**Salary\_Data: Build a prediction model for Salary\_hike**

Output (Y) : Salary

Input (X) : YearsExperience

**R Code:**

**# Load the library**

library(readr)

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

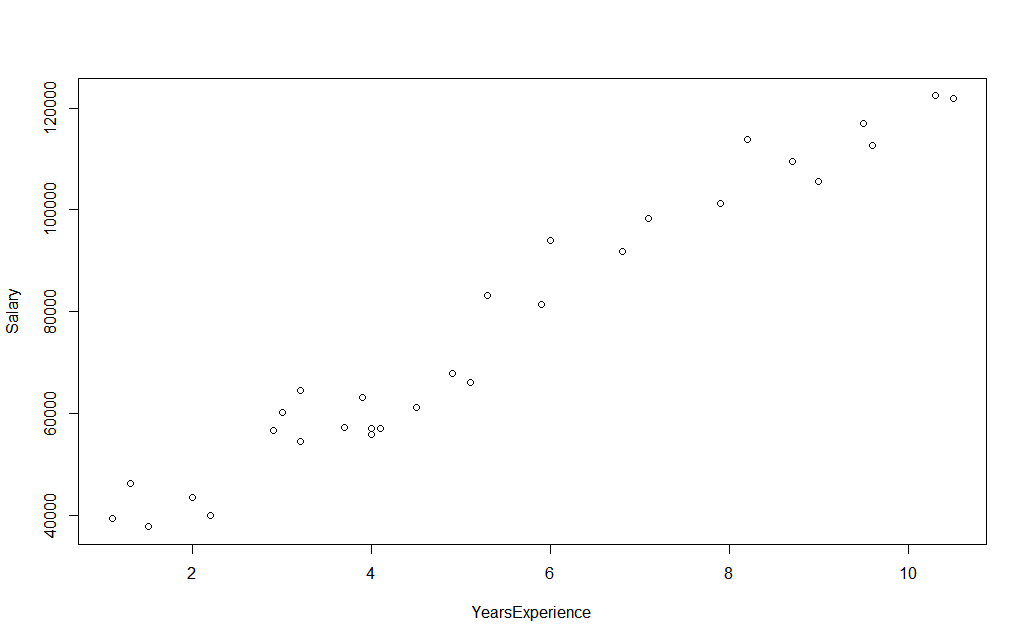
salary\_data <- read\_csv("D:/ALL Assignments/3.Simple Linear Regression/Salary\_Data.csv

")

**# attach the object**

attach(salary\_data)

**# Draw scatter diagram**

plot(YearsExperience,Salary)

It tell following things:

I) Direction : positive correlation

II) Strength : Strong

III) Linearity :Linear relationship

**#Correlation coefficient r :**

cor(YearsExperience,Salary)

It give r = 0.9782

As |r| between > 0.85 => Strong strength

**#Linear regression technique and its summary**

salary\_data\_model <- lm(Salary~YearsExperience)

summary(salary\_data\_model)

It gives:

Call:

lm(formula = Salary ~ YearsExperience)

**Residuals:**

**Min 1Q Median 3Q Max**

**-7958.0 -4088.5 -459.9 3372.6 11448.0**

Coefficients:

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

(Intercept) 25792.2 2273.1 11.35 5.51e-12 **\*\*\***

YearsExperience 9450.0 378.8 24.95 < 2e-16 **\*\*\***

---

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

Residual standard error: 5788 on 28 degrees of freedom

Multiple **R-squared:**  **0.957**, Adjusted R-squared: 0.9554

F-statistic: 622.5 on 1 and 28 DF, **p-value: < 2.2e-16**

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 :

**Salary = 25792.2 + 9450(YearExp)**

**R-squared = 0.957**

**Confidence estimate:**

**confint(salary\_data\_model,level = 0.95)**

2.5 % 97.5 %

(Intercept) 21136.061 30448.34

YearsExperience 8674.119 10225.81

Lower Limit: Salary = 21136.061 + 8674(YearExp)

Upper Limit: Salary = 30448.34 + 10225.81(YearExp)