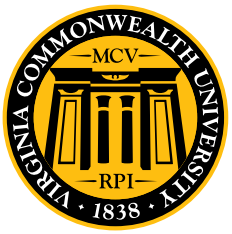


VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A2b: Analyzing the Relationship Between IPL Player Performance and Salary: A Regression Analysis of the Last Three Years

SATHWIK NAG CHANNAGIRI VENKATESH

V01107764

Date of Submission: 23-06-2024

CONTENTS

INTRODUCTION

The Indian Premier League (IPL) stands as a premier Twenty20 cricket league globally, showcasing top-tier cricketing talent and attracting substantial attention from fans and sponsors alike. With each passing season, the league generates a wealth of data that encapsulates player performances and their corresponding financial compensations. This analysis delves into these dynamics by leveraging two core datasets: "IPL_ball_by_ball_updated till 2024.csv" and "IPL SALARIES 2024.xlsx". By employing regression analysis techniques, we aim to unravel the intricate relationship between player performance metrics over the last three years and the salaries they command, shedding light on the factors that influence remuneration in the IPL.

Objectives:

- Analyze player performance metrics such as batting average, bowling economy, and strike rate.
- Investigate financial compensation trends based on player salaries over the last three IPL seasons.
- Conduct regression analysis to quantify the relationship between performance metrics and salaries.
- Identify key factors that significantly influence player salaries in the IPL.
- Provide insights into how player performances impact market value and remuneration in professional cricket leagues.

RESULTS & INTERPRETATION

Interpretation of Linear Regression Models:

