Regression Analyses with Cricket Data

In week 1, we took a brief look at the cricket match of statistics of the Indian Premier league in 2018 (IPL2018teams dataset). In this week, we will look at the player level statistics. In particular, we are interested in whether the player performance impact their salaries.

Import useful libraries

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.formula.api as sm
```

Import cricket data

In our data repository, there is a data set "IPL18Player.csv" which contains performance statistics as well as salary information of cricket players in the Indian Premier League in 2018.

In [2]:
```
IPLPlayer=pd.read_csv("../../Data/Week 4/IPL18Player.csv")
IPLPlayer.head()
```

Out[2]:

	player_id	long_scorecard_name	Salary	team	matches	wins	team_runs_for	team_runs_against	matches_keeper	byes_conceded	...	bowling_d
0	8931	AT Rayudu	343750.0	Chennai Super Kings	16	11	2809	2750	0	0	...	
1	254771	D Shorey	31250.0	Chennai Super Kings	1	1	128	127	0	0	...	
2	44613	DJ Bravo	1000000.0	Chennai Super Kings	16	11	2809	2750	0	0	...	
3	214425	DJ Willey	NaN	Chennai Super Kings	3	2	484	483	0	0	...	
4	258155	DL Chahar	125000.0	Chennai Super Kings	12	9	2117	2068	0	0	...	

5 rows × 35 columns

Data Exploration and Preparation

In [3]:
```
IPLPlayer.shape
```

Out[3]: (149, 35)

Missing Values

In [4]:
```
IPLPlayer.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149 entries, 0 to 148
Data columns (total 35 columns):
player_id                149 non-null int64
long_scorecard_name      149 non-null object
Salary                   141 non-null float64
team                     149 non-null object
matches                  149 non-null int64
wins                     149 non-null int64
team_runs_for            149 non-null int64
team_runs_against        149 non-null int64
matches_keeper           149 non-null int64
byes_conceded            149 non-null int64
moms                     149 non-null int64
innings                  149 non-null int64
not_outs                 149 non-null int64
runs                     149 non-null int64
balls_faced              149 non-null int64
fours                    149 non-null int64
sixes                    149 non-null int64
matches_bowled           149 non-null int64
balls_bowled             149 non-null int64
wickets                  149 non-null int64
runs_conceded            149 non-null int64
catches                  149 non-null int64
stumpings                149 non-null int64
run_outs                 149 non-null int64
batting_dot_balls        149 non-null int64
bowling_dot_balls        149 non-null int64
bowling_sixes            149 non-null int64
no_balls                 149 non-null int64
balls_bowled_1_to_6      149 non-null int64
runs_conceded_1_to_6     149 non-null int64
balls_bowled_7_to_14     149 non-null int64
runs_conceded_7_to_14    149 non-null int64
balls_bowled_15_to_20    149 non-null int64
runs_conceded_15_to_20   149 non-null int64
event_winner             149 non-null int64
dtypes: float64(1), int64(32), object(2)
memory usage: 40.8+ KB
```

There are missing values in the salary variable. We will drop observations with missing values.

In [5]:
```
IPLPlayer=IPLPlayer.dropna()
IPLPlayer.shape
```

Out[5]: (141, 35)

Create useful variables

Create dummy variables to indicate the role of the players.

- Create a variable to indicate whether a player had played as a batsman.

  The variable "innings" indicates how many innings a player had batted in.

In [6]:
```
IPLPlayer['batsman']=np.where(IPLPlayer['innings']> 0, 1, 0)
IPLPlayer['batsman'].describe()
```

Out[6]:
```
count    141.000000
mean       0.943262
std        0.232165
min        0.000000
25%        1.000000
50%        1.000000
75%        1.000000
max        1.000000
Name: batsman, dtype: float64
```

- Create a variable to indicate bowler.

In [7]:
```
IPLPlayer['bowler']=np.where(IPLPlayer['matches_bowled']> 0, 1, 0)
IPLPlayer['bowler'].describe()
```

Out[7]:
```
count    141.000000
mean       0.631206
std        0.484198
min        0.000000
25%        0.000000
50%        1.000000
75%        1.000000
max        1.000000
Name: bowler, dtype: float64
```

The last type of player that is not captured by either batsman or bowler is wicket keeper. In the dataset, the variable "matches_keeper" indicates the number of matches that a player is a wicket keeper.

Performance Measures

1. batting average = runs / the numbers of outs
2. batting strike rate = (runs * 100) / balls faced
3. bowling average = runs conceded / wicket taken
4. bowling strike rate = number of balls bowled / wicket taken

Notice that if a batsman has scored runs but not been dismissed, his batting average is technically infinite. Similarly, if a player did not face any ball, his batting strike would be infinite and if a player did not lose any wicket, his bowling average or bowling strike would be infinite.

We will not be able to run a regression when our variables have some infinite values.

There are two alternatives we will consider to deal with this issue.

1. Add 1 to the number of outs, balls faced, andn wickets taken in calculating the above variables.
2. Instead of creating the above measures, we can simply include total runs, total number of outs, and balls faced to measure a batsman's performance, and include runs conceded, number of balls bowled, and wickets taken to measure a bowler's performance.

In [8]:
```
IPLPlayer['outs']=np.where(IPLPlayer['batsman']==1, IPLPlayer['innings']-IPLPlayer['not_outs'], 0)
IPLPlayer['outs'].describe()
```

Out[8]:
```
count    141.000000
mean       5.000000
std        4.605897
min        0.000000
25%        1.000000
50%        4.000000
75%        9.000000
max       16.000000
Name: outs, dtype: float64
```

Create batting average, batting strke rate, bowling average, and bowling strike rate variables. Add 1 to the number of outs, balls faced, andn wickets taken in calculating these variables.

In [9]:
```
IPLPlayer['batting_average']=IPLPlayer['runs']/(IPLPlayer['outs']+1)
IPLPlayer['batting_strike']=IPLPlayer['runs']/((IPLPlayer['balls_faced']+1))*100
IPLPlayer['bowling_average']=IPLPlayer['runs_conceded']/(IPLPlayer['wickets']+1)
IPLPlayer['bowling_strike']=IPLPlayer['balls_bowled']/(IPLPlayer['wickets']+1)
```

In [10]:
```
IPLPlayer['batting_average'].describe()
```

Out[10]:
```
count    141.000000
mean      15.093066
std       13.761819
min        0.000000
25%        4.000000
50%       12.500000
75%       23.000000
max       65.000000
Name: batting_average, dtype: float64
```

In [11]:
```
IPLPlayer['batting_strike'].describe()
```

Out[11]:
```
count    141.000000
mean     104.164456
std       53.873378
min        0.000000
25%       73.913043
50%      118.446602
75%      139.669421
max      250.000000
Name: batting_strike, dtype: float64
```

In [12]:
```
IPLPlayer['bowling_average'].describe()
```

Out[12]:
```
count    141.000000
mean      17.493864
std       16.108488
min        0.000000
25%        0.000000
```

