**STATISTICAL ANALYSIS OF**

**IPL PLAYERS PERFORMANCE**

Submitted by

Dhruv Dadheech-22030121070

Diya Jain-22030121076

Gautam Bhalerao-22030121081

Hardik Gandhi-22030121085

BCA Students, Sem-III,

Symbiosis Institute of Computer Studies & Research

Symbiosis International (Deemed) University

Model colony, Atur Centre,

PUNE -411016

2022-25 Batch

**ACKNOWLEDGEMENT** to

Director SICSR-Dr.Jatinderkumar R. Saini

Deputy Director-Dr. Parag Ravikant Kaveri

Admin officer-Dr. Deepali Prakash Suryavanshi

Course in-charge-Dr. Priti Kulkarni

Faculty in-charge-Dr. Shilpa Mujumdar

Data set downloaded - Kaggle Website

**INDEX**

|  |  |  |
| --- | --- | --- |
| SR. NO. | CONTENTS | PAGE NO. |
| 1 | Abstract | 1 |
| 2 | Introduction | 2 |
| 3 | Methodology | 3 |
| 4 | Data Analysis | 4 |
| 5 | Conclusions | 9 |
| 6 | References | 10 |
| 7 | Annexure | 11 |

**Abstract**

Statistical tools are used to analyse the data on the **“STATISTICAL ANALYSIS OF IPL PLAYERS PERFORMANCE”.**

We have used different Statistical tools like Arithmetic Mean (AM), Geometric Mean (GM), Harmonic Mean (HM), Standard Deviation (S.D), Coefficient of Variation, Coefficient of kurtosis, Correlation, Regression, First Quartile Q1, Second Quartile Q2, Third Quartile Q3, Largest Value, Smallest Value, Range, Coefficient of Skewness and Interpreting the values.

In this statistical analysis project, we have examined four key variables: Runs, Boundaries, Wickets and Catches.

We found that the following conclusions and interpretation of data can be drawn from the data set -

1. We can conclude that the data follows normal distribution as the absolute skewness of all variables is less than 2 and absolute kurtosis of all variables is less than 7.

2. Data is Leptokurtic (i.e., Pearsonian Coefficient of kurtosis >0).

3. Data is positively skewed.

4. Positive slope is observed between runs and boundaries which indicates the direct relationship between these two variables. We have predicted runs using boundaries.

5. Correlation between runs and boundaries is positive.

1

**Introduction**

• This mini project is a part of internal assessment of Elements of Statistics of our course.

• The reason for choosing this topic is to analyse the performance of players in IPL on the basis of different variables such as -Runs, Boundaries, Wickets and Catches.

• The objective of the IPL player statistic dataset is to provide a comprehensive set of performance metrics for players participating in the Indian Premier League (IPL). The dataset contains information on various key aspects of a player's performance, including their batting, bowling, and fielding statistics.

• Data is collected after research from Kaggle Website.

<https://www.kaggle.com/datasets/mohammadzamakhan/ipl-players-statistics>

• The method of data collection is Research/Internet.

• The data is collected for the period 2008-2023.

• The data is collected for the following variables - Runs, Boundaries, Wickets and Catches.

• Population for the project - The data exist from 2008-2023 consisting of various variables as well as attributes having multiple numbers of records.

• Type of sample - Non-Random Sample

• Sampling method - As the type of sample is non-random therefore the sampling method is not applicable.

• It is a small type of data containing 98 observations, 4 variables and 1 attribute.

• We have not taken any references into consideration from ebsco and google scholar websites.

2

**Methodology for data analysis**:

The various tools used for data analysis are -

Data Analysis tool kit (Descriptive Statistics), inbuilt excel formulas and conclusion is drawn after carefully studying the data. Data is extracted from the Kaggle website.

3

**Data Analysis:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameters | Runs | Boundaries | Wickets | Catches |
| Mean | 1313.683673 | 170.6428571 | 42.01020408 | 27.94897959 |
| Median | 408.5 | 52 | 27 | 19.5 |
| Mode | 19 | 3 | 5 | 2 |
| Standard Deviation | 1654.355904 | 214.4705901 | 45.6949759 | 29.11207368 |
| Kurtosis | 1.35909851 | 1.02623584 | 1.310211477 | 2.422633522 |
| Skewness | 1.436973557 | 1.35018794 | 1.428280828 | 1.660614533 |
| Range | 6622 | 837 | 182 | 134 |
| Minimum | 12 | 1 | 1 | 1 |
| Maximum | 6634 | 838 | 183 | 135 |

