# VIRGINIA COMMONWEALTH UNIVERSITY

# Statistical Analysis and Modelling (SCMA 632)

## A2:

Multiple Regression Analysis and Diagnostic Examination in Maharashtra State: Findings, Corrections, and Insights.

Analyzing the Relationship Between IPL Player Performance and Salary

**ANJELINA SAJAN**

**V01107258**

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

## CONTENTS

## INTRODUCTION:

Using R and Python, this study explores the complex dynamics of multiple regression analysis and diagnostic assessment in Maharashtra State, highlighting insights into socio-economic factors influencing food expenditures. Simultaneously, it examines the relationship between IPL player performance and salary across three seasons (2021-2023), employing statistical methods to reveal significant factors driving player compensation in professional cricket.

### Analyzing Multiple Regression Analysis and Diagnostics in Maharashtra State: Insights, Corrections, and Findings

In Maharashtra State, conducting a comprehensive multiple regression analysis and diagnostic examination aims to elucidate the intricate relationship between socio-economic variables and food expenditures. By refining the model and addressing diagnostic issues, this study seeks to provide deeper insights into the factors influencing food consumption patterns in the region.

### Analyzing the Relationship Between IPL Player Performance and Salary

Exploring the connection between IPL player performance and salary involves using regression analysis to uncover how runs scored by batsmen and wickets taken by bowlers relate to their respective salaries over the past three seasons (2021-2023). This analysis aims to shed light on the significant determinants of player compensation in professional cricket, utilizing robust statistical techniques in R and Python to enhance understanding beyond surface-level correlations.

## OBJECTIVE

The objective of this study is as follows:

1. Conduct multiple regression analysis to elucidate the relationship between socio-economic variables (such as income, age, education) and food expenditures in Maharashtra State.
2. Perform diagnostic assessments to identify and correct any model deficiencies or assumptions in the regression analysis of food expenditure data.
3. Explore the impact of monthly per capita expenditures (MPCE), age, meals consumed at home, possession of a ration card, and education level on food expenditures.
4. Analyze the relationship between IPL player performance metrics (runs scored for batsmen and wickets taken for bowlers) and salary to identify significant factors influencing player compensation over the last three seasons (2021-2023).

## BUSINESS SIGNIFICANCE:

### Objective 1: Multiple Regression Analysis in Maharashtra State

1. **Consumer Behavior Insights:** Understanding how socio-economic factors influence food expenditures can provide valuable insights into consumer behavior patterns in Maharashtra. This knowledge is crucial for businesses in the food industry to tailor their marketing strategies and product offerings effectively.
2. **Policy Implications:** Insights from the regression analysis can inform policymakers about the impact of policies related to income distribution, education, and food subsidies. This understanding can aid in designing more targeted interventions to support vulnerable populations and optimize resource allocation.
3. **Market Segmentation:** Businesses can use the findings to segment their target market based on socio-economic characteristics, thereby improving market penetration strategies and enhancing customer satisfaction by offering products that align with consumer spending behaviors.
4. **Operational Efficiency:** By identifying key drivers of food expenditures, businesses can optimize their supply chain management, inventory planning, and pricing strategies to better meet consumer demand and improve operational efficiency.

### Objective 2: Analyzing IPL Player Performance and Salary

1. **Player Recruitment and Retention:** Understanding the relationship between performance metrics and salaries helps IPL teams in making informed decisions regarding player recruitment and retention strategies. Teams can prioritize players who deliver value for money based on performance metrics.
2. **Strategic Investment:** Analysis of player performance and salary can guide team investment decisions in player development, training programs, and coaching staff to maximize player performance and overall team success.
3. **Fan Engagement and Sponsorship:** High-performing players tend to attract more fan engagement and sponsorship opportunities. Teams and sponsors can leverage this data to create more compelling marketing campaigns and sponsorship deals, ultimately increasing revenue streams.
4. **League Competitiveness:** By analyzing player performance and salary trends over time, IPL administrators can ensure the league remains competitive and attractive to both players and fans, fostering long-term growth and sustainability.

a) **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. [data "NSSO68.csv"]**

