**Statistical Analysis of FIFA Game Players**

**Contents**

1. Abstract
2. Introduction
3. Details of the Work
4. Conclusion

**Abstract**

Insights and correlations between player value, age, special attributes (which include sprint speed and stamina rating) , and performance can be derived from the dataset. This uninterpreted data can be converted into information by analysing it. We have derived summary statistics for teams, clubs, & players. Through extensive football experience: the insights provided in our results, alongwith understanding, and contextualized information enables users to act smartly when playing FIFA, picking a better team for say Fantasy Premier league, or increase their betting odd.

**Details of the Work**

1. The data on which this project is based is a sample, as we have taken randomly 250 players out of all the football players in the world.
2. The data we used is covered over 5 years from where we get stamina rating of player , sprint speed of the player. These ratings change every year based on the previous year ratings.

1. The Qualitative and the Quantitative data which we have taken are as follows shown in the Table 1 .

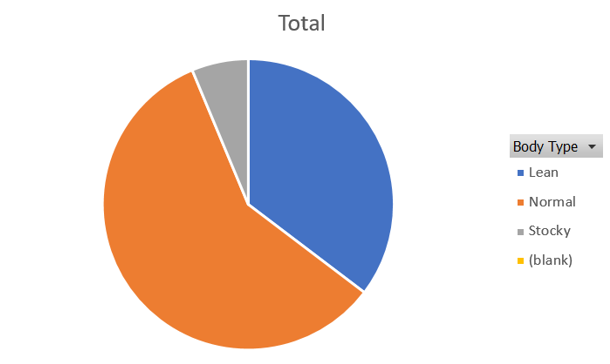
**Table 1. Variables Selected for the Analysis**

|  |  |  |
| --- | --- | --- |
| **Variable Name in the Data Set** | **Description**  (See the data dictionary for describing the variables.) | **Type of Variable**  (Qualitative or Quantitative) |
| Variable 1: “Nationality” | Gives the detail of the **Nationality** of the football player respectively. | Qualitative |
| Variable 2: “Preferred Foot ” | Gives the detail of the **Foot** used by the football player respectively. | Qualitative |
| Variable 3:  “ Preferred position of the player” | Gives the detail of the **position** of the player on which he plays . | Qualitative |
| Variable 4:  “Body Type ” | Gives the details of the **Body Type** of the football player. | Qualitative |
| Variable 5:  “Age” | **Age** of the player. | Quantitative |
| Variable 6: “Potential/Overall Rating” | Give the detail of the **Potential/Overall rating** of the player . | Quantitative |
| Variable 7: “Height” | **Height** of the player. | Quantitative |
| Variable 8: “Weight” | **Weight** of the player. | Quantitative |
| Variable 9: “Sprint Speed” | **Sprint Speed** of the player. | Quantitative |
| Variable 10: “Stamina” | **Stamina (Rating)** of the player. | Quantitative |

**Data Set Description**

1. Variable :- **Body Type (Qualitative )**

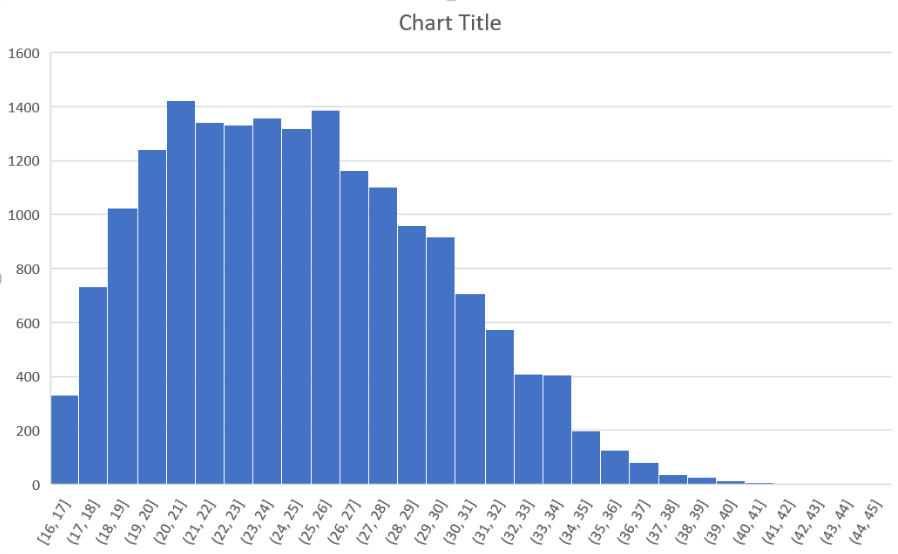
Suitable Graph for this variable is **pie chart** which is shown below :-



