2) Delivery\_time -> Predict delivery time using sorting time

**Ans :**

**R Code :**

## Simple Linear Regression

########## Delivery Time Data Set #########

delivery <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Simple Linear Regression\\delivery\_time.csv')

DT <- delivery$Delivery.Time

ST <- delivery$Sorting.Time

boxplot(DT, col="dodgerblue4")

boxplot(ST,col="dodgerblue4")

plot(ST,DT,main="Scatter Plot",

col="Dodgerblue4",

col.main="Dodgerblue4",

col.lab="Dodgerblue4",

xlab="Delivery Time",

ylab="Sorting Time", pch=20)

reg.model<-lm(DT~ST, data=delivery)

summary(reg.model)

plot(ST,DT,main="Line of Best Fit",col="Dodgerblue4", col.main="Dodgerblue4")

abline(reg.model, col="red")

res <- signif(residuals(reg.model))

pre <- predict(reg.model)

segments(ST, DT, ST, pre)

predict(reg.model,newdata= data.frame(ST=c(15,20)))

###### Squared Data Transformation ###########

STsqr <- ST \* ST

deliverysqr <- cbind(delivery,STsqr)

reg.modelsqr<-lm(DT~STsqr, data=deliverysqr)

summary(reg.modelsqr)

plot(STsqr,DT,main="Line of Best Fit",col="Dodgerblue4", col.main="Dodgerblue4")

abline(reg.modelsqr, col="red")

res <- signif(residuals(reg.modelsqr))

pre <- predict(reg.modelsqr)

segments(STsqr, DT, STsqr, pre)

###### Square-Root Data Transformation ###########

STsqrt <- sqrt(ST)

deliverysqrt <- cbind(delivery,STsqrt)

reg.modelsqrt<-lm(DT~STsqrt, data=deliverysqrt)

summary(reg.modelsqrt)

plot(STsqrt,DT,main="Line of Best Fit",col="Dodgerblue4", col.main="Dodgerblue4")

abline(reg.modelsqrt, col="red")

res <- signif(residuals(reg.modelsqrt))

pre <- predict(reg.modelsqrt)

segments(STsqrt, DT, STsqrt, pre)

###### Log Data Transformation ###########

STlog <- log(ST)

deliverylog <- cbind(delivery,STlog)

reg.model.log<-lm(DT~STlog, data=deliverylog)

summary(reg.model.log)

plot(STlog,DT,main="Line of Best Fit",col="Dodgerblue4", col.main="Dodgerblue4")

abline(reg.model.log, col="red")

res <- signif(residuals(reg.model.log))

pre <- predict(reg.model.log)

segments(STlog, DT, STlog, pre)

**Results :**

> reg.model<-lm(DT~ST, data=delivery)

> summary(reg.model)

Call:

lm(formula = DT ~ ST, data = delivery)

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 \*\*

ST 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

> predict(reg.model,newdata= data.frame(ST=c(15,20)))

1 2

31.31803 39.56313

> reg.modelsqr<-lm(DT~STsqr, data=deliverysqr)

> summary(reg.modelsqr)

Call:

lm(formula = DT ~ STsqr, data = deliverysqr)

Residuals:

Min 1Q Median 3Q Max

-5.824 -2.485 -0.331 1.644 7.141

Coefficients:

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

(Intercept) 11.23721 1.19563 9.399 1.42e-08 \*\*\*

STsqr 0.12487 0.02194 5.691 1.74e-05 \*\*\*

---

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

Residual standard error: 3.166 on 19 degrees of freedom

Multiple R-squared: 0.6303, Adjusted R-squared: 0.6108

F-statistic: 32.39 on 1 and 19 DF, p-value: 1.739e-05

> reg.modelsqrt<-lm(DT~STsqrt, data=deliverysqrt)

> summary(reg.modelsqrt)

Call:

lm(formula = DT ~ STsqrt, data = deliverysqrt)

Residuals:

Min 1Q Median 3Q Max

-4.6789 -1.7277 -0.3694 0.8023 6.4211

Coefficients:

