**LABOUR COST PREDICTION USING MULTI LINEAR REGRESSION**

We are going to build a model to predict a labor cost using multi linear regression. We have a vehicle.csv file as a data source to build this model. This file contains 7 variables below is the below variable description:

**Vehicle No: It’s a code of Vehicle.**

**Fm: Functional machine number**

**Mileage: It tells about mileage of the vehicle**

**Lh : Labor hour**

**Lc : Labor cost**

**Mc : Maintenance cost of vehicle**

**State : State where it is produced**

**DS no: Distribution number.**

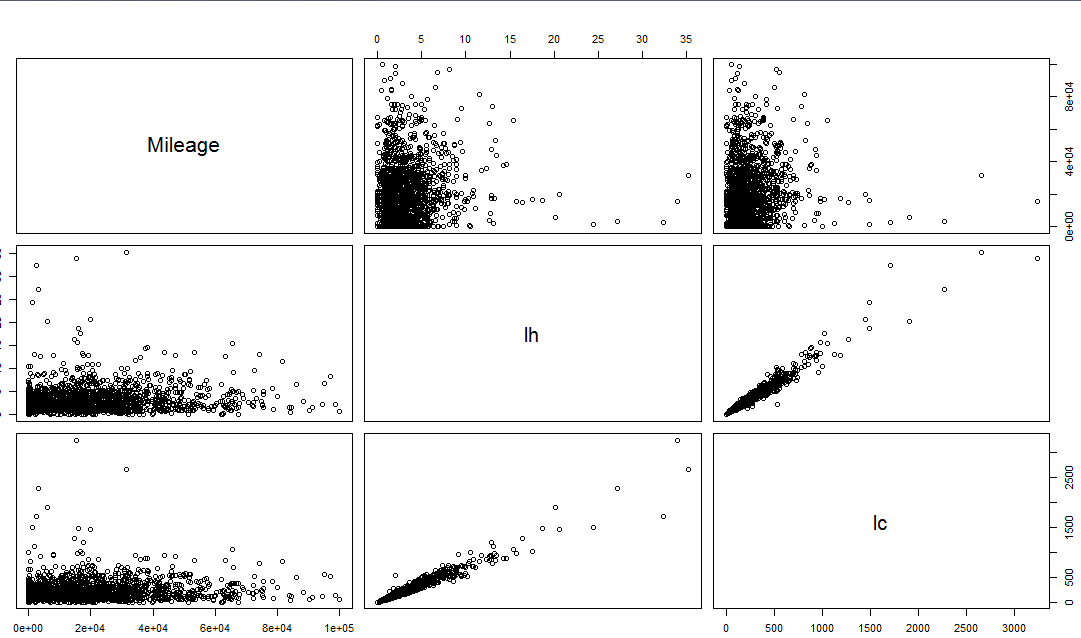
**PT no. Part No.**

**Equation :**

**LC = b0 + b1lh + b2mileage**

Now depending on weather labor hours and mileage have significant contribution on overall model ,if any of the variable do not have significant contribution on model they can be drop eg. Mileage and we may be left with a simple linear regression model where labor cost is dependent only on labor hour.

Now we will see scatter plot by using variable mileage, labor hour, labor cost to get an idea if any relation is there or no



In the above scatter plot we can clearly see that labor hour vs. labor cost are related to each other in stronger way rather than mileage vs. labor cost similarly mileage vs. labor hour.

**Multiple regression model:**

**results <- lm(lc~Mileage+lh,Vehicle)**

**After running the above code we got the below result:**

Coefficients:

(Intercept) Mileage lh

1.375e+00 -8.475e-05 7.355e+01

So our model will look like below:

Lc = 1.375

Mileage = -8.475

-8.475 times ten to the power of -5

-0.0000847

**Lc = 1.375 - -0.0000847mileage + 7.355lh**

**The above is the linear model we are fitting with the data.**

So now we will check more i.e we will check the summary of our model in R.

Call:

lm(formula = lc ~ Mileage + lh, data = Vehicle)

Residuals:

Min 1Q Median 3Q Max

-672.79 -14.73 -0.62 12.89 741.05

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.375e+00 2.218e+00 0.62 0.535

Mileage -8.475e-05 6.622e-05 -1.28 0.201

lh 7.355e+01 4.155e-01 177.01 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48.62 on 1621 degrees of freedom

Multiple R-squared: 0.951, Adjusted R-squared: 0.951

F-statistic: 1.574e+04 on 2 and 1621 DF, p-value: < 2.2e-16

In this summary we can see the P value where we can see that which variable plays a significant role in predicting labor cost. In this we can see that mileage is not playing significant role whereas labor hour do. So in further analysis we will drop this mileage variable.

Here P value is 0.001 from which we can derive a confidence level as below:

(1 – P) = 1 – 0.001

= 0.999 or 99.99 % Confidence.

This means if we treat labor hour as statistical significant variable then our confidence level is 99.99% and that’s the high level of confidence which we look for, whereas if we see the p value of mileage its 0.2 which after subtracting will be 0.8 i.e 80% which is less then 95% and this shows that its not a significant variable.

Here we can also see R-squared:0.951 which means variable in the model contribute 95% to the overall variability which is a very good number. Even though if we add more Variable in the model it will be for remaining 5 %.

After removing mileage variable we will get the final model :

**results <- lm(lc~lh,Vehicle)**

Coefficients:

(Intercept) lh

-0.2359 73.5088

**Conclusion and final model:**

**Lc =** -0.2359 + 73.5088 lh

**The above model is the significant model after removing the variables.**

**Comparing 2 models**

We fill compare 2 models i.e one with reduced variables and one including full variables by doing analysis of variance (ANOVA).

**reduced <- lm(lc~lh,Vehicle)**

**full <- lm(lc~Mileage+lh,Vehicle)**

**anova(reduced,full)**

Analysis of Variance Table

Model 1: lc ~ lh

Model 2: lc ~ Mileage + lh

Res.Df RSS Df Sum of Sq F Pr(>F)

1 1622 3835760

2 1621 3831889 1 3871 1.6376 0.2008

Above Pvalue which is 0.2 itself shows that adding mileage variable is not gaining any statistical significance.

**PREDICTION:**

**We want to predict laborcost from results for labor hours of 10 hours with the confidence interval prediction of 95%:**

**predict(results,data.frame(lh=10),interval = 'confidence')**

fit lwr upr

1 734.8522 728.9201 740.7844

Above we can see the labor cost :

Fit cost(Avg Cost) = $ 734.85

Low cost = $ 728.92

Upper cost = $ 740.

**Business Implications:**

**we can say that labor costing plays a crucial role in profitability of the company so predicting a labor cost helps business to take important decisions for future.**