Regression

Simple Linear Regression

Assignment 3

BP: Salary\_hike -> Build a prediction model for Salary\_hike

1. After observing the plot between delivery time vs sorting time, the plot is in a linear relationship.
2. Now r value is = 0.978 and value of R2= 0.957
3. First well known procedure is doing RMSE ( root mean square error ) for predicting the values .

Now finding the confidence interval   
using , confint(reg,level=0.95)

at 2.5% -> 21136.061- X (8674.119)

At 95% -> 30448.34- x(10225.81)

and now we can get the predicted values in the interval using ,   
  
Predict ( reg , interval=”predict “)

fit lwr upr

1 36187.16 23698.92 48675.40

2 38077.15 25628.63 50525.67

3 39967.14 27556.52 52377.76

4 44692.12 32368.22 57016.03

5 46582.12 34289.64 58874.59

6 53197.09 40999.70 65394.48

7 54142.09 41956.37 66327.80

8 56032.08 43868.25 68195.91

9 56032.08 43868.25 68195.91

10 60757.06 48639.42 72874.70

11 62647.05 50544.46 74749.65

12 63592.05 51496.24 75687.86

13 63592.05 51496.24 75687.86

14 64537.05 52447.52 76626.57

15 68317.03 56247.70 80386.36

16 72097.02 60039.93 84154.10

17 73987.01 61933.05 86040.96

18 75877.00 63824.18 87929.82

19 81546.98 69485.57 93608.39

20 82491.97 70427.39 94556.56

21 90051.94 77944.06 102159.83

22 92886.93 80754.66 105019.20

23 100446.90 88228.15 112665.65

24 103281.89 91022.76 115541.02

25 108006.87 95670.98 120342.77

26 110841.86 98454.30 123229.42

27 115566.84 103084.00 128049.68

28 116511.84 104008.59 129015.09

29 123126.81 110468.27 135785.35

30 125016.80 112309.98 137723.63

Now we will try to transform the R2 values:

Now applying different methods,

The value before and after transformation are :

OUTPUT INPUT COR METHOD R2 RMSE

Salary Years’ Experience 0.978 SLR 0.957 5592.044

Salary log(Years Experience) 0.924 logarithmic 0.8539 10302.89

Log(salary) Years’Experiennce 0.9653 exponential 0.932 14762

salary sqrt(YearsExperience) 0.957 square root 0.9175 8019.74  
  
  
Therefor among all the four , considering for small change of value in r we can consider exponential method as the best transformation among all.  
  
  
At 95% = -30174.34+X(37713.66)

At 2.5 % = -7335.428+ X(47621.798)

The predicted values for Salary hike are:

fit lwr upr

1 25995.41 7764.924 44225.89

2 29893.81 11802.081 47985.54

3 33502.20 15529.691 51474.71

4 41586.40 23847.955 59324.84

5 44531.59 26866.789 62196.38

6 53905.64 36432.710 71378.57

7 55147.79 37695.349 72600.23

8 57571.47 40155.636 74987.30

9 57571.47 40155.636 74987.30

10 63318.13 45971.151 80665.11

11 65507.13 48179.689 82834.57

12 66580.57 49261.361 83899.79

13 66580.57 49261.361 83899.79

14 67640.68 50328.718 84952.65

15 71757.04 54464.939 89049.14

16 75694.16 58408.715 92979.61

17 77602.42 60315.844 94888.99

18 79473.61 62183.175 96764.04

19 84884.67 67567.738 102201.60

20 85759.28 68435.930 103082.63

21 92508.94 75116.177 109901.71

22 94936.80 77510.516 112363.09

23 101171.04 83638.201 118703.88

24 103426.90 85848.316 121005.48

25 107096.84 89435.788 124757.89

26 109248.30 91534.323 126962.29

27 112755.89 94948.434 130563.35

28 113446.24 95619.348 131273.13

29 118181.28 100212.006 136150.55

30 119504.36 101492.508 137516.22