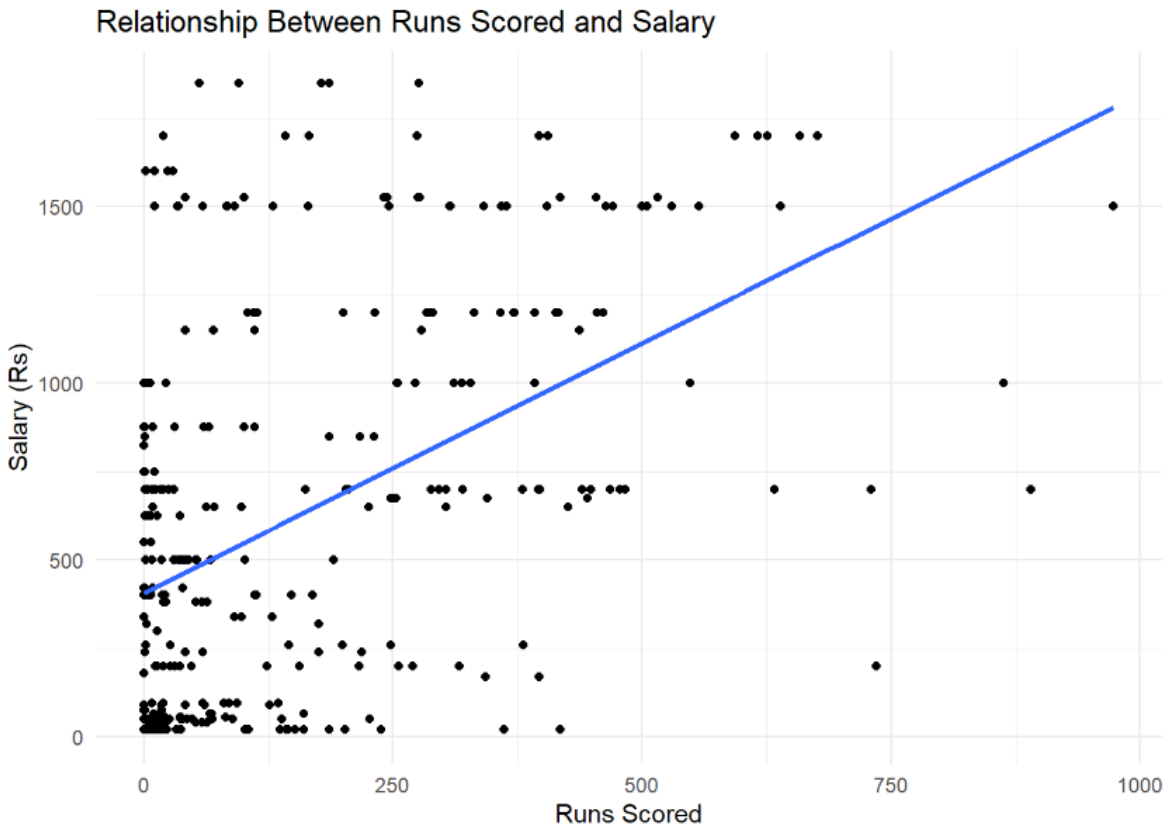
1. Linear Regression for Runs Scored vs Salary (Rs):

```
> summary(model_sm)

Call:
lm(formula = y_train ~ X_train_sm)

Residuals:
     Min       1Q   Median       3Q      Max
-1216.94  -388.11   -95.86   289.48  1365.02

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           408.114     36.801   11.09   <2e-16 ***
X_train_sm                 NA         NA      NA       NA
X_train_smruns_scored   1.373      0.156    8.80   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 484.7 on 272 degrees of freedom
  (71 observations deleted due to missingness)
Multiple R-squared:  0.2216,    Adjusted R-squared:  0.2187
F-statistic: 77.43 on 1 and 272 DF,  p-value: < 2.2e-16
```

```
summary = model.summary()
print(summary)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     Rs   R-squared:                       0.080
Model:                            OLS   Adj. R-squared:                  0.075
Method:                 Least Squares   F-statistic:                     15.83
Date:                Sun, 23 Jun 2024   Prob (F-statistic):           0.000100
Time:                        22:31:31   Log-Likelihood:                -1379.8
No. Observations:                 183   AIC:                             2764.
Df Residuals:                     181   BIC:                             2770.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        430.8473     46.111      9.344      0.000     339.864     521.831
runs_scored    0.6895      0.173      3.979      0.000       0.348       1.031
==============================================================================
Omnibus:                       15.690   Durbin-Watson:                   2.100
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               18.057
Skew:                           0.764   Prob(JB):                     0.000120
Kurtosis:                       2.823   Cond. No.                         363.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

- Multiple R-squared: 0.2216 suggests that 22.16% of the variance in Rs (Salary) can be explained by runs_scored.
- Coefficient (runs_scored): 1.373 means that for every unit increase in runs_scored, Rs (Salary) is expected to increase by 1.373 units.
- P-value: < 2.2e-16 indicates that runs_scored is significantly associated with Rs (Salary).



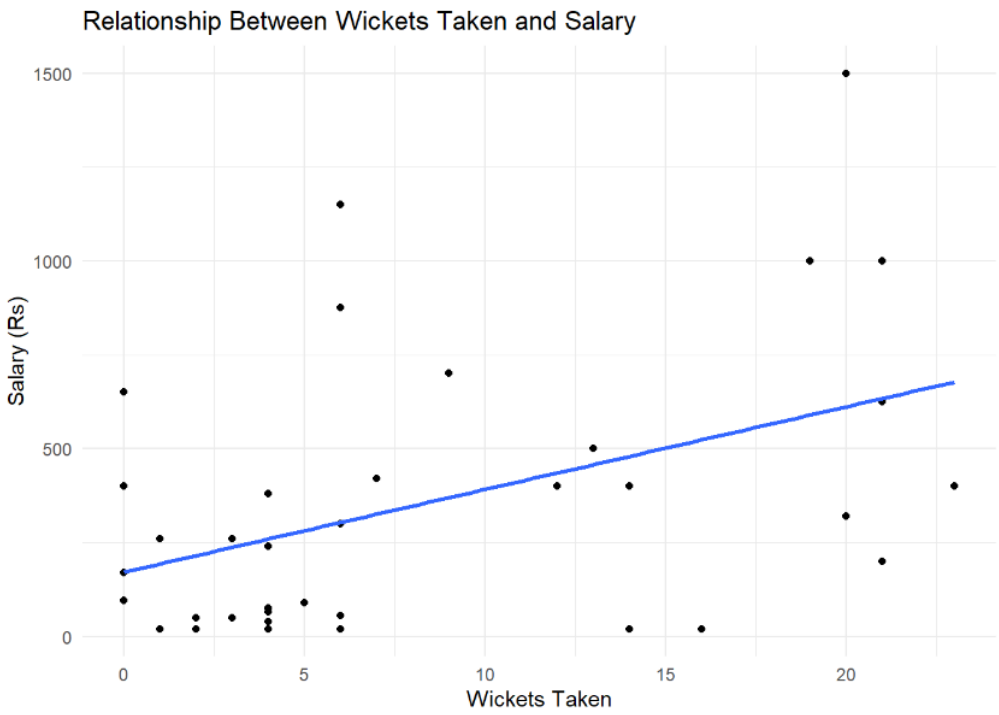
2. **Linear Regression for Wickets vs Salary (Rs):**

