> data<-read.table("C:/Users/Kathie/Desktop/Last Semester 2020/Applied Predictive Analytics/APCuv.txt", header = TRUE)

> set.seed(19)

> R=runif(nrow(data))

> data$R=R

> data$type<-ifelse(data$R<=.6,"train","test")

> train<-data[data$type=="train",]

> test<-data[data$type=="test",]

> regmodel<-lm(Log10APC~UV,data=train)

> summary(regmodel)

Call:

lm(formula = Log10APC ~ UV, data = train)

Residuals:

Min 1Q Median 3Q Max

-1.1662 -0.4515 -0.1176 0.4161 1.8021

Coefficients:

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

(Intercept) 1.4727 0.2506 5.877 3.01e-07 \*\*\*

UV 9.7026 1.5563 6.234 8.20e-08 \*\*\*

---

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

Residual standard error: 0.7049 on 52 degrees of freedom

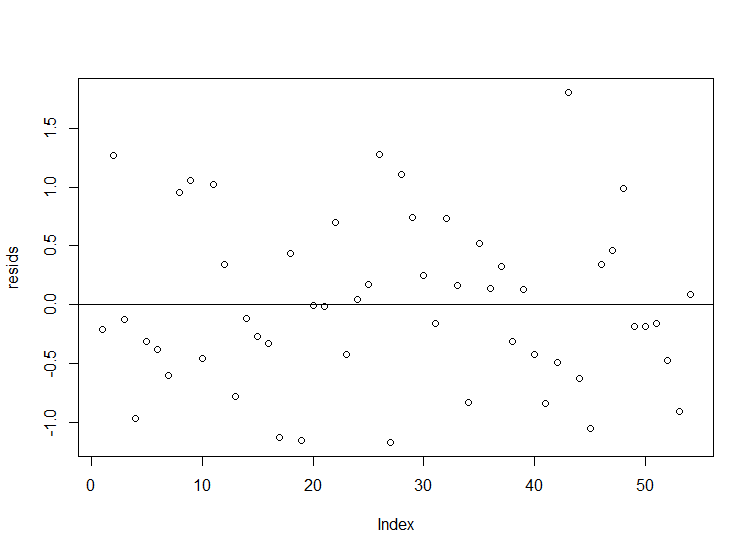
Multiple R-squared: 0.4277, Adjusted R-squared: 0.4167

F-statistic: 38.87 on 1 and 52 DF, p-value: 8.196e-08

> resids<-resid(regmodel)

> plot(resids)

> abline(0,0)



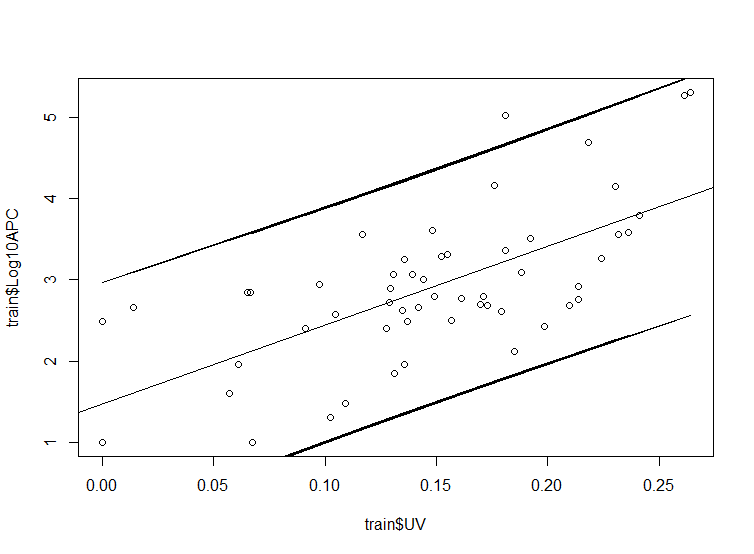
> plot(train$UV,train$Log10APC)

> abline(regmodel)

> pred\_interval<-predict(regmodel,newdata = train,interval = "prediction",level=0.95)

> lines(train$UV,pred\_interval[,2])

> lines(train$UV,pred\_interval[,3])



> predicted<-predict.lm(regmodel,newdata = test)

> mape<-predict.lm(regmodel,newdata = test)

> mape<-100\*mean(abs(test$Log10APC-predicted)/test$Log10APC)

> mape

[1] 16.07458

**Discussions**

Based on the residuals summary we can observe that the coefficient and intercept are significant. The p-value of the F-test shows that the regression model is significant as well. We can estimate that that for UV = 0.25, LogAPC will be:

logAPC estimate = 1.4727 + 9.7026\*(0.20)

logAPC estimate = 3.41322

using a confidence interval of 95% we can say we can expect that 95% of 0.20 UV will yield between around 1.5 and 4.5 LogAPC(based on graph).

Based on the residual plot we can identify a random pattern. This random pattern indicates that a linear model can provide a good fit to the data. The average of absolute percent error values tells that we can expect to miss the true value by 16% when we use our model.