```
50%       20.052632
75%       27.466667
max       72.000000
Name: bowling_average, dtype: float64
```

In [13]:
```
IPLPlayer['bowling_strike'].describe()
```

Out[13]:
```
count    141.000000
mean      11.478621
std       10.295591
min        0.000000
25%        0.000000
50%       12.500000
75%       19.600000
max       42.000000
Name: bowling_strike, dtype: float64
```

Regression Analyses

First let's run a regression of the salary on the type of player, batsman, bowler, and all-rounder.

In [14]:
```
reg_IPL1=sm.ols(formula = 'Salary ~ batsman+ bowler+ batsman*bowler', data= IPLPlayer, missing="drop").fit()
print(reg_IPL1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.060
Model:                            OLS   Adj. R-squared:                  0.046
Method:                 Least Squares   F-statistic:                     4.379
Date:                Mon, 26 Jul 2021   Prob (F-statistic):             0.0143
Time:                        00:22:00   Log-Likelihood:                -2069.2
No. Observations:                 141   AIC:                             4144.
Df Residuals:                     138   BIC:                             4153.
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept      2.859e+05   1.11e+05      2.577      0.011    6.65e+04    5.05e+05
batsman         4.61e+05   1.11e+05      4.156      0.000    2.42e+05     6.8e+05
bowler        -1.668e+05   1.11e+05     -1.504      0.135   -3.86e+05    5.25e+04
batsman:bowler  8340.1549    1.2e+05      0.070      0.945   -2.29e+05    2.45e+05
==============================================================================
Omnibus:                       29.772   Durbin-Watson:                   2.052
```

Next we will first focus on performance of batsman.

We will first simply use the total number of runs, number of not outs, and number of balls faced to measure players' performance.

In [15]:
```
reg_IPL2=sm.ols(formula = 'Salary ~ runs', data= IPLPlayer).fit()
print(reg_IPL2.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.267
Model:                            OLS   Adj. R-squared:                  0.261
Method:                 Least Squares   F-statistic:                     50.57
Date:                Mon, 26 Jul 2021   Prob (F-statistic):           5.54e-11
Time:                        00:22:01   Log-Likelihood:                -2051.7
No. Observations:                 141   AIC:                             4107.
Df Residuals:                     139   BIC:                             4113.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   3.878e+05   5.39e+04      7.198      0.000    2.81e+05    4.94e+05
runs        1737.9337    244.400      7.111      0.000    1254.711    2221.157
==============================================================================
Omnibus:                       19.056   Durbin-Watson:                   2.116
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               22.155
Skew:                           0.913   Prob(JB):                     1.55e-05
Kurtosis:                       3.660   Cond. No.                         277.
==============================================================================