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

(Intercept) -2.519 2.995 -0.841 0.411

STsqrt 7.937 1.204 6.592 2.61e-06 \*\*\*

---

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

Residual standard error: 2.872 on 19 degrees of freedom

Multiple R-squared: 0.6958, Adjusted R-squared: 0.6798

F-statistic: 43.46 on 1 and 19 DF, p-value: 2.611e-06

> reg.model.log<-lm(DT~STlog, data=deliverylog)

> summary(reg.model.log)

Call:

lm(formula = DT ~ STlog, data = deliverylog)

Residuals:

Min 1Q Median 3Q Max

-4.0829 -2.0133 -0.1965 0.9351 7.0171

Coefficients:

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

(Intercept) 1.160 2.455 0.472 0.642

STlog 9.043 1.373 6.587 2.64e-06 \*\*\*

---

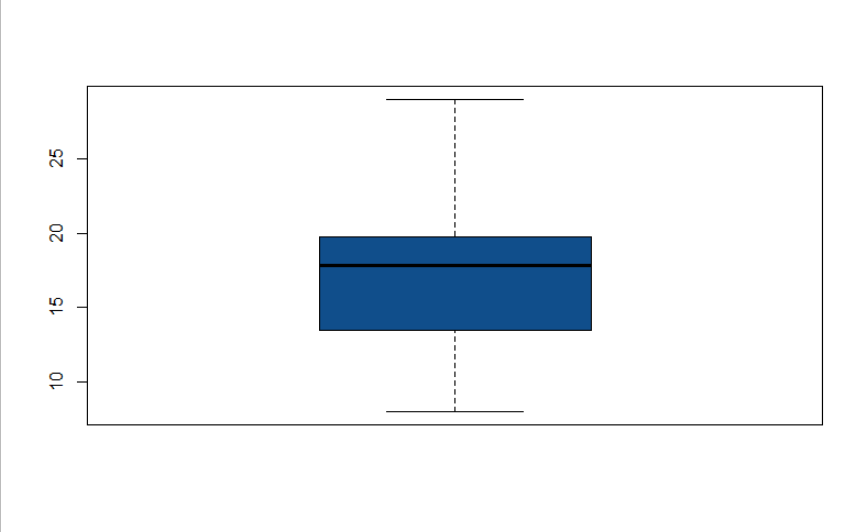
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

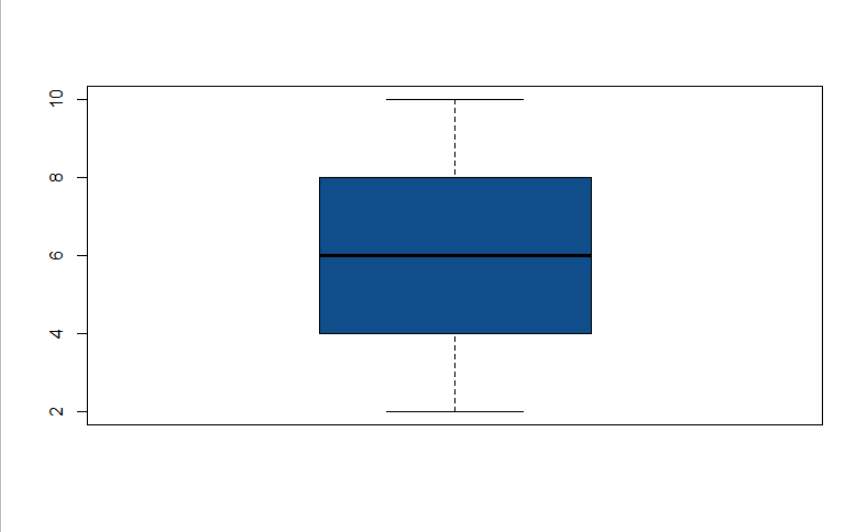
Residual standard error: 2.873 on 19 degrees of freedom

