 and 4.

References

FOXSports, 2013. 2013 Goalkeeping Statistics. *World Cup Statistics*. Available at: http://www.foxsports.com/soccer/stats?competition=31&season=20130&category=GOALKEEPING&pos=Goalie&team=0&isOpp=0&splitType=0&sort=9&sortOrder=0.

FOXSports, 2015. 2015 Goalkeeping Statistics. *World Cup Statistics*. Available at: http://www.foxsports.com/soccer/stats?competition=31&season=20150&category=GOALKEEPING&pos=Goalie&team=0&isOpp=0&splitType=0&sort=9&sortOrder=0.