**Description of Findings :** From this pie chart we get to know most of the players have a normal body (58.3%) , percentage players with lean body type is in between the players who have stocky and normal body type respectively.Lean body type players are the stocky with 6.2%.

1. Variable:- **Age(Quantitative)**

Suitable Graph for this variable is **Histogram** which is shown below :-



**Description of Findings :-** From this histogram we infer that most football players are in the age category 20-21(years of age), most least is the range 40-41(years in age).Taking a rough look at the graph (Histogram) we get to know that most players have age in between 18-30(age in years).

6) (a) Variable 1**: “Sprinting Speed”**

Numerical Summary:-

**Table 2. Descriptive Analysis for Variable 1(Sprinting Speed)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **n** | Measure(s) of Central Tendency | Measure(s) of Dispersion |
| Variable: | 18159 | Mean = 64.727 | S.D = 14.649 |

**Description of Findings :-** So we get to know that the central tendency for “Sprinting Speed” i.e. mean = 64.727 which means that most players have Sprinting speed rating close 64 out of 100 , and by the Measure(s) of Dispersion i.e. S.D = 14.649 through which we get know that the players have their sprinting speed rating 64.727±14.649.

(b) Variable 2**: “Stamina”**

Numerical Summary:-

**Table 2. Descriptive Analysis for Variable 1(Sprinting Speed)**

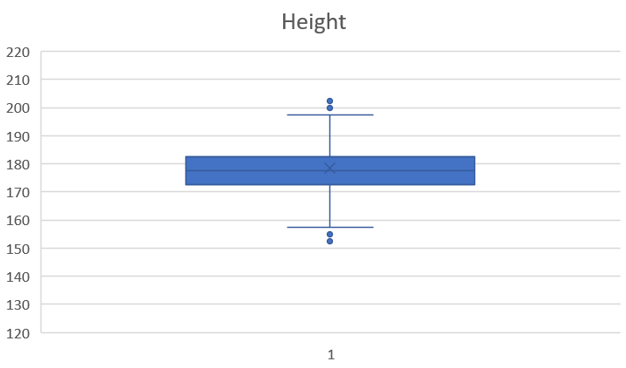
|  |  |  |  |
| --- | --- | --- | --- |
| Variable | n | Measure(s) of Central Tendency | Measure(s) of Dispersion |
| Variable: | 18159 | Mean = 63.212 | S.D = 15.894 |

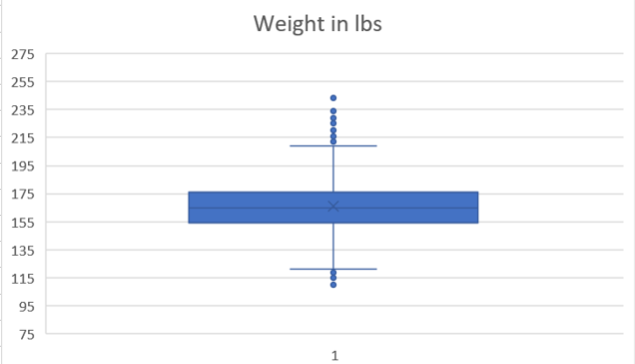
**Description of Findings:-** So we get to know that the central tendency for “Stamina” i.e. mean = 63.212 which means that most players have stamina rating close 63 out of 100 , and by the Measure(s) of Dispersion i.e. S.D = 14.649 through which we get know that the players have their stamina rating 63.212±15.894.