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

In [16]:
```
reg_IPL3=sm.ols(formula = 'Salary ~ runs+not_outs', data= IPLPlayer).fit()
print(reg_IPL3.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.318
Model:                            OLS   Adj. R-squared:                  0.308
Method:                 Least Squares   F-statistic:                     32.15
Date:                Mon, 26 Jul 2021   Prob (F-statistic):           3.45e-12
Time:                        00:22:01   Log-Likelihood:                -2046.6
No. Observations:                 141   AIC:                             4099.
Df Residuals:                     138   BIC:                             4108.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    2.88e+05   6.07e+04      4.747      0.000    1.68e+05    4.08e+05
runs        1491.1582    248.725      5.995      0.000     999.353    1982.963
not_outs    8.955e+04   2.79e+04      3.215      0.002    3.45e+04    1.45e+05
==============================================================================
Omnibus:                       16.162   Durbin-Watson:                   2.092
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               18.001
Skew:                           0.806   Prob(JB):                     0.000123
Kurtosis:                       3.682   Cond. No.                         333.
==============================================================================

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

In [17]:
```
reg_IPL4=sm.ols(formula = 'Salary ~ runs+not_outs+balls_faced', data= IPLPlayer).fit()
print(reg_IPL4.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.321
Model:                            OLS   Adj. R-squared:                  0.306
Method:                 Least Squares   F-statistic:                     21.60
Date:                Mon, 26 Jul 2021   Prob (F-statistic):           1.62e-11
Time:                        00:22:01   Log-Likelihood:                -2046.3
No. Observations:                 141   AIC:                             4101.
```

```
Df Residuals:                   137   BIC:                             4112.
Df Model:                         3
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept    3.013e+05   6.29e+04      4.791      0.000    1.77e+05    4.26e+05
runs         2871.9166   1711.997      1.678      0.096    -513.440    6257.273
not_outs     8.945e+04   2.79e+04      3.207      0.002    3.43e+04    1.45e+05
balls_faced -2044.5867   2508.113     -0.815      0.416   -7004.207    2915.033
==============================================================================
Omnibus:                       16.922   Durbin-Watson:                   2.095
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               19.142
Skew:                           0.819   Prob(JB):                     6.97e-05
Kurtosis:                       3.758   Cond. No.                         418.
==============================================================================

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

In the next regressions, we will use the modified batting average and batting strike variables to measure player performance.

In [18]:
```
reg_IPL5=sm.ols(formula = 'Salary ~ batting_average', data= IPLPlayer).fit()
print(reg_IPL5.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.233
Model:                            OLS   Adj. R-squared:                  0.227
Method:                 Least Squares   F-statistic:                     42.13
Date:                Mon, 26 Jul 2021   Prob (F-statistic):           1.40e-09
Time:                        00:22:01   Log-Likelihood:                -2054.9
No. Observations:                 141   AIC:                             4114.
Df Residuals:                     139   BIC:                             4120.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      3.072e+05   6.52e+04      4.716      0.000    1.78e+05    4.36e+05
batting_average  2.074e+04   3194.873      6.491      0.000    1.44e+04    2.71e+04
==============================================================================
Omnibus:                       18.526   Durbin-Watson:                   2.083
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               21.526
Skew:                           0.929   Prob(JB):                     2.12e-05
Kurtosis:                       3.457   Cond. No.                         30.4
==============================================================================

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

In [19]:
```
reg_IPL6=sm.ols(formula = 'Salary ~ batting_average+batting_strike', data= IPLPlayer).fit()
print(reg_IPL6.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.234
Model:                            OLS   Adj. R-squared:                  0.223
Method:                 Least Squares   F-statistic:                     21.12
Date:                Mon, 26 Jul 2021   Prob (F-statistic):           9.96e-09
Time:                        00:22:02   Log-Likelihood:                -2054.7
No. Observations:                 141   AIC:                             4115.
Df Residuals:                     138   BIC:                             4124.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      2.668e+05   9.69e+04      2.754      0.007    7.52e+04    4.58e+05
batting_average  1.903e+04   4409.440      4.315      0.000    1.03e+04    2.77e+04
batting_strike    635.3041   1126.380      0.564      0.574   -1591.892    2862.500
==============================================================================
Omnibus:                       17.736   Durbin-Watson:                   2.061
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               20.405
Skew:                           0.908   Prob(JB):                     3.71e-05
Kurtosis:                       3.416   Cond. No.                         262.
==============================================================================

