

FOOTBALL FANTASY PREMIER LEAGUE

a_group117

***OLUWATOBI ELIJAH AKANNI, . NIMRA, TAIWO FUNMILAYO ADEBAYO, MANEESH
SATHEESH, MUHAMMAD SAJAWAL KHAN***

Thursday, 13 January 2024

TABLE OF CONTENTS

ABSTRACT	3
1.0.INTRODUCTION	3
1.1.RESEARCH QUESTION	3
1.2.VARIABLES.....	3
1.3.HYPOTHESES.....	4
1.4.DATASET DETAILS	4
1.5.METHODOLOGY	4
2.0. VISUALISATION	4
2.1. BOXPLOT.....	4
2.2. HISTOGRAM	5
3.0. ANALYSIS	6
4.0. CONCLUSION	6
REFERENCES	6
APPENDIX	6

FIGURES

Figure 1: Fantasy Premier League: https://fantasy.premierleague.com/	3
Figure 2: Boxplot Of Football Fantasy Premier League	4
Figure 3: Histogram Of Football Fantasy Premier League	5
Figure 4: Football Fantasy Premier League Database	7
Figure 5: Football Fantasy Premier League Research Question.....	7
Figure 6: Football Fantasy Premier League Visualisation (Boxplot and Histogram).....	8
Figure 7: Football Fantasy Premier League Analysis(pairwise.t.test).....	8

ABSTRACT

This report presents a statistical analysis of the Fantasy Premier League data (O'BrienID, et al., 2021) to investigate the difference in the means of total points scored across seasons from 2016 to 2021. The study utilised a dataset obtained from Kaggle (ID: K088) and performed statistical analysis using R (Austria, 2020). The report presents the research question, independent and dependent variables, null and alternative hypotheses, and the dataset details. Furthermore, it includes the code output for loading the dataset into R and conducting statistical analysis.

1.0. INTRODUCTION

Fantasy Premier League, or FPL, is the most liked fantasy game in the whole world. It is the official fantasy football game for England's top league or soccer system, run annually by the Premier League's website. The exciting side of Football games makes FPL very interesting, and the chance in football gives people a special thrill like that of an adventure activity. (Pratik, et al., 2022)

Fantasy Premier League (*see Figure 1*) is a fun online game where you pick real-life football players to be on your team. You get points if they play well in actual games, too! This report will examine the total points made by players during different seasons. It wants to find out if there is a significant change in their average scores.



Figure 1: Fantasy Premier League: <https://fantasy.premierleague.com/>

1.1 RESEARCH QUESTION

The research question for this study is: Is there a difference in the means of total points scored across seasons (ranging from 2016 to 2021)?

1.2 VARIABLES

Independent Variable: Season (season_x)

Dependent Variable: Total Points (total_points)

1.3 HYPOTHESES

Null Hypothesis (H0): There is no difference in means between total points and seasons (ranging from 2016 to 2021).

Alternative Hypothesis (H1): There is a difference in means between total points and seasons (ranging from 2016 to 2021).

1.4 DATASET DETAILS

Dataset ID: K088

Dataset URL: [Kaggle Fantasy Football

Dataset](<https://www.kaggle.com/datasets/joebeachcapital/fantasy-football>)

Dataset File: cleaned_merged_seasons.csv

(Arvidsson, 2023)

1.5 METHODOLOGY

The data was put into R, and we took the crucial parts for the study. To check the results, we made a boxplot and Histogram to look at how total points were spread across different seasons. The points average and spread were worked out. Then, a line graph for normal distribution was made. Then, a form of pairwise. t-test called the Bonferroni method, was used to compare the average total points between seasons.

However, the results of this study will inform us how many points players score in the Fantasy Premier League in different seasons. The study done using R programming language will help us know the differences in average values and how important they are (*see Figures 6 & 7*). This report is helpful for those who like fantasy football and people studying sports numbers.