7.) Variables (**Quantitative**) used for making the **box plot and whisker plot** are :-

**A) Height**

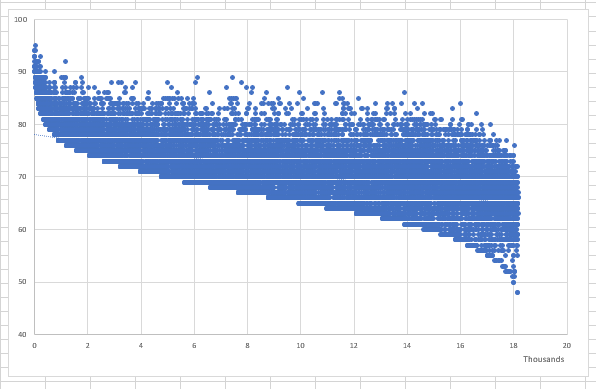
1. **Weight**

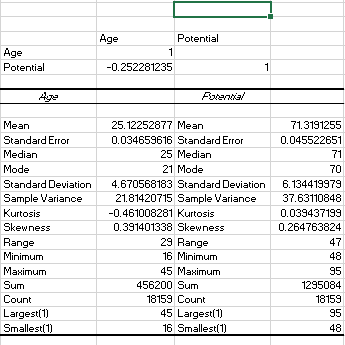




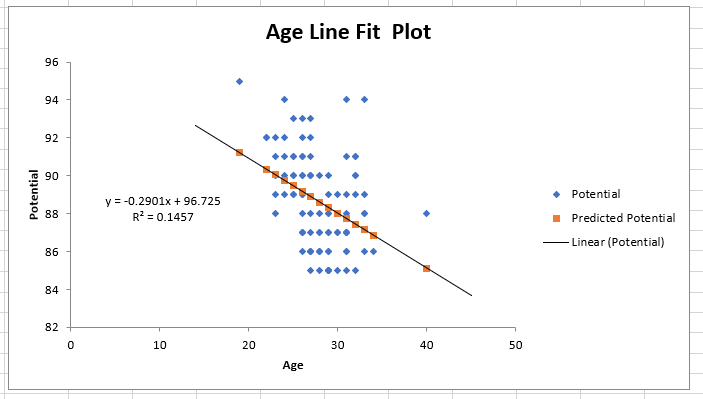
**Description of Findings:-** From the above whisker box plot we infer that number of outliers in case of weight of the football players are way more than the number of outliers in case of the height of the players.

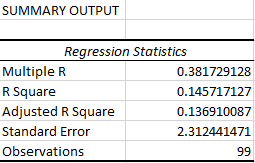


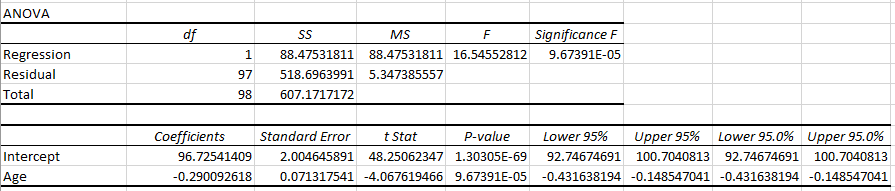




Age and Potential have weak and negative potential that means with increase in age potential of a player decreases which is very obvious.