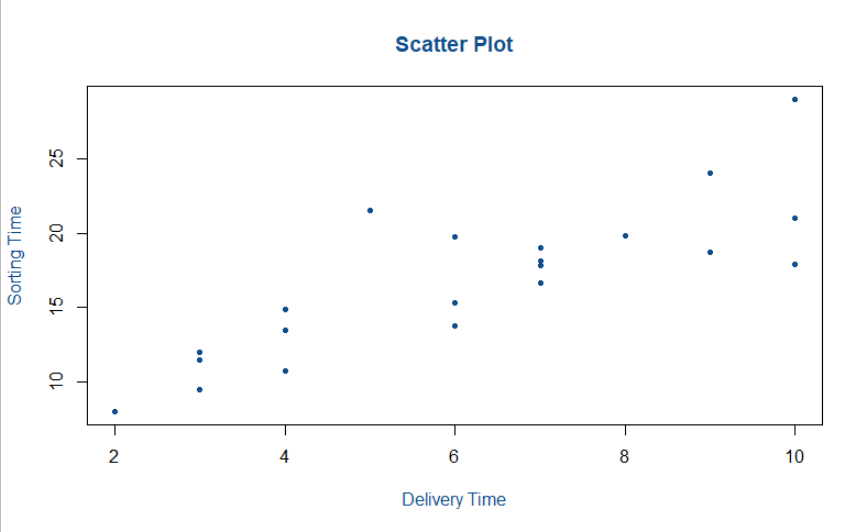
Multiple R-squared: 0.6954, Adjusted R-squared: 0.6794

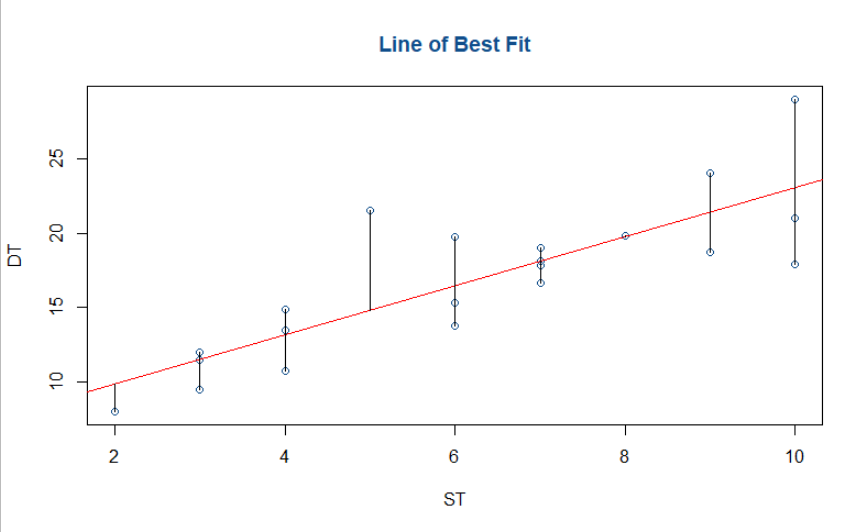
F-statistic: 43.39 on 1 and 19 DF, p-value: 2.642e-06