➢ Multiple regression analysis and the regression diagnostics

```
> # Fit the regression model
> model <- lm(foodtotal_q~ MPCE_MRP+MPCE_URP+Age+Meals_At_Home+Possess_ration_card+Education, data = subset_data)
> # Print the regression results
> print(summary(model))

Call:
lm(formula = foodtotal_q ~ MPCE_MRP + MPCE_URP + Age + Meals_At_Home +
    Possess_ration_card + Education, data = subset_data)

Residuals:
    Min      1Q  Median      3Q     Max
-92.328  -3.488  -0.517   2.764 145.583

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          7.552e+00  5.815e-01  12.988  < 2e-16 ***
MPCE_MRP             1.039e-03  4.697e-05  22.110  < 2e-16 ***
MPCE_URP             2.476e-04  4.284e-05   5.779  7.8e-09 ***
Age                  6.709e-02  5.626e-03  11.924  < 2e-16 ***
Meals_At_Home        1.131e-01  5.745e-03  19.682  < 2e-16 ***
Possess_ration_card -5.210e-01  1.811e-01  -2.877  0.00402 **
Education            2.374e-01  2.253e-02  10.538  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.164 on 7852 degrees of freedom
  (184 observations deleted due to missingness)
Multiple R-squared:  0.2505,    Adjusted R-squared:  0.2499
F-statistic: 437.4 on 6 and 7852 DF,  p-value: < 2.2e-16
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            foodtotal_q   R-squared:                       0.183
Model:                            OLS   Adj. R-squared:                  0.183
Method:                 Least Squares   F-statistic:                     300.8
Date:                Sun, 23 Jun 2024   Prob (F-statistic):               0.00
Time:                        20:29:44   Log-Likelihood:                -26978.
No. Observations:                8043   AIC:                         5.397e+04
Df Residuals:                    8036   BIC:                         5.402e+04
Df Model:                           6
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                   8.2016      0.650     12.622      0.000       6.928       9.475
MPCE_MRP                0.0007   4.81e-05     13.613      0.000       0.001       0.001
MPCE_URP                0.0003   4.49e-05      6.372      0.000       0.000       0.000
Age                     0.1069      0.006     17.336      0.000       0.095       0.119
Meals_At_Home           0.0990      0.006     15.361      0.000       0.086       0.112
Possess_ration_card    -1.8801      0.199     -9.462      0.000      -2.270      -1.491
Education               0.2907      0.025     11.621      0.000       0.242       0.340
==============================================================================
Omnibus:                     4208.018   Durbin-Watson:                   1.687
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           415336.270
Skew:                           1.581   Prob(JB):                         0.00
Kurtosis:                      38.062   Cond. No.                     4.12e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.12e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
In [38]: # Check for multicollinearity using Variance Inflation Factor (VIF)
         vif_data = pd.DataFrame()
         vif_data["feature"] = X.columns
         vif_data["VIF"] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
         print(vif_data)  # VIF Value more than 8 is problematic

                feature        VIF
         0            const  70.730547
         1         MPCE_MRP   2.434452
         2         MPCE_URP   2.324550
         3              Age   1.172517
         4    Meals_At_Home   1.032731
         5  Possess_ration_card   1.115256
         6        Education   1.234124

In [39]: # Extract the coefficients from the model
         coefficients = model.params

In [40]: # Construct the equation
         equation = f"y = {round(coefficients[0], 2)}"
         for i in range(1, len(coefficients)):
             equation += f" + {round(coefficients[i], 6)}*x{i}"

In [41]: # Print the equation
         print(equation)

         y = 8.2 + 0.000655*x1 + 0.000286*x2 + 0.106924*x3 + 0.098999*x4 + -1.88013*x5 + 0.29071*x6
```

**Regression Model Summary and Interpretation**

- *Model Summary:*

The model explains a significant portion of the variance in foodtotal_q, with an R-squared of 0.2505 and an adjusted R-squared of 0.2499. This indicates that approximately 25.05% of the variability in food expenditure can be explained by the predictors included in the model.

- *Coefficients Interpretation*:

Intercept (7.55): This represents the estimated food expenditure when all predictor variables are zero.