**Table 1.1: Descriptive Statistics**

4

**Conclusion:**

1. On an average a player has scored 1313.68 runs, 170,64 boundaries, 42.01 wickets and 27.95 catches.
2. 50% of the players have scored more than 408.5 runs, 52 boundaries, 27 wickets, 19.5 catches and 50% of the players have scored less than 408.5 runs, 52 boundaries, 27 wickets, 19.5 catches.
3. Most of the players have scored 19 runs, 3 boundaries, 5 wickets and 2 catches.
4. The average difference of runs, boundaries, wickets and catches of each player with respect to Arithmetic Mean is 1654.36 runs, 214.47 boundaries, 45.69 wickets, 29.11 catches.
5. Data is Leptokurtic (i.e., Pearsonian Coefficient of kurtosis >0)
6. Data is positively skewed (i.e., Karl Pearson’s Coefficient of Skewness >0) and AM>Median>Mode.
7. The range with respect to data is 6622 runs, 837 boundaries ,182 wickets and 134 catches.
8. The minimum data values are 12 runs, 1 boundary, 1 wicket, 1 catch.
9. The maximum data values are 6634 runs, 838 boundaries, 183 wickets, 135 catches.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Runs | Boundaries | Wickets | Catches |
| Runs | 1 |  |  |  |
| Boundaries | 0.995132682 | 1 |  |  |
| Wickets | -0.14251683 | -0.14867594 | 1 |  |
| Catches | 0.879844181 | 0.8542608 | 0.086936886 | 1 |

**Table 1.2: Correlation**

5

**Conclusion:**

1. The correlation between runs and boundaries is positive.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Runs | Boundaries | Wickets | Catches |
| Geometic Mean | 337.9766329337.9766329 | 45.57884013 | 20.55074288 | 15.55484774 |
| Harmonic Mean | 42.91798628 | 8.393288197 | 8.688132988 | 7.09416404 |
| Quartile 1 | 66.75 | 8 | 6 | 8 |
| Quartile 3 | 2101.75 | 287 | 59 | 36 |
| Mild Outlier(lower limit) | -2985.75 | -410.5 | -73.5 | -34 |
| Mild Outlier(Upper limit) | 5154.25 | 705.5 | 138.5 | 78 |
| Extreme Outlier(lower limit) | -6038.25 | -829 | -153 | -76 |
| Extreme Outlier(Upper limit) | 8206.75 | 1124 | 218 | 120 |
| Coefficient of variation | 125.9325922 | 125.6838954 | 108.7711353 | 104.1614904 |

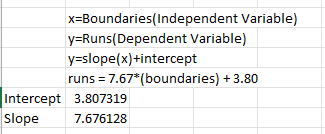
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

6

**Table 1.3: GM, HM, Q1, Q3. Outliers, Coefficient of variation**

1. The geometric mean as per the data set is 337.98 runs, 45.58 boundaries, 20.55 wickets and 15.55 catches.
2. The harmonic mean as per the data set is 42.92 runs, 8.39 boundaries, 8.68 wickets and 7.09 catches.
3. 25% of the players have scored less than 66.75 runs, 8 boundaries, 6 wickets , 8 catches and the remaining the 75% of the players have scored more than 66.75 runs, 8 boundaries, 6 wickets , 8 catches.
4. 75% of the players have scored less than 2101.75 runs, 287 boundaries,59 wickets and 36 catches and the remaining 25% of the players have scored more than 2101.75 runs, 287 boundaries,59 wickets and 36 catches.
5. Since, observation in the data is not within the calculated range ( [Q1 - 1.5(Q3-Q1)],[Q3+1.5(Q3-Q1)] ) , Mild Outliers exist in the following variables - Runs, Boundaries, Wickets and Catches.
6. Since observation in the data is not within the calculated range ( [Q1 - 3(Q3-Q1)],[Q3+3(Q3-Q1)] ) , Extreme Outliers exist in the following variables - Catches.
7. The coefficient of variation is 125.93 for runs ,125.68 for boundaries,108.77 for wickets and 104.16 for catches; it means that variation is quite high which signifies that values are spread across a wide range.

7

****

**Table 1.4: Intercept, Slope (Regression)**

1. Positive slope represents that there is a direct relationship between two variables (runs and boundaries).
2. The value of a dependent variable (runs) is determined by another independent variable (boundaries).
3. As these variables show a positive slope, we can predict that variable- Runs are dependent on an independent variable- boundaries, it means that if the number of boundaries increases, therefore the number of runs will increase.

8

**Conclusions:**

1. Data helps us to conclude that the data follows normal distribution as the absolute skewness of all variables (runs, boundaries, wickets, and catches) is less than 2 and absolute kurtosis of all variables is less than 7.
2. The descriptive statistics of data as per different variables which help us to observe that data is leptokurtic and positively skewed as the values of all the variables is greater than 0.
3. The positive correlation is observed between variables – runs and boundaries , as per the dataset.
4. The values of G.M, HM, Q1, Q3, Outliers and Coefficient of Variations are calculated and analysed .
5. Mild Outliers exist in the following variables - Runs, Boundaries, Wickets and Catches and Extreme Outliers exist in the following variables - Catches.
6. The positive slope is observed between dependent (runs) and independent (boundaries) variables.

