**Emp\_data -> Build a prediction model for Churn\_out\_rate**

Output(Y) : Churn out rate

Input(X) : Salary hike

**R Code:**

**# Load the library**

library(readr)

**# Load the csv file and stored in object calories\_consumed**

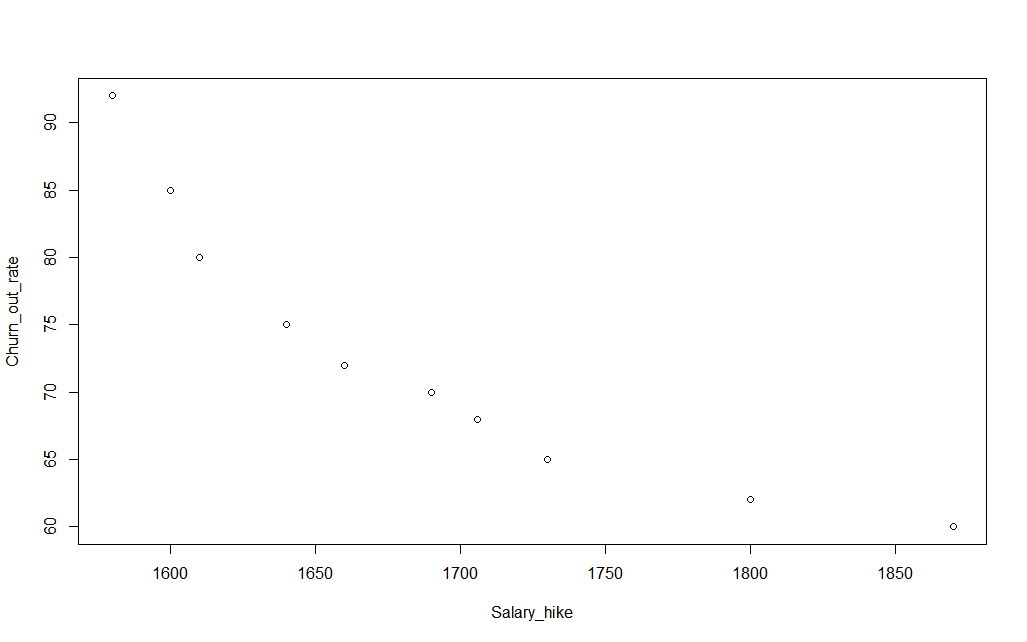
emp\_data <- read\_csv("D:/ALL Assignments/3.Simple Linear Regression/emp\_data.csv")

**# attach the object**

attach(emp\_data)

**# Draw scatter diagram**

plot(Salary\_hike,Churn\_out\_rate)



It tell following things:

I) Direction : negative correlation

II) Strength : Strong

III) Linearity :Linear relationship

**#Correlation coefficient r :**

cor(Salary\_hike,Churn\_out\_rate)

It give r = -0.9117

As |r| between > 0.85 => Strong strength

**#Linear regression technique and its summary**

emp\_data\_model <- lm(Churn\_out\_rate~Salary\_hike)

summary(emp\_data\_model)

It gives:

Call:

lm(formula = Churn\_out\_rate ~ Salary\_hike)

Residuals:

Min 1Q Median 3Q Max

-3.804 -3.059 -1.819 2.430 8.072

Coefficients:

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

(Intercept) 244.36491 27.35194 8.934 1.96e-05 \*\*\*

Salary\_hike -0.10154 0.01618 -6.277 0.000239 \*\*\*

---

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

Residual standard error: 4.469 on 8 degrees of freedom

Multiple R-squared: 0.8312, Adjusted R-squared: 0.8101

F-statistic: 39.4 on 1 and 8 DF, p-value: 0.0002386

As we are getting two and three stars(probability of getting wrong is less) and R-squared value is greater than 0.8.

Prediction model equation :

**Churn\_out\_rate = 244.36491 - 0.10154(Salary\_hike)**

**R-squared = 0.8312**

**Lets try some variants of regression for better R-squared value :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Regression Type** | **Y** | **X** | **r** | **R-squared** | **P-value** |
| Simple Linear | **Y** | **X** | **-0.911** | **0.8312** | 0.0002386 |
| Logarithm | **Y** | **Log(x)** | **-0.921** | **0.8486** | 0.0001532 |
| Exponential | **Log(Y)** | **x** | **-0.9346** | **0.8735** | 7.377e-05 |

The better model is given by exponential regression :

#exp

**plot(Salary\_hike,log(Churn\_out\_rate))**

**cor(Salary\_hike,log(Churn\_out\_rate))**

**emp\_data\_model\_3 <- lm(log(Churn\_out\_rate)~Salary\_hike)**

**summary(emp\_data\_model\_3)**

**Outtput:**

lm(formula = log(Churn\_out\_rate) ~ Salary\_hike)

Residuals:

Min 1Q Median 3Q Max

-0.04825 -0.03519 -0.01909 0.02942 0.08970

Coefficients:

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

(Intercept) 6.6383000 0.3175983 20.902 2.88e-08 \*\*\*

Salary\_hike -0.0013963 0.0001878 -7.434 7.38e-05 \*\*\*

---

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

Residual standard error: 0.0519 on 8 degrees of freedom

Multiple R-squared: 0.8735, Adjusted R-squared: 0.8577

F-statistic: 55.26 on 1 and 8 DF, p-value: 7.377e-05

**Prediction model equation :**

**Churn\_out\_rate = 6.638300 - 0.0013963(Salary\_hike)**

**R-squared = 0.8735**

**P value = 7.377e-05**

**Confidence estimation:**

**confint(emp\_data\_model\_3,level = 0.95)**

2.5 % 97.5 %

(Intercept) 5.905917079 7.3706828388

Salary\_hike -0.001829477 -0.0009631923

Lower Limit: churn out rate = exp(5.90 - 0.0018(salary hike))

Upper limit: churn out rate = exp(7.37 - 0.00096(salary hike))