2.0 VISUALISATION

Many of us have been in an unpleasant circumstance when only a data set's mean and standard deviation are presented. However, we are supposed to know everything about the dataset as if those two figures were all that was required. We would learn more if we could quickly build and share graphical representations and interpretations of the complete raw data set. (Rachel, et al., 2020)

The study results will be shown below (*see Figures 2 & 3*). It will include things like number facts, pictures, and tests of figures to check if there is a big difference between seasons for total points scored.

2.1 BOXPLOT

The boxplot shows how the total points scored are spread over different seasons. It lets you compare the average and changes in scores quickly (*see Figure 2*). The boxplot adds this by showing how often each total points scored happens.

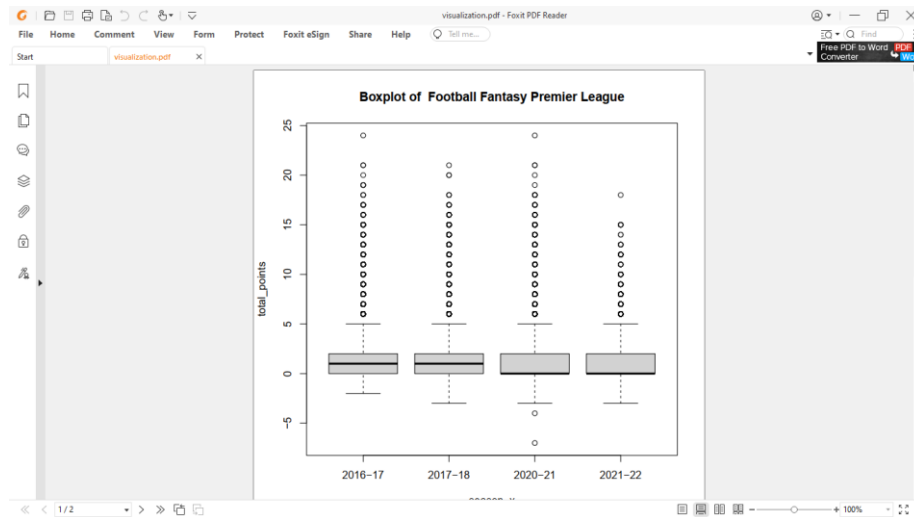


Figure 2: Boxplot Of Football Fantasy Premier League

2.2 HISTOGRAM

The average and standard deviation are significant numbers that help us understand the middle score and spread of all points earned. The normal distribution line shows how many points were scored in total and gives information on the shape of this distribution. (see figure 3)

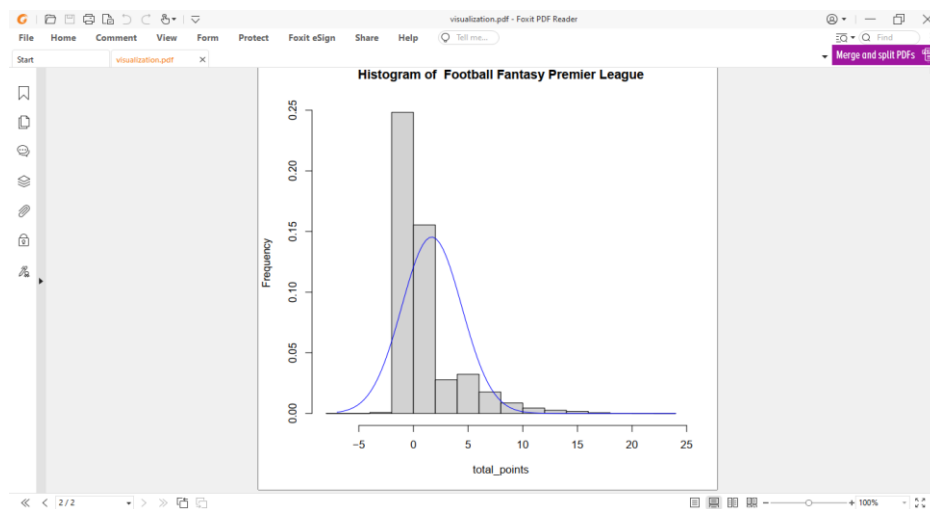


Figure 3: Histogram Of Football Fantasy Premier League

Overall, these visualisations and summary statistics will aid in understanding the performance trends of players in the Fant.