- *Predictor Variables:*

MPCE_MRP and MPCE_URP: Higher values of both MPCE_MRP and MPCE_URP are associated with increased food expenditures. Specifically, a unit increase in MPCE_MRP is linked to a 0.001039 unit rise in food expenditure, while MPCE_URP increases food expenditure by 0.000248 units, holding other variables constant.

Age: Each year increase in age is associated with a 0.067089 unit increase in food expenditure, suggesting that older individuals tend to spend more on food, potentially due to dietary needs or lifestyle changes.

Meals_At_Home: More meals consumed at home positively impact food expenditure, with each additional meal increasing food expenditure by 0.113071 units. This likely reflects higher grocery spending associated with home-cooked meals.

Possess_ration_card: Possession of a ration card is negatively associated with food expenditure, indicating a decrease of 0.52103 units in food expenditure. This suggests that ration card holders may access subsidized food, leading to lower overall food expenses.

Education: Higher education levels correlate positively with food expenditure, with each unit increase in education level (on a scale from 1 to 16) associated with a 0.237418 unit increase in food expenditure. This highlights the role of education in influencing economic status and lifestyle choices.

**Model Fit and Diagnostics:**

F-statistic (437.4): The model's F-statistic is significant with a very low p-value (< 2.2e-16), indicating that the overall model is statistically significant.

Residual Standard Error (6.164): This measures the average distance that observed values deviate from the regression line, providing insight into the model's accuracy in predicting food expenditures.

Residuals Analysis: While specific plots (like residual plots and QQ plots) are not detailed here, assessing residuals for homoscedasticity and normality is crucial for validating model assumptions.

Multicollinearity (VIF < 3): The VIF values for predictor variables are relatively low, suggesting no significant multicollinearity issues, which enhances the reliability of the regression coefficients.

**Summary:**

The combined findings indicate that socio-economic factors significantly influence food expenditures in Maharashtra. Higher monthly per capita expenditures (both major and minor items), older age, more meals consumed at home, and higher education levels are associated with increased food expenditures. Conversely, possession of a ration card is linked to reduced food expenditures, likely due to subsidized food access. These insights underscore the complex interplay of economic factors and individual characteristics in shaping food consumption patterns.

In conclusion, the model provides a robust framework for understanding the drivers of food expenditures in Maharashtra, highlighting the importance of socio-economic variables in predicting consumer behaviour related to food consumption.

- **Differences Observed**

```
> head(subset_data$MPCE_MRP,1)
[1] 2877.09
> head(subset_data$MPCE_URP,1)
[1] 2636.75
> head(subset_data$Age,1)
[1] 32
> head(subset_data$Meals_At_Home,1)
[1] 56
> head(subset_data$Possess_ration_card,1)
[1] 2
> head(subset_data$Education,1)
[1] 10
> head(subset_data$foodtotal_q,1)
[1] 22.90039
```

```
In [41]: # Print the equation
         print(equation)

y = 8.2 + 0.000655*x1 + 0.000286*x2 + 0.106924*x3 + 0.098999*x4 + -1.88013*x5 + 0.29071*x6

In [42]: print(subset_data['MPCE_MRP'].head(1).values[0])
         print(subset_data['MPCE_URP'].head(1).values[0])
         print(subset_data['Age'].head(1).values[0])
         print(subset_data['Meals_At_Home'].head(1).values[0])
         print(subset_data['Possess_ration_card'].head(1).values[0])
         print(subset_data['Education'].head(1).values[0])
         print(subset_data['foodtotal_q'].head(1).values[0])

2877.09
2636.75
32
56.0
2.0
10.0
22.9003875
```

After revisiting the regression analysis and diagnostics, the significant differences observed are:

1. **Adjusted R-squared**: The adjusted R-squared value after revisiting is 0.2499, which indicates that approximately 25% of the variability in the dependent variable (food expenditure) is explained by the independent variables (MPCE_MRP, MPCE_URP, Age, Meals_At_Home, Possess_ration_card, Education) included in the model. This suggests a moderate fit of the model to the data.
2. **Coefficients**:
   - **MPCE_MRP**: For each unit increase in MPCE_MRP (Monthly Per Capita Expenditure at market rates), there is a corresponding increase of approximately 0.001039 units in food expenditure, holding other variables constant.
   - **MPCE_URP**: Similarly, an increase in MPCE_URP (Monthly Per Capita Expenditure at rural prices) by one unit is associated with a 0.000248 unit increase in food expenditure, all else being equal.
   - **Age**: Age has a positive coefficient of 0.067089, indicating that older individuals tend to have higher food expenditures, assuming other factors are constant.
   - **Meals_At_Home**: A unit increase in the number of meals prepared at home (Meals_At_Home) leads to a 0.113071 unit increase in food expenditure.

