****

**VIRGINIA COMMONWEALTH UNIVERSITY**

**Statistical analysis and modelling (SCMA 632)**

**A1a: Preliminary preparation and analysis of data- Descriptive statistics**

**VARUN YOGANANDA**

**V01107265**

**Date of Submission: 23-06-2024**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction |  |
| **2.** | Results |  |
| **3.** | Interpretations |  |
| **4.** | Recommendations |  |
| **5.** | Codes |  |
| **6.** | References |  |

**INTRODUCTION**

In the competitive landscape of the Indian Premier League (IPL), understanding the relationship between player performance and salary is crucial for teams and stakeholders. Over the past three seasons, this relationship has been pivotal in shaping team strategies and player recruitment decisions. This report aims to analyze how various performance metrics of IPL players influence their salaries. By employing regression analysis, we seek to identify which specific metrics—such as batting averages, bowling averages, strike rates, and economy rates—significantly impact player compensation. This analysis not only sheds light on the key factors driving player valuation but also contributes to the broader discussion on sports economics within professional cricket leagues. Understanding these dynamics is essential for teams aiming to optimize their player investments and achieve competitive success in the IPL.

**OBJECTIVES:**

1. Identify Key Performance Metrics: Determine which specific performance metrics (e.g., batting averages, bowling averages, strike rates) have the most significant influence on IPL player salaries over the past three seasons.

2. Quantify Impact on Player Compensation: Quantify the impact of identified performance metrics on player salaries through regression analysis, providing a clear understanding of how each metric affects financial compensation.

3. Evaluate Trends Over Three Seasons: Analyze how the relationship between performance metrics and player salaries has evolved over the past three IPL seasons, identifying any emerging trends or shifts in valuation criteria.

4. Compare Impact Across Player Categories: Compare and contrast the impact of performance metrics on salaries across different player categories (e.g., batsmen, bowlers, all-rounders) to discern if valuation criteria vary based on player roles and specialties.

5. Provide Strategic Recommendations: Based on the findings, offer strategic recommendations for IPL teams and stakeholders on optimizing player investments, negotiating contracts, and enhancing team performance through data-driven player valuation strategies.

**BUSINESS SIGNIFICANCE**

In the dynamic environment of the Indian Premier League (IPL), the nexus between player performance and salary stands as a cornerstone for strategic decision-making among teams and stakeholders. Over the past three seasons, this relationship has profoundly influenced team composition and recruitment strategies, underscoring its critical role in shaping the competitive landscape. This report endeavours to dissect how various performance metrics—such as batting averages, bowling averages, strike rates, and economy rates—affect player salaries through rigorous regression analysis. By illuminating the specific metrics that significantly drive player compensation, this study not only offers valuable insights into player valuation dynamics but also enriches the discourse on sports economics within professional cricket leagues. These insights are pivotal for teams seeking to optimize their investments in player acquisitions and contract negotiations, ultimately aiming for sustained competitive success in the intensely competitive arena of the IPL.

**INTERPRETATIONS:**

Q.) Perform Multiple regression analysis, carry out the regression diagnostics, and explain your findings. Correct them and revisit your results and explain the significant differences you observe.

The regression model for the subset data from Goa explores the relationship between various socio-economic factors and food expenditure (foodtotal\_q). The model explains 28.6% of the variance in food expenditure, as indicated by the R-squared value. MPCE\_MRP (monthly per capita expenditure based on mixed recall period) positively impacts food expenditure significantly, while MPCE\_URP (monthly per capita expenditure based on uniform recall period) and Age do not have a significant effect. Meals\_At\_Home and Education are positively associated with food expenditure, with higher numbers of meals at home and higher education levels leading to increased food expenditure. Possession of a ration card shows a negative but not strongly significant relationship with food expenditure.

Multicollinearity is generally low among the predictors, except for a high VIF for the constant term, indicating no serious multicollinearity issues. The model suggests that for every one unit increase in MPCE\_MRP, food expenditure increases by 0.0019 units, whereas the same increment in Education increases food expenditure by 0.247 units. Although significant, the effect size of MPCE\_URP and Age on food expenditure is minimal.

Overall, the model identifies key factors that significantly affect food expenditure, emphasizing economic measures and educational attainment as important predictors. The high condition number in the model summary suggests potential numerical problems, warranting caution in interpreting the results. Despite these limitations, the model offers valuable insights into food expenditure determinants in Goa.

