Communicatin results project 2

Matthew Hurworth

2022-05-18

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
dataCR<-read.csv("Ass2data.csv")
attach(dataCR)

dataCR<-na.omit(dataCR)
```

```
set.seed(100)
trainCR<- dataCR[1:165,]
testCR<- dataCR[166:200,]
```

```
plot(dataCR)

corplot<- cor(dataCR)

library(car)

library(corrplot)

corrplot(corplot, method="square")
```

```
ggplot(data=dataCR, mapping = aes(x=C1, y = Y)) + geom_point(colour="blue") + geom_point(aes(x=C4, y=Y)
```

```
m1<- lm(Y ~ C1+C2+C3+C4+C5+C6, data=trainCR)
summary(m1)
plot(m1)
```

##step model

```
Final_step_model<- step(m1)

Final_step_model
```

##Manual model

```
m1<- lm(Y ~ C1+C2+C3+C4+C5+C6, data=trainCR)
summary(m1)
plot(m1)
```

```
m2<- lm(Y ~ C1+C3+C4+C5+C6, data=trainCR)
summary(m2)
```

```
m3<- lm(Y ~ C1+C3+C4+C5, data=trainCR)
summary(m3)
```

```
m4<- lm(Y ~ C1+C4+C5, data=trainCR)
summary(m4) ##Final model - same as the automodel
```

add 1 model which is in predictive modelling two

```
mii<- lm(Y ~ C1+C4+C6, data=trainCR)
summary(mii) ##Final model - same as the automodel
```

```
mi4<- lm(Y ~ C1+C4, data=trainCR)
summary(mi4)

AIC(mi4)
AIC(m4)
AIC(mii)
```

```
Anova(m4, mii)
Anova(mii, m4)

vif(m4)
```

```
par(mfrow=c(2,2))
plot(m4)
```

```
addm1<- glm(Y ~ 1, data=trainCR)
summary(addm1)

add1(addm1, scope=trainCR, test="Chisq")
```

```
addm2<- glm(Y~C1, data=trainCR)
add1(addm2, scope=trainCR, test="Chisq")
```

```
addm3<- glm(Y~C1 + C4, data=trainCR)
add1(addm3, scope=trainCR, test="Chisq")
```

```
addm4<- glm(Y~C1 + C4 + C5, data=trainCR)  ##Final model using add1 also produces the same model as man
add1(addm4, scope=trainCR, test="Chisq")
```

##Model validation

```r
1-(addm4$deviance / addm4$null.deviance)
```

```r
###RMSE
RMSE<-function(predicted, true)mean((predicted-true)^2)^.5

testCR<-na.omit(testCR)

RMSE(predict(m4,testCR),testCR$Y)
RMSE(predict(mii,testCR),testCR$Y)

###Predicted vs Actual
par(mfrow=c(2,2))

p1<-plot(x=predict(m4), y=Y)+ title((main="Model 4 - Prediction vs Actual for Y"),cex.main=1.5, cex.lab=
abline(0,1, col="red")

###Correlation tests
cor(testCR["Y"],predict(m4,testCR))

###VIF

vif(m4)
```

```r
ggplot(m4, aes(x=predict(m4), y=Y)) +
  geom_point() +
  geom_abline(intercept=0, slope=1)
```

```r
##Histogram plots of data variables to analyse spread
c1<-ggplot(data=data, mapping = aes(C1))+geom_histogram(binwidth = 1.5)
c2<-ggplot(data=data, mapping = aes(C2))+geom_histogram(binwidth = 1)
c3<-ggplot(data=data, mapping = aes(C3))+geom_histogram(binwidth = 1.5)
c4<-ggplot(data=data, mapping = aes(C4))+geom_histogram(binwidth = 1.5)
c5<-ggplot(data=data, mapping = aes(C5))+geom_histogram(binwidth = 1.5)
c6<-ggplot(data=data, mapping = aes(C6))+geom_histogram(binwidth = 0.5)
Y<-ggplot(data=data, mapping = aes(Y))+geom_histogram(binwidth = 15)

grid.arrange(c1,c2,c3,c4,c5,c6,Y)
```

```r
##Scatter plots of variables against Y to look at correlations
C1<-ggplot(data=data, mapping=aes(x=Y, y=C1))+geom_point()+ geom_smooth(se=FALSE, col="red")
C2<-ggplot(data=data, mapping=aes(x=Y, y=C2))+geom_point()+ geom_smooth(se=FALSE, col="red")
C3<-ggplot(data=data, mapping=aes(x=Y, y=C3))+geom_point()+ geom_smooth(se=FALSE, col="red")
C4<-ggplot(data=data, mapping=aes(x=Y, y=C4))+geom_point()+ geom_smooth(se=FALSE, col="red")
C5<-ggplot(data=data, mapping=aes(x=Y, y=C5))+geom_point()+ geom_smooth(se=FALSE, col="red")
C6<-ggplot(data=data, mapping=aes(x=Y, y=C6))+geom_point()+ geom_smooth(se=FALSE, col="red")
Yi<-ggplot(data=data, mapping=aes(x=Y, y=Y))+geom_point()+ geom_smooth(se=FALSE, col="red")

grid.arrange(C1,C2,C3,C4,C5,C6, Yi)
```