- **Possess_ration_card**: Possession of a ration card (Possess_ration_card) is associated with a decrease in food expenditure, with a coefficient of -0.52103. This might suggest that those with ration cards spend less on food, possibly due to subsidies or access to government-supported food supplies.
- **Education**: Higher levels of education (Education) are positively associated with food expenditure, with a coefficient of 0.237418. This implies that more educated individuals tend to spend more on food, all else being equal.

3. **Variance Inflation Factors (VIF)**: The VIF values for all variables are well below 8, indicating no problematic multicollinearity among the independent variables. This suggests that the predictors in the model are not excessively correlated with each other.
4. **Residuals**:
    - **Residual Standard Error**: The residual standard error is 6.164, which represents the average amount that the observed values deviate from the predicted values by the model.
    - **Residuals Distribution**: The residuals appear reasonably normally distributed, as inferred from the minimum, 1st quartile, median, 3rd quartile, and maximum values provided.

In summary, the regression model explains a significant portion of the variability in food expenditure using the selected predictors. Age, education, and factors related to income (MPCE_MRP, MPCE_URP) and living conditions (Meals_At_Home, Possess_ration_card) appear to be significant drivers of food expenditure in the analyzed dataset from Maharashtra (MH). The model's fit is reasonable, and there are no apparent issues with multicollinearity among the predictors.

## b) 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)

- *Establish the relationship between the player's performance and payment he receives*

```
# Create a new column in df_salary with matched names from df_runs
df_salary['Matched_Player'] = df_salary['Player'].apply(lambda x: match_names(x, df_runs['Striker'].tolist()))

# Merge the DataFrames on the matched names
df_merged = pd.merge(df_salary, df_runs, left_on='Matched_Player', right_on='Striker')

In [11]: df_original = df_merged.copy()

In [12]: #susbsets data for last three years
df_merged = df_merged.loc[df_merged['Season'].isin(['2021', '2022', '2023'])]

In [13]: df_merged.Season.unique()
Out[13]: array(['2023', '2022', '2021'], dtype=object)

In [14]: df_merged.head()
Out[14]:
```

| | Player | Salary | Rs | international | iconic | Matched_Player | Season | Striker | runs_scored |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Abhishek Porel | 20 lakh | 20 | 0 | NaN | Abishek Porel | 2023 | Abishek Porel | 33 |
| 3 | Anrich Nortje | 6.5 crore | 650 | 1 | NaN | A Nortje | 2022 | A Nortje | 1 |
| 4 | Anrich Nortje | 6.5 crore | 650 | 1 | NaN | A Nortje | 2023 | A Nortje | 37 |
| 13 | Axar Patel | 9 crore | 900 | 0 | NaN | AR Patel | 2021 | AR Patel | 40 |
| 14 | Axar Patel | 9 crore | 900 | 0 | NaN | AR Patel | 2022 | AR Patel | 182 |

The salary data was merged with the runs and wickets datasets using the matched player names. This merge allowed for analysis of the relationship between a player's performance (in terms of runs scored and wickets taken) and their salary.

### Analysis of Batsmen Performance and Salary

- Linear Regression for Runs Scored vs. Salary:
    - A linear regression model was created to establish the relationship between runs_scored and Salary (Rs) for the last three years (2021-2023).
    - The model, fitted on 183 observations, had an R-squared value of 0.080, indicating that 8% of the variance in salary can be explained by runs scored.
    - The coefficient for runs_scored was 0.6895 with a p-value of 0.000, showing a statistically significant positive relationship between runs scored and salary.
    - For instance, a player scoring 100 runs in a season is associated with an average increase of approximately Rs 68.95 lakh in their salary.