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

We will now turn to bowlers' performance.

Again, we will first use number of runs conceded, number of balls bowled, and number of wickets taken to measure bowlers' performance.

In [20]:
```
reg_IPL7=sm.ols(formula = 'Salary ~ runs_conceded', data= IPLPlayer).fit()
print(reg_IPL7.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.023
Model:                            OLS   Adj. R-squared:                  0.015
Method:                 Least Squares   F-statistic:                     3.200
Date:                Mon, 26 Jul 2021   Prob (F-statistic):            0.0758
Time:                        00:22:02   Log-Likelihood:                -2072.0
No. Observations:                 141   AIC:                             4148.
Df Residuals:                     139   BIC:                             4154.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      5.438e+05   6.53e+04      8.322      0.000    4.15e+05    6.73e+05
runs_conceded   569.3099    318.271      1.789      0.076     -59.968    1198.587
==============================================================================
Omnibus:                       39.245   Durbin-Watson:                   2.073
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               62.394
Skew:                           1.414   Prob(JB):                     2.83e-14
Kurtosis:                       4.621   Cond. No.                         271.
==============================================================================

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

In [21]:
```
reg_IPL8=sm.ols(formula = 'Salary ~ runs_conceded+balls_bowled', data= IPLPlayer).fit()
print(reg_IPL8.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.042
```

```
Model:                            OLS   Adj. R-squared:                  0.028
Method:                 Least Squares   F-statistic:                     3.026
Date:                Mon, 26 Jul 2021   Prob (F-statistic):             0.0518
Time:                        00:22:02   Log-Likelihood:                -2070.5
No. Observations:                 141   AIC:                             4147.
Df Residuals:                     138   BIC:                             4156.
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept      5.565e+05   6.54e+04      8.514      0.000    4.27e+05    6.86e+05
runs_conceded -2749.4793   2004.845     -1.371      0.172   -6713.666    1214.707
balls_bowled   4574.8787   2729.042      1.676      0.096    -821.265    9971.023
==============================================================================
Omnibus:                       41.062   Durbin-Watson:                   2.075
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               67.476
Skew:                           1.449   Prob(JB):                     2.23e-15
Kurtosis:                       4.756   Cond. No.                         336.
==============================================================================

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

In [22]:
```
reg_IPL9=sm.ols(formula = 'Salary ~ runs_conceded+balls_bowled+wickets', data= IPLPlayer).fit()
print(reg_IPL9.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.049
Model:                            OLS   Adj. R-squared:                  0.028
Method:                 Least Squares   F-statistic:                     2.329
Date:                Mon, 26 Jul 2021   Prob (F-statistic):             0.0772
Time:                        00:22:02   Log-Likelihood:                -2070.1
No. Observations:                 141   AIC:                             4148.
Df Residuals:                     137   BIC:                             4160.
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept      5.543e+05   6.54e+04      8.472      0.000    4.25e+05    6.84e+05
runs_conceded -3049.2063   2029.065     -1.503      0.135   -7061.542     963.129
balls_bowled   6343.2378   3284.524      1.931      0.056    -151.683    1.28e+04
wickets       -2.737e+04   2.83e+04     -0.968      0.335   -8.33e+04    2.85e+04
==============================================================================
Omnibus:                       41.167   Durbin-Watson:                   2.050
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               67.717
Skew:                           1.453   Prob(JB):                     1.97e-15
Kurtosis:                       4.757   Cond. No.                         337.
==============================================================================

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

In the next regression, we will use the modified bowling average and bowling strike variables to measure player performance.

In [23]:
```
reg_IPL10=sm.ols(formula = 'Salary ~ bowling_average+bowling_strike', data= IPLPlayer).fit()
print(reg_IPL10.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.054
Model:                            OLS   Adj. R-squared:                  0.040
Method:                 Least Squares   F-statistic:                     3.912
Date:                Mon, 26 Jul 2021   Prob (F-statistic):             0.0223
Time:                        00:22:03   Log-Likelihood:                -2069.7
No. Observations:                 141   AIC:                             4145.
Df Residuals:                     138   BIC:                             4154.
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept        6.535e+05   7.33e+04      8.918      0.000    5.09e+05    7.98e+05
bowling_average -3.415e+04   1.25e+04     -2.740      0.007   -5.88e+04   -9507.025
bowling_strike   4.914e+04   1.95e+04      2.520      0.013    1.06e+04    8.77e+04
==============================================================================
Omnibus:                       33.704   Durbin-Watson:                   2.040
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               48.909
Skew:                           1.293   Prob(JB):                     2.40e-11
Kurtosis:                       4.280   Cond. No.                         42.4
==============================================================================