9

**References:**

1. Kaggle

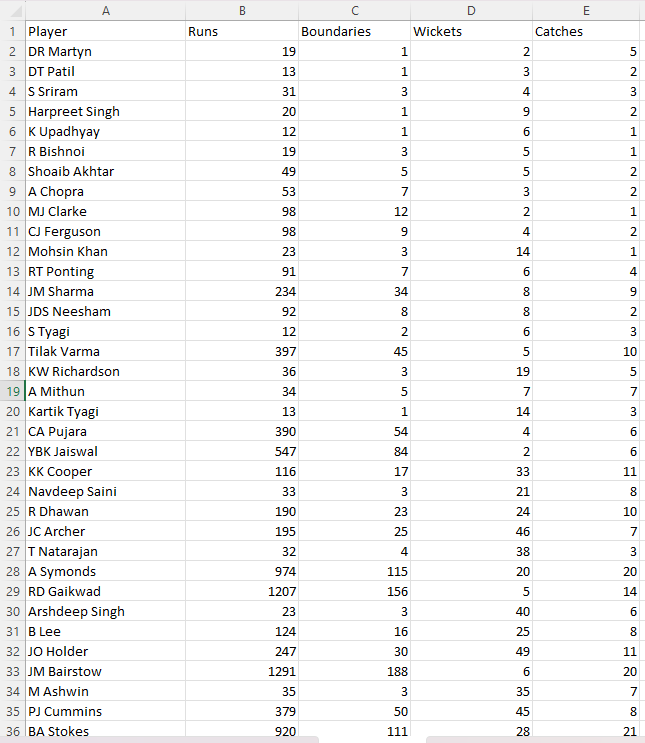
<https://www.kaggle.com/datasets/mohammadzamakhan/ipl-players-statistics>

1. Notes provided by faculty in-charge
2. Class work

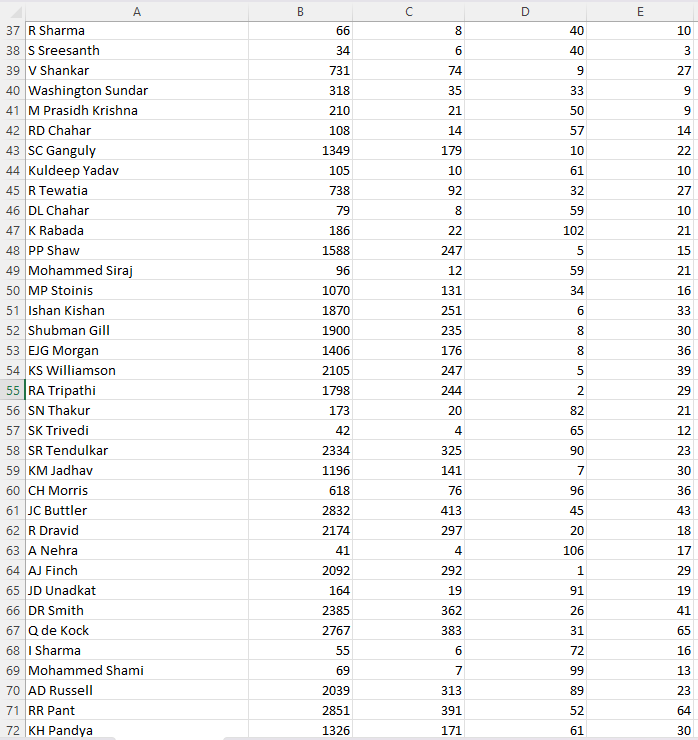
10

**Annexure:**

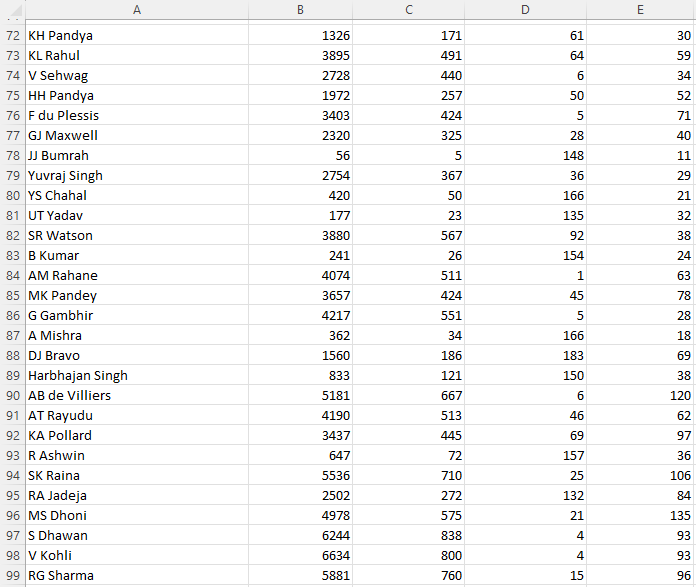
Original Data



11



12



13