### Analysis of Bowlers Performance and Salary

- Linear Regression for Wickets Taken vs. Salary:
    - A similar linear regression model was created to examine the relationship between wicket_confirmation and Salary (Rs) for the year 2022.
    - This model, fitted on 48 observations, had an R-squared value of 0.074, indicating that 7.4% of the variance in salary can be explained by wickets taken.

- The coefficient for wicket_confirmation was 17.6635 with a p-value of 0.061, suggesting a positive but not statistically significant relationship at the 0.05 level between wickets taken and salary.
- For example, taking 10 wickets in a season is associated with an average increase of approximately Rs 1.766 crore in salary, though this result is not statistically significant.

**Discussion of Findings**

- **Batsmen Performance**:
  - The linear regression results show a significant positive relationship between the runs scored by a batsman and their salary. This indicates that, generally, higher-scoring batsmen tend to receive higher salaries.
  - However, the low R-squared value (0.080) suggests that while runs scored is a significant factor, many other factors influence a player's salary that are not captured in this model.
- **Bowlers Performance**:
  - The relationship between wickets taken and salary is positive, indicating that taking more wickets tends to be associated with a higher salary. However, this relationship is not statistically significant at the 0.05 level (p-value = 0.061).
  - Similar to batsmen, the low R-squared value (0.074) indicates that additional factors beyond just wickets taken impact a bowler's salary.
- **Overall Insights**:
  - Player salaries in IPL are influenced by performance metrics such as runs scored and wickets taken, but these metrics alone do not fully explain the variation in salaries.
  - Factors such as player experience, marketability, team strategy, and roles within the team likely also play significant roles in determining salaries.
  - The models suggest that while performance metrics are important, they are part of a broader set of considerations that teams use to determine player payments.

➢ *Relationship Between Salary and Performance Over the Last Three Years(Regression Analysis)*

```python
import pandas as pd
from sklearn.model_selection import train_test_split
import statsmodels.api as sm

# Assuming df_merged is already defined and contains the necessary columns
X = df_merged[['runs_scored']] # Independent variable(s)
y = df_merged['Rs'] # Dependent variable

# Split the data into training and test sets (80% for training, 20% for testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Add a constant to the model (intercept)
X_train_sm = sm.add_constant(X_train)

# Create a statsmodels OLS regression model
model = sm.OLS(y_train, X_train_sm).fit()

# Get the summary of the model
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:                        19:45:12   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 specifie
```

```r
# Convert X to a data frame
X <- data.frame(runs_scored = df_merged$runs_scored)

# Dependent variable
y <- df_merged$Rs

# Create train/test split indices
trainIndex <- createDataPartition(y, p = .8, list = FALSE, times = 1)

# Split the data
X_train <- X[trainIndex, , drop = FALSE]  # use drop = FALSE to maintain data frame structure
X_test <- X[-trainIndex, , drop = FALSE]  # use drop = FALSE to maintain data frame structure

# Print dimensions to verify
print(dim(X_train))
print(dim(X_test))


X_train <- X[trainIndex]
X_test <- X[-trainIndex]
y_train <- y[trainIndex]
y_test <- y[-trainIndex]

# Check dimensions
length(X_train)
length(X_test)
length(y_train)
length(y_test)



X_train <- cbind(1, X_train)  # Adding constant
model <- lm(y_train ~ ., data = data.frame(X_train))
summary(model)
```

```
> summary(model)

Call:
lm(formula = y_train ~ ., data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.26350 -0.44191  0.00687  0.44555  1.21189 

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -0.51504    0.20841  -2.471   0.0269 *
X1                NA         NA      NA       NA  
X2           0.22382    0.18610   1.203   0.2490  
X3           0.03762    0.24793   0.152   0.8816  
X4           0.09322    0.21253   0.439   0.6676  
X5          -0.03006    0.19586  -0.153   0.8802  
X6           0.29955    0.26894   1.114   0.2841  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7476 on 14 degrees of freedom
Multiple R-squared:  0.1621,	Adjusted R-squared:  -0.1371 
F-statistic: 0.5418 on 5 and 14 DF,  p-value: 0.7419
```

### Summary of the Analysis

To understand the relationship between player salaries and their performance, we analyzed the data from the last three years, focusing on both batsmen and bowlers. Two main metrics were considered:

Runs Scored for batsmen.