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

Lastly, we will incorporate performance measures of both batsman and bowler in the same regression.

We will first use the original variables, total number of runs, number of not outs, number of balls faced, number of runs conceded, number of balls bowled, and number of wickets in the regression.

In [24]:
```
reg_IPL11=sm.ols(formula = 'Salary ~ runs+not_outs+balls_faced+runs_conceded+balls_bowled+wickets', data= IPLPlayer).fit()
print(reg_IPL11.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.408
Model:                            OLS   Adj. R-squared:                  0.382
Method:                 Least Squares   F-statistic:                     15.41
Date:                Mon, 26 Jul 2021   Prob (F-statistic):           2.20e-13
Time:                        00:22:03   Log-Likelihood:                -2036.6
No. Observations:                 141   AIC:                             4087.
Df Residuals:                     134   BIC:                             4108.
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept      1.458e+05   7.32e+04      1.993      0.048    1123.895    2.91e+05
runs           2089.9774   1629.015      1.283      0.202   -1131.930    5311.885
not_outs       5.95e+04   2.82e+04      2.110      0.037    3731.096    1.15e+05
balls_faced    -354.0166   2412.450     -0.147      0.884   -5125.422    4417.389
runs_conceded -1737.5721   1672.019     -1.039      0.301   -5044.535    1569.391
balls_bowled   5030.2797   2648.309      1.899      0.060    -207.614    1.03e+04
wickets       -2.231e+04   2.26e+04     -0.989      0.325   -6.69e+04    2.23e+04
==============================================================================
Omnibus:                       17.606   Durbin-Watson:                   2.049
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               23.546
Skew:                           0.709   Prob(JB):                     7.71e-06
```

```
Kurtosis:                       4.412   Cond. No.                         538.
==============================================================================

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

We will also use the modified batting average, batting strike, bowling average, and bowling strike variables to measure the player performance.

In [25]:
```
reg_IPL12=sm.ols(formula = 'Salary ~ batting_average+batting_strike+bowling_average+bowling_strike', data= IPLPlayer).fit()
print(reg_IPL12.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Salary   R-squared:                       0.308
Model:                            OLS   Adj. R-squared:                  0.288
Method:                 Least Squares   F-statistic:                     15.16
Date:                Mon, 26 Jul 2021   Prob (F-statistic):           2.85e-10
Time:                        00:22:04   Log-Likelihood:                -2047.6
No. Observations:                 141   AIC:                             4105.
Df Residuals:                     136   BIC:                             4120.
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept         1.37e+05   1.14e+05      1.204      0.231   -8.81e+04    3.62e+05
batting_average  2.418e+04   4617.989      5.237      0.000    1.51e+04    3.33e+04
batting_strike    -61.2223   1094.543     -0.056      0.955   -2225.747    2103.303
bowling_average -3.186e+04   1.08e+04     -2.951      0.004   -5.32e+04   -1.05e+04
bowling_strike   5.941e+04    1.7e+04      3.492      0.001    2.58e+04    9.31e+04
==============================================================================
Omnibus:                       13.575   Durbin-Watson:                   2.103
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               14.473
Skew:                           0.732   Prob(JB):                     0.000720
Kurtosis:                       3.565   Cond. No.                         324.
==============================================================================

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

Self Test

- Run a regression of salary as a function of the interaction of batsman and runs and the interaction of bowler and wickets taken.
- Interpret your regression results.

In [ ]:
```
```