* **Delivery\_time -> Predict delivery time using sorting time**

> delivery\_time <- read.csv("delivery\_time.csv")

> summary(delivery\_time)

Delivery Time Sorting Time

Min. : 8.00 Min. : 2.00

1st Qu.:13.50 1st Qu.: 4.00

Median :17.83 Median : 6.00

Mean :16.79 Mean : 6.19

3rd Qu.:19.75 3rd Qu.: 8.00

Max. :29.00 Max. :10.00

> delivery\_time\_model <- lm(delivery\_time$`Delivery Time` ~ delivery\_time$`Sorting Time`, delivery\_time)

> summary(delivery\_time\_model)

Call:

lm(formula = delivery\_time$`Delivery Time` ~ delivery\_time$`Sorting Time`,

data = delivery\_time)

Residuals:

Min 1Q Median 3Q Max

-5.1729 -2.0298 -0.0298 0.8741 6.6722

Coefficients:

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

(Intercept) 6.5827 1.7217 3.823 0.00115 \*\*

delivery\_time$`Sorting Time` 1.6490 0.2582 6.387 3.98e-06 \*\*\*

---

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

Residual standard error: 2.935 on 19 degrees of freedom

Multiple R-squared: 0.6823, Adjusted R-squared: 0.6655

F-statistic: 40.8 on 1 and 19 DF, p-value: 3.983e-06

> plot(delivery\_time\_model)









p-value: 3.983e-06(significant), Multiple R-squared: 0.6823(Insignificant)

**Observation**:

By observing above model plots, the data points 5,9,21 are outliers. So further improve the model predictions capability we will discard those data points from the model and test it again.

> delivery\_time\_model1 <- lm(delivery\_time$`Delivery Time` ~ delivery\_time$`Sorting Time` , data = delivery\_time[c(-5,-9,-21),])

> summary(delivery\_time\_model1)

Call:

lm(formula = delivery\_time$`Delivery Time` ~ delivery\_time$`Sorting Time`,

data = delivery\_time[c(-5, -9, -21), ])

Residuals:

Min 1Q Median 3Q Max

-2.3407 -1.5027 0.2275 0.9328 3.6815

Coefficients:

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

(Intercept) 6.0240 1.1751 5.126 0.000102 \*\*\*

Sorting.Time 1.6741 0.1872 8.941 1.27e-07 \*\*\*

---

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

Residual standard error: 1.839 on 16 degrees of freedom

Multiple R-squared: 0.8332, Adjusted R-squared: 0.8228

F-statistic: 79.94 on 1 and 16 DF, p-value: 1.273e-07







**Conclusion**: As P value is 1.273e-07, which is <0.05, x variable is significant, also the R-squared value is 0.8332, which is >0.85 it shows strong linear relationship. Which means model predicts 83.32% output correct.