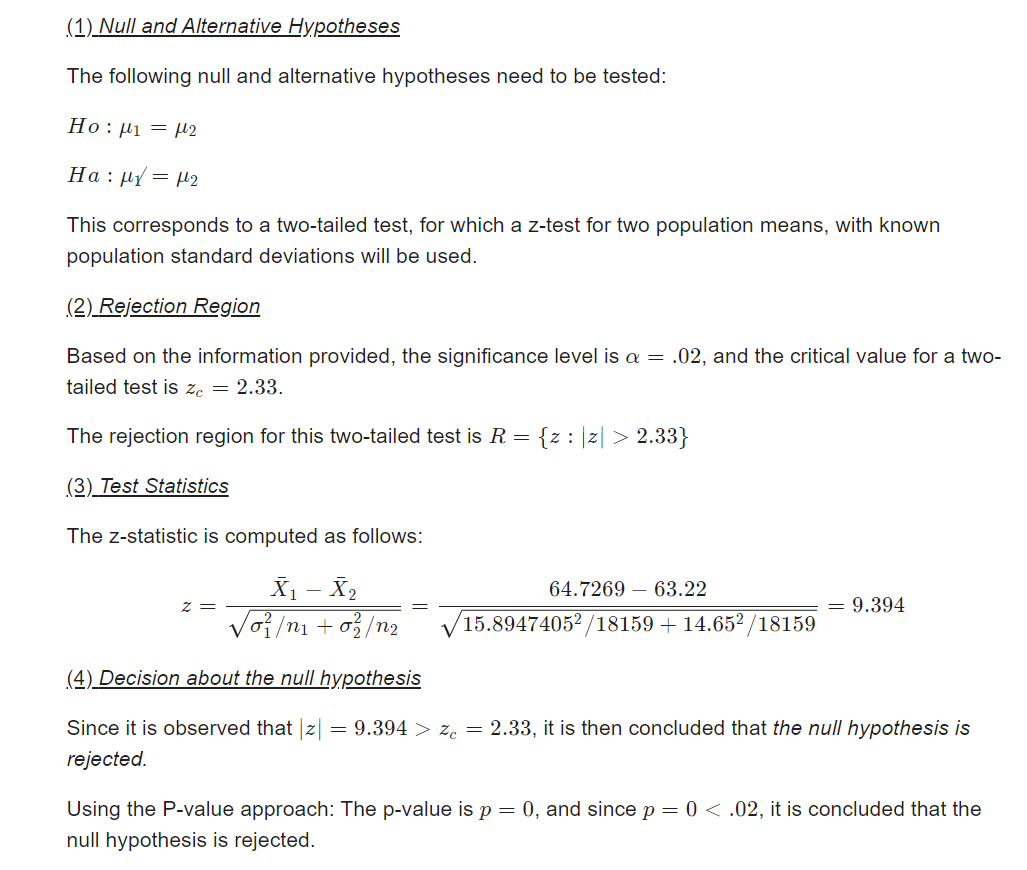


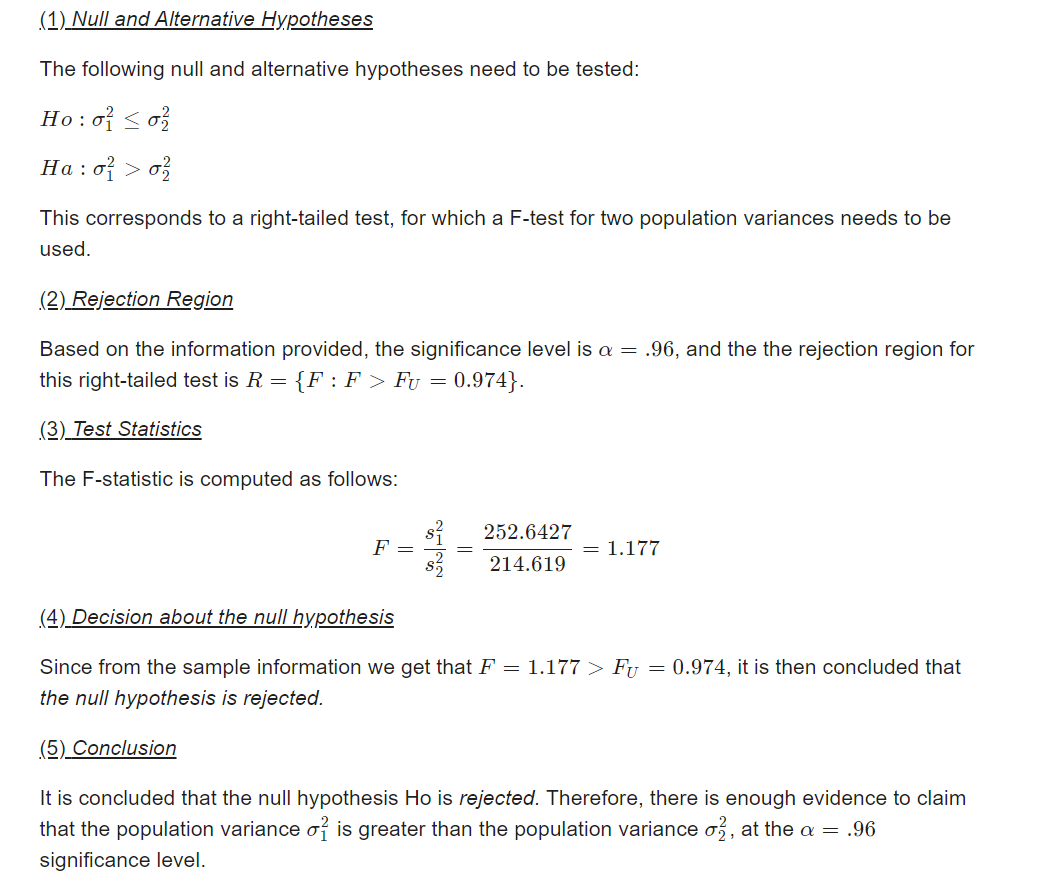
9)