3.0. ANALYSIS

The results of the pairwise.t.test analysis comparing the means of total points scored across seasons are as follows:

Pairwise comparisons using t-tests with pooled SD

data: data\$total_points and data\$season_x

	2016	2017	2018	2019	2020
2017	0.066	-	-	-	-
2018	1.000	0.016	-	-	-
2019	0.049	1.000	0.008	-	-
2020	0.681	0.032	1.000	0.031	-
2021	0.011	1.000	0.002	1.000	0.011

P value adjustment method: Bonferroni

The p-values indicate the significance of the differences in means between the seasons, with Bonferroni correction for multiple comparisons.

CONCLUSION

The pairwise.t.test analysis with Bonferroni correction shows essential changes in total points scored during different seasons of Fantasy Premier League. This finding backs up the alternative hypothesis that there is a difference in means between total points and seasons (ranging from 2016 to 2021).

References

- Arvidsson, J., (2023). *Kaggle*. [Online]
Available at: <https://www.kaggle.com/datasets/joebeachcapital/fantasy-football>
- Austria, V., (2020). *R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing*. [Online]
Available at: <https://www.R-project.org/>(<https://www.R-project.org/>)
[Accessed 20 october 2023].
- Dr. Noll, J., (2023). *analysis.R*, University Of Hertfordshire: s.n.
- Dr. Noll, J., (2023). *HistogramPlusBoxplot*, University Of Hertfordshire: s.n.
- O'BrienID, J. D., Gleeson, J. P. & O'Sullivan, D. J. P., (2021). Identification of skill in an online game: The. *PLOS ONE* , Volume 3, p. 16.
- Pratik, P., Panday, A. T. S. & Acharya, B., (2022). Fantasy Premier League-Performance Prediction. *Department of Electronics and Computer Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*.
- Rachel, H., Chen, Z., Rashid, M. & Sarkar, J., (2020). Visualizing the mean and the standard deviation using R/RStudio Shiny package. *hufnagel2020visualizing*.

APPENDIX

AutoSave

cleaned_merged_seasons

Saved to this PC

Search

akanni oluwatobi

FileHomeInsertPage LayoutFormulasDataReviewViewAutomateHelp

Paste

Clipboard

Calibri

11

A⁺

B

I

U

Font

Alignment

Number

General

Format as Table

Cell Styles

Conditional Formatting

Format as Table

Cell Styles

Number

Insert

Delete

Format

Insert

Delete

Format

Cells

Sort & Find

Filter & Select

Editing

Sort & Find

Filter & Select

Editing

Filter & Select

Editing

Sensitivity

Add-ins

Analyze Data

Sensitivity

Add-ins

Analyze Data

Sensitivity

Add-ins

Analyze Data

Comments

Share

A1

season_x

name

position

team_x

assists

bonus

bps

clean_sheets

creativity

element

fixture

goals

goals_conceded

scout_index

influence

kickoff

minutes

opponent

own_goals

goal_penalties

2016-17

Aaron Creech

DEF

NA

0

0

0

0

0

454

10

0

0

0

0

0

4

Chelsea

0

2016-17

Aaron Lennon

MID

NA

0

0

6

0

0.3

142

3

0

0

0.9

8.2

2016-08-11

15

17

Spurs

2016-17

Aaron Ramsdale

MID

NA

0

0

5

0

4.9

16

8

3

 0 | 3 | 2.2 | 2016-08-11 | 60 | 9 | Liverpool |

2016-17

Abdoulaye Doucoure

MID

NA

0

0

0

0

0

482

7

0

 0 | 0 | 0 | 2016-08-11 | 0 | 13 | Southampton |

2016-17

Adam Forster

MID

NA

0

0

 0 | 3 | 0 | 1.3 | 286 | 6 | 1 | 0 | 0.3 | 2 | 2016-08-11 | 69 | 14 | Stoke |

2016-17

Adam Lallana

MID

NA

1 2 | 33 | 0 | 33.7 | 205 | 8 | 3 | 1 | 14.2 | 51.2 | 2016-08-11 | 75 | 1 | Arsenal |