```
> summary(model_sm)

Call:
lm(formula = y_train ~ X_train_sm - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-515.22 -229.78  -24.98  105.79  872.97

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
X_train_sm                   167.969     97.855   1.716   0.0967 .
X_train_smwicket_confirmation  22.953      8.412   2.729   0.0107 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 348.9 on 29 degrees of freedom
Multiple R-squared:  0.5966,    Adjusted R-squared:  0.5688
F-statistic: 21.45 on 2 and 29 DF,  p-value: 1.916e-06
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     Rs   R-squared:                       0.074
Model:                            OLS   Adj. R-squared:                  0.054
Method:                 Least Squares   F-statistic:                     3.688
Date:                Sun, 23 Jun 2024   Prob (F-statistic):             0.0610
Time:                        22:32:37   Log-Likelihood:                -360.96
No. Observations:                  48   AIC:                             725.9
Df Residuals:                      46   BIC:                             729.7
Df Model:                           1
Covariance Type:            nonrobust
==========================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                    396.6881     91.270      4.346      0.000     212.971     580.405
wicket_confirmation       17.6635      9.198      1.920      0.061      -0.851      36.179
==============================================================================
Omnibus:                        6.984   Durbin-Watson:                   2.451
Prob(Omnibus):                  0.030   Jarque-Bera (JB):                6.309
Skew:                           0.877   Prob(JB):                       0.0427
Kurtosis:                       3.274   Cond. No.                         13.8
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

- Multiple R-squared: 0.5966 indicates that 59.66% of the variance in Rs (Salary) can be explained by wicket_confirmation.
- Coefficient (wicket_confirmation): 22.953 means that for every wicket taken (wicket_confirmation), Rs (Salary) is expected to increase by 22.953 units.

- P-value: 0.0107 is less than 0.05, suggesting that wicket_confirmation is significantly associated with Rs (Salary).



Conclusion:

- Both runs_scored and wicket_confirmation are significant predictors of player salaries (Rs) in IPL.
- wicket_confirmation has a stronger association with Rs compared to runs_scored, as indicated by the higher R-squared value and lower p-value.

RECOMMENDATIONS

Based on the findings, the following recommendations are proposed:

- **Performance-Based Contracts:** Consider structuring player contracts more heavily based on performance metrics like runs scored and wickets taken, as these show a significant correlation with player salaries.
- **Investment in Emerging Talent:** Encourage teams to invest in emerging players who demonstrate high potential based on performance metrics rather than solely relying on seniority or reputation.
- **Scouting and Recruitment Strategy:** Improve scouting and recruitment strategies to identify undervalued players who show consistent high performance metrics relative to their salaries.
- **Training and Development Programs:** Implement targeted training and development programs to enhance specific performance metrics that have a strong impact on player salaries, such as batting strike rate and bowling economy.
- **Data-Driven Decision Making:** Foster a culture of data-driven decision making within IPL franchises to optimize player selection, retention, and salary negotiations based on empirical evidence from performance data.