

IPL\_MLR\_MODEL\_BUILDIN  
G

## The functional form for the MLR Model Building is given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots \dots \dots \beta_{18} X_{18} + \beta_{19} X_{19} + \beta_{20} X_{20}$$

Where

Y = SOLD PRICE

$X_1$  = AGE

$X_2$  = COUNTRY

$X_3$  = PLAYING ROLE

$X_4$  = T-RUNS

$X_5$  = T-WKTS

$X_6$  = ODI-RUNS-S

$X_7$  = ODI-SR-B

$X_8$  = ODI-WKTS

$X_9$  = ODI-SR-BL

$X_{10}$  = CAPTAINCY EXP

$X_{11}$  = RUNS-S

$X_{12}$  = HIGH SCORE

$X_{13}$  = AVE

$X_{14}$  = SR -B

$X_{15}$  = SIXERS

$X_{16}$  = RUNS-C

$X_{17}$  = WKTS

$X_{18}$  = AVE-BL

$X_{19}$  = ECON

$X_{20}$  = SR-BL

## Regression Model 1 Summary

SOLD PRICE =  $3.75e^5 - 53.7 * T-RUNS - 132.5 * T-WKTS + 57.9 * ODI-RUNS-S - 524.1 * ODI-SR-B +$   
 $815.3 * ODI-WKTS + \dots + 7.57e^4 * PLAYING$   
ROLE\_Batsman +  
 $1.54e^4 * PLAYING\_ROLE\_Bowler - 7.13e^4 * PLAYING\_ROLE\_W.\ Keeper +$   
 $1.64e^5 * CAPTAINCY\_EXP\_1$

### R-Square

The adjusted R-Square value for the Model is 0.362 which shows weak correlation between the Dependent and independent variables considered for the Model Building. The R-Square value of 0.362 indicates that only 36.2% of the variation in Sold Price of the players is explained by the parameters considered in Model Building

### P-Value:

The Model Output at a confidence level of 95%(i.e p-value<0.05) indicates that only the features “High Score”, “AGE\_2”, “AVE”, “COUNTRY\_ENGLAND” have come out significant in influencing “Sold Price”. However this is not very intuitive and could be a result of multi-collinearity effect of variables

# **Multi-Collinearity Analysis**

When the data has large no. of independent variables, it is possible that few of these variables may be highly correlated. The existence of a high correlation between independent variables is called Multi-Collinearity.

The following variables were highly correlated by the observations made from the Heat Map between variables having  $VIF > 4$

- T-RUNS and ODI-RUNS-S
- ODI-WKTS and T-WKTS
- RUNS-S and HIGH SCORE, AVE, SIXERS
- HIGH SCORE AND AVE, SIXERS
- RUNS-C AND WKTS
- AVE-BL AND ECON, SR-BL
- SR-BL AND ECON

Using Domain Knowledge the variables removed are T-RUNS, T-WKTS, RUNS-S, HIGH SCORE, AVE, RUNS-C, AVE-BL, ECON.

## **REGRESSION MODEL 2 SUMMARY**

$$\begin{aligned} \text{SOLD PRICE} = & 1.04e^5 * \text{COUNTRY\_PAK} - 3.91e^4 * \text{COUNTRY\_WI} + \\ & \dots\dots\dots + \\ & 1.26e^5 * \text{COUNTRY\_NZ} - 2.02e^4 * \text{PLAYING\_ROLE\_BOWLER} \end{aligned}$$

R-Square is 0.726 showing high correlation.

The p-value indicates the variables “COUNTRY\_IND”, “COUNTRY\_ENG”, “SIXERS” AND “CAPTAINCY EXP\_1” as the most significant variables at p-value<0.05.

## **REGRESSION MODEL 3 SUMMARY**

$$\begin{aligned} \text{SOLD PRICE} = & 3.87e^5 * \text{COUNTRY\_IND} + 7.31e^5 * \text{COUNTRY\_ENG} + 8637.83 * \text{SIXERS} + \\ & 3.57e^5 * \text{CAPTAINCY\_EXP\_1} \end{aligned}$$

R-Square is 0.704 showing high correlation.

# **RESIDUAL ANALYSIS**

## **Test for Normality**

The graph shows a plot between standardized residuals on x-axis and Normally distributed standardized residuals on y-axis. Since the graph is a linear plot at 45 degrees, it can be concluded that the important assumption of OLS that the residuals should be normally distributed is adhered to.

## **Test for Homoscedasticity**

The graph shows a plot between standardized predicted values on x-axis and standardized residuals on y-axis. As there is no specific pattern observed in the plot, it may be concluded that the assumption of Homoscedasticity is adhered to.