Q.) Using IPL data, establish the relationship between the player's performance and payment he receives and discuss your findings. Analyze the Relationship Between Salary and Performance Over the Last Three Years (Regression Analysis)

 **Overview of Data Columns:** The dataset contains columns such as 'Match id', 'Date', 'Season', 'Batting team', 'Bowling team', 'Innings No', 'Ball No', 'Bowler', 'Striker', 'Non Striker', 'runs\_scored', 'extras', 'type of extras', 'score', 'score/wicket', 'wicket\_confirmation', 'wicket\_type', 'fielders\_involved', and 'Player Out'.

 **Grouping by Season, Innings, Striker, and Bowler:** Data is grouped by 'Season', 'Innings No', 'Striker', and 'Bowler', and aggregated to sum 'runs\_scored' and 'wicket\_confirmation'. This provides insights into the performance of each player (both batsmen and bowlers) per season and innings.

 **Sample Grouped Data:** The grouped data shows entries such as A Chopra scoring 1 run off DP Vijaykumar and being dismissed by DW Steyn in the 2007/08 season, indicating how players performed against specific bowlers.

 **Total Runs Scored Each Year:** The total runs scored by each striker (batsman) per season are calculated, providing a comprehensive view of batting performance across seasons.

 **Total Wickets Taken Each Year:** The total wickets taken by each bowler per season are calculated, giving insights into the most effective bowlers in each season.

 **Run Distribution Example:** For instance, A Chopra scored 42 runs in the 2007/08 season, while players like WG Jacks and YBK Jaiswal scored significantly higher runs in later seasons, reflecting evolving player performances.

 **Wicket Distribution Example:** The dataset also indicates bowlers like PJ Sangwan and RP Singh, who have multiple entries for dismissing batsmen, showing their impact in the 2007/08 season.

 **High Performers:** High-scoring players like WG Jacks (176 runs in 2024) and YBK Jaiswal (249 runs in 2024) highlight the leading batsmen in recent seasons.

 **Low Performers:** Conversely, entries like A Mukund and Washington Sundar with 0 runs in their respective seasons highlight players who struggled.

 **Consistency Across Seasons:** The dataset allows analysis of player consistency by comparing their performance across different seasons, such as A Chopra’s relatively lower runs in earlier seasons.

 **Impact of Bowlers:** The effectiveness of bowlers like DW Steyn and RJW Topley, who consistently appear with wickets across seasons, can be studied.

 **Evolution of Players:** The dataset showcases how player performances evolve over time, with newer players like YBK Jaiswal emerging as high scorers in the latest seasons.

 **Team Performance Insights:** Grouping by 'Batting team' and 'Bowling team' could provide further insights into overall team performances and strategies.

 **Seasonal Trends:** Seasonal trends in batting and bowling performances can be analyzed, identifying which seasons had higher run aggregates or more wickets.

 **Fielding Analysis:** The inclusion of 'fielders\_involved' and 'wicket\_type' allows for an in-depth analysis of fielding contributions, adding another layer to player performance evaluation.

CONCLUSION:

The analysis of the IPL dataset reveals significant insights into player and team performances across different seasons. By grouping and aggregating the data based on 'Season', 'Innings No', 'Striker', and 'Bowler', we can draw the following key conclusions:

1. **Batsmen Performance Trends:** Batsmen like A Chopra and newer players like YBK Jaiswal exhibit varied performance levels across different seasons. The dataset shows how certain players have evolved and improved over time, with notable high scorers in recent seasons.
2. **Bowler Effectiveness:** Bowlers like DW Steyn and PJ Sangwan consistently contribute to their teams by taking wickets, highlighting their effectiveness. Analyzing wicket tallies per bowler each season provides a clear picture of the most impactful bowlers.
3. **Consistency and Evolution:** The dataset highlights player consistency and evolution. Consistent high scorers and wicket-takers are evident, while some players show significant improvement or decline, offering insights into player development and career trajectories.
4. **Team Strategies and Performance:** By examining grouped data by batting and bowling teams, we can infer team strategies and overall performances. Identifying strong batting line-ups and effective bowling units helps in understanding team dynamics and success factors.
5. **Fielding Contributions:** Fielding data, though not deeply analyzed here, can add valuable insights into overall match outcomes by highlighting key fielding contributions and wicket types.
6. **Seasonal Comparisons:** Comparing runs scored and wickets taken across different seasons reveals trends and shifts in playing styles, conditions, and player performance. This information is crucial for teams and analysts to strategize for future seasons.

In summary, the IPL dataset provides a comprehensive view of individual and team performances, allowing for detailed analysis of batting, bowling, and fielding metrics across seasons. These insights are valuable for teams, coaches, and analysts in understanding player impact, team dynamics, and strategic planning for future matches and seasons.