2016-17

Adam Smith

DEF

NA

 0 | 0 | 23 | 0 | 4.3 | 34 | 9 | 3 | 1 | 7.4 | 42.4 | 2016-08-11 | 90 | 11 | Man Utd |

2016-17

Adrian S

FW

NA

 0 | 0 | 16 | 0 | 0 | 450 | 10 | 2 | 0 | 3 | 29.8 | 2016-08-11 | 90 | 4 | Chelsea |

2016-17

Alex Iwobi

MID

NA

 0 | 0 | 12 | 0 | 17.5 | 21 | 8 | 3 | 0 | 3.4 | 16.6 | 2016-08-11 | 38 | 9 | Liverpool |

2016-17

Alex McCosker

GW

NA

 0 | 0 | 0 | 0 | 0 | 101 | 7 | 0 | 0 | 0 | 0 | 0 | 2016-08-11 | 0 | 18 | Watford |

2016-17

Alex Oxlade-Chamberlain

MID

NA

 0 | 0 | 23 | 0 | 6.5 | 18 | 8 | 1 | 1 | 6.5 | 39.4 | 2016-08-11 | 32 | 9 | Liverpool |

2016-17

Andreas Pereira

MID

NA

 0 | 0 | 0 | 0 | 0 | 263 | 9 | 0 | 0 | 0 | 0 | 0 | 2016-08-11 | 0 | 2 | Bournemouth |

2016-17

Andrew Robertson

DEF

NA

 0 | 0 | 14 | 0 | 1.8 | 152 | 4 | 1 | 0 | 1.7 | 14.8 | 2016-08-11 | 90 | 8 | Leicester |

2016-17

Andre Gray

FW

NA

 0 | 0 | -3 | 0 | 2.3 | 68 | 1 | 1 | 0 | 4.3 | 0 | 2016-08-11 | 90 | 16 | Swansea |

2016-17

Andros Townsend

MID

NA

 0 | 0 | 17 | 0 | 32.8 | 120 | 2 | 1 | 0 | 10.2 | 23.2 | 2016-08-11 | 90 | 19 | West Brom |

2016-17

Andre Schurrle

FW

NA

 0 | 0 | 5 | 0 | 1.2 | 374 | 10 | 0 | 0 | 0.5 | 0.2 | 2016-08-11 | 34 | 4 | Chelsea |

2016-17

Andy Carroll

FW

NA

 0 | 0 | 7 | 0 | 0 | 17 | 468 | 10 | 2 | 0 | 3.2 | 10.6 | 2016-08-11 | 90 | 4 | Chelsea |

2016-17

Angelo Ogbonna

DEF

NA

 0 | 0 | 0 | 0 | 0 | 456 | 10 | 0 | 0 | 0 | 0 | 0 | 2016-08-11 | 0 | 4 | Chelsea |

2016-17

Anthony Martial

FW

NA

 2 | 3 | 35 | 0 | 37.3 | 267 | 9 | 1 | 0 | 10.4 | 38.8 | 2016-08-11 | 84 | 2 | Bournemouth |

2016-17

Arthur Mboya

DEF

NA

 0 | 0 | 19 | 0 | 1.4 | 505 | 10 | 2 | 0 | 1.8 | 16.4 | 2016-08-11 | 90 | 4 | Chelsea |

2016-17

Ashley Barnes

FW

NA

 0 | 0 | 0 | 0 | 0 | 70 | 1 | 0 | 0 | 0 | 0 | 0 | 2016-08-11 | 0 | 16 | Swansea |

2016-17

Ashley Young

DEF

NA

 0 | 0 | 0 | 0 | 0 | 260 | 9 | 0 | 0 | 0 | 0 | 0 | 2016-08-11 | 0 | 2 | Bournemouth |