We choose two Quantitative variables **Stamina and Sprint** speed .Mean and standard deviation of first variable (Stamina) is **63.22 and 15.894 .**

Mean and standard deviation of second variable (Sprint Speed ) is **64.7269 and 14.65**. The difference between their means is 1.507. Two-sided confidence interval of confidence level of **97%** for the difference of their population averages is**(−1.855,−1.159)** while two-sided confidence interval for the ratios of their population variances for the confidence level of **99% is (1.133,1.223).**

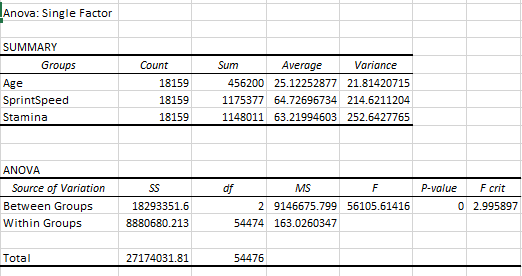
10)



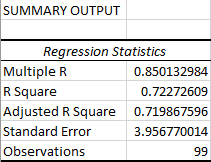


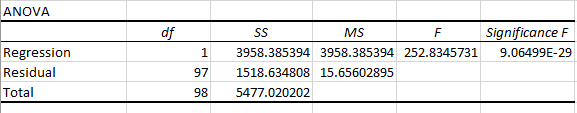
**The 4% confidence interval is (1.176,1.178).**

1. **We want to examine the potential of Fifa Players based on Age, Stamina, Sprint Speed.We collected sample for these factors and using data we will test the potential of the players (In the Fifa Game), using α=5%**

****



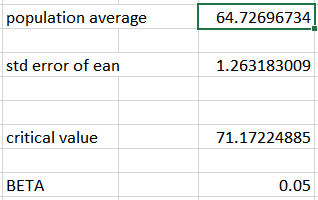
****



reg3

**Type 1 error** is the same as the 1-α which has the numerical value of 0.05 as the chosen α has the value of 0.95.

**Type 2 error :- Beta = 0.05 shown below .**

****

**Conclusion**

This project was divided into two part , which helped us to work on variables in two categories. In this project we were able to find a lot of things which include the relationship between the age of player and the potential of the player (using linear regression),then we also found out the confidence interval for the alpha as 95% which was a two sided interval . We also defined and tested the hypothesis , then calculated the type 1 and type 2 error on a quantitative variable and also defined anova problem and made a scattered plot for regression.