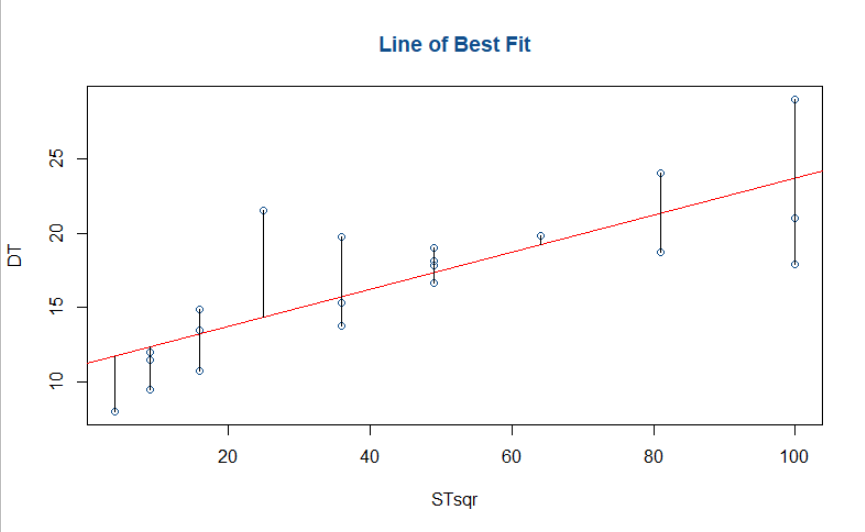
**Plots :**

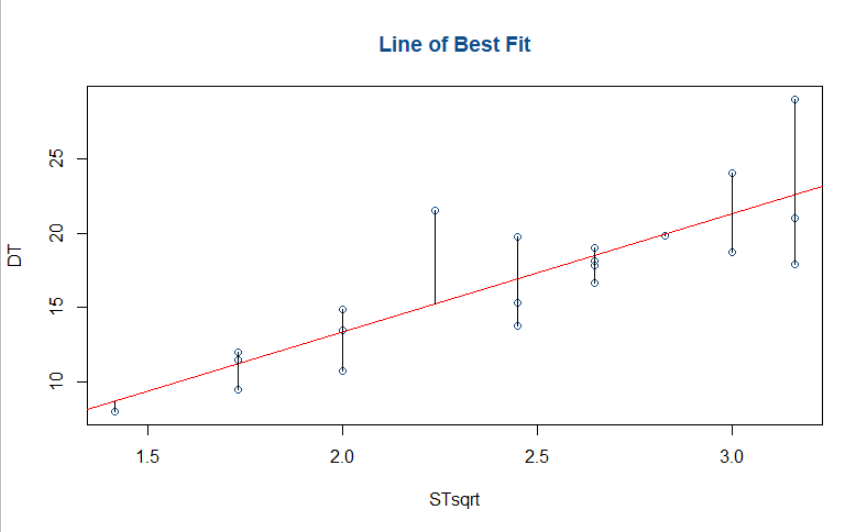


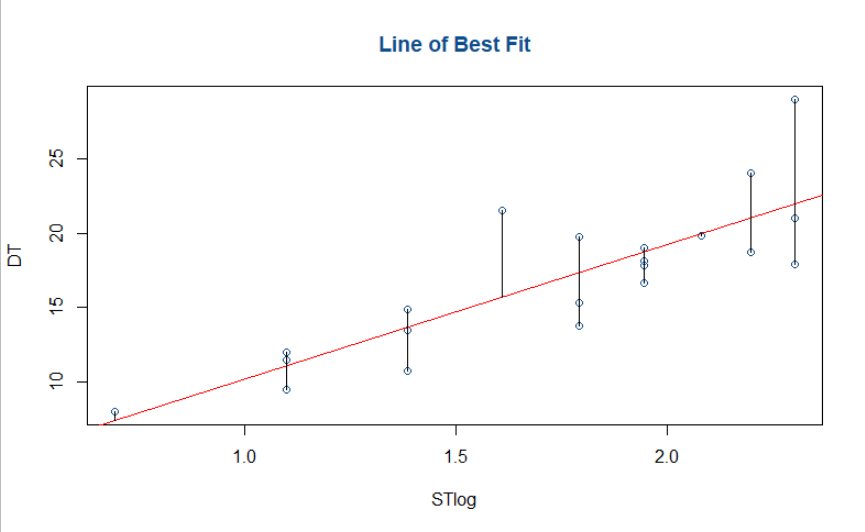












**Inference :**

Performed regression models using various data transformations.

The best results we are getting in Square Root Data Transformation.

|  |  |  |
| --- | --- | --- |
| **Data Transformation** | **Multiple R-squared** | **Adjusted R-squared** |
| Normal | 0.6823 | 0.6655 |
| Squared | 0.6303 | 0.6108 |
| Square Root | 0.6958 | 0.6798 |
| Log | 0.6954 | 0.6794 |