2016-17

Ramir Leal

MID

NA

 0 | 0 | 6 | 0 | 16.9 | 198 | 3 | 1 | 0 | 3 | 7.7 | 2016-08-11 | 90 | 6 | Fulham |

cleaned_merged_seasons

Ready

Accessibility

100%

Figure 4: Football Fantasy Premier League Database

```

# group name
# group: a_group 117
# group members name
# members: [Oluwatobi Elijah Akanni, - Wura, Taiwo Fumilayo Adedokun, Harshesh Satheesh, Muhamad Sa'adul Khan]
# general topic
# topic: Football Fantasy Premier League
# research question
# question:
# Q: Is there a difference in the means of total points scored across seasons (ranging from 2016 - 2021)?
# write your independent and dependent variables.
# independent-var: season_x
# dependent-var: total_points
# null hypothesis for your research question.
# null_hypothesis: there is no difference of means between total points and seasons (ranging from 2016 - 2021)
# alternative hypothesis for your research question.
# alt_hypothesis: there is difference of means between total points and seasons (ranging from 2016 - 2021)
# the dataset ID (name):
# dataset-id: 1028
# write the dataset URL.
# dataset-url: https://www.kaggle.com/datasets/josebachopital/fantasy-football
# the dataset CSV filename.
# dataset-filename: cleaned_merged_seasons.csv
# Paste the output of loading your dataset into R.
# columns:
# [1] "season_x" "name" "position" "team_x"
# [4] "assists" "bonus" "bps" "clean_sheets"
# [8] "creativity" "element" "fixture" "goals_conceded"
# [12] "goals_scored" "influence" "kickoff_time"
# [16] "minutes" "opponent_team" "own_goals"
# [20] "penalties_missed" "penalties_saved" "red_cards" "round"
# [24] "shots" "shots_on_target" "team_x_score" "team_y_score"

```

Figure 5: Football Fantasy Premier League Research Question

```

1 #The following set of code is inspired by "Dr. Noll J. (2023) Academic section: practical 4, Histogramplusboxplot"
2
3 # Load base R
4 library(base)
5
6 #setting the output to pdf
7 pdf("visualization.pdf")
8
9 #reading our dataset
10 mydata <- read.csv("cleaned_merged_seasons.csv")
11
12 #creating Boxplot graph for the dataset
13 boxplot(mydata$total_points ~ mydata$season_x, main = "Boxplot of Football Fantasy Premier League", xlab = "season_x", ylab = "total_points")
14
15 #creating a histogram for the dataset
16 hist(mydata$total_points, breaks = , main = "Histogram of Football Fantasy Premier League", xlab = "total_points", ylab = "Frequency", freq=FALSE)
17
18 #calculates the mean of the variable "total_points"
19 # in the dataset "mydata" and stores it in the variable "mean_total"
20 mean_total <- mean(mydata$total_points)
21
22 #the standard deviation of the variable "total_points"
23 #in the dataset "mydata" and stores it in the variable "sd_total".
24 sd_total <- sd(mydata$total_points)
25
26 #creating a sequence of 100 evenly spaced values between
27 #the minimum and maximum values of the variable "total_points"
28 # in the dataset "mydata" and stores it in the variable "x"
29 x <- seq(min(mydata$total_points), max(mydata$total_points), length = 100)
30
31 #calculating the normal distribution values for the sequence of "x"
32 # values, using the mean and standard deviation calculated earlier, and stores it in the variable "y".
33 y <- dnorm(x, mean = mean_total, sd = sd_total)
34
35 #plots a line graph using the "x" and "y" values, with a blue color.
36 lines(x, y, col = "blue")
37
38 dev.off()
39

```

Figure 6: Football Fantasy Premier League Visualisation (Boxplot and Histogram) (Noll, 2023)

```

1 #The following set of code is inspired by "Dr. Noll J. (2023) Academic section: analysis-assignment, analysis.R"
2
3 data <- read.csv("cleaned_merged_seasons.csv")
4
5 stdout <- pairwise.t.test(data$season_x, data$total_points, p.adjust.method = "bonferroni")
6
7 print(stdout)
8

```

Figure 7: Football Fantasy Premier League Analysis(pairwise.t.test) (Noll, 2023)