Wickets Taken for bowlers.

We used linear regression models to explore the relationship between these performance metrics and the players' salaries, given in crores (Rs). The data was subsetted to include only the last three years: 2021, 2022, and 2023.

### Batsmen Performance Analysis

*Linear Regression Model Summary for Batsmen:*

- Dependent Variable: Rs (Salary in crores)
- Independent Variable: runs_scored (Total runs scored)

### RESULTS

- R-squared: 0.080. This indicates that 8% of the variation in salaries can be explained by the runs scored.
- Coefficient (runs_scored): 0.6895. For every additional run scored, the salary increases by approximately 0.69 crores.
- P-value for runs_scored: 0.000. The low p-value indicates that runs scored is a statistically significant predictor of salary.
- Constant: 430.8473. This is the base salary in crores when no runs are scored.
- F-statistic: 15.83 with a p-value of 0.000100, indicating the model is statistically significant.
- Durbin-**Watson:** 2.100, suggesting no strong autocorrelation in the residuals.

## INTERPRETATION

The linear regression analysis reveals that runs scored by batsmen significantly influence their salaries. For every additional run scored, a batsman's salary increases by approximately 0.69 crores, holding other factors constant. The model, though statistically significant with a low p-value, explains only 8% of the variability in salaries. This suggests that while runs scored are an important factor, other variables such as experience, player role, and market demand also contribute to determining batsmen's salaries in cricket.

### Bowlers Performance Analysis

*Linear Regression Model Summary for Bowlers*:

- Dependent Variable: Rs (Salary in crores)
- Independent Variable: wicket_confirmation (Total wickets taken)

## RESULTS

- R-squared: 0.074. This indicates that 7.4% of the variation in salaries can be explained by the wickets taken.
- Coefficient (wicket_confirmation): 17.6635. For every additional wicket taken, the salary increases by approximately 17.66 crores.
- P-value for wicket_confirmation: 0.061. The p-value is slightly above the conventional threshold of 0.05, suggesting that wickets taken is marginally significant as a predictor of salary.
- Constant: 396.6881. This is the base salary in crores when no wickets are taken.
- F-statistic: 3.688 with a p-value of 0.0610, indicating the model is marginally statistically significant.
- Durbin-Watson: 2.451, suggesting no strong autocorrelation in the residuals.

## INTERPRETATION:

In the context of bowlers, the analysis indicates that the number of wickets taken modestly affects their salaries. Each additional wicket taken is associated with an increase of about 17.66 crores in salary, assuming other factors remain unchanged. However, the model's explanatory power is limited, with wickets taken explaining only 7.4% of the variability in salaries. This implies that while wicket-taking ability is valued, other factors such as bowling style, match conditions, and team strategy also heavily influence bowlers' earnings in professional cricket.

**CONCLUSION:**

The multiple regression analysis conducted on food expenditure in Maharashtra reveals insightful relationships between socio-economic factors and spending patterns. Key findings include the positive associations of higher monthly per capita expenditures (MPCE), older age, more meals consumed at home, and higher education levels with increased food expenditures. Conversely, possession of a ration card shows a negative correlation, likely due to subsidized food access. The model's robustness is supported by a significant F-statistic, low residual standard error, and satisfactory diagnostics for multicollinearity. These findings underscore the complex interplay of economic factors and individual characteristics influencing food consumption patterns in Maharashtra.

Analyzing the relationship between IPL player performance and salary over the last three years reveals nuanced insights. For batsmen, there is a statistically significant positive correlation between runs scored and salary, indicating that higher-scoring players tend to command higher earnings, though runs scored explain only a modest proportion of salary variance. For bowlers, while there is a positive relationship between wickets taken and salary, it is not statistically significant at the conventional level, suggesting other factors beyond wicket-taking influence earnings. These findings highlight that while performance metrics play a role in determining player salaries in the IPL, other factors such as experience, market demand, and team strategies also significantly contribute to compensation decisions.