DMDA(Week - 8 to 12)

Week 8

a) Write R program to find R-Mean, Median & Mode with the sample data.

```
get_mode <- function(v) {
  v <- na.omit(v)

  uniq <- unique(v)
  uniq[which.max(tabulate(match(v, uniq)))]
}

x <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 9, NA)

cat("Mean for x: ", mean(x, na.rm = TRUE), "\n")
cat("Mode for x: ", get_mode(x), "\n")
cat("Median for x: ", median(x, na.rm = TRUE), "\n")
```

b) Write R program to find Analysis and Covariance with the sample data and visualize the regression graphically.

```
group <- factor(rep(c("A","B","C"),each=5))
score <- c(12,14,15,13,12,18,15,30,1,5,17,22,31,25,15)

model <- aov(score ~ group)
summary(model)

age <- 20:35
treatment <- factor(rep(c("A","B"),each=8))
score <- c(60,65,73,75,78,79,80,65,60,88,63,55,66,77,88,99)

model <- lm(score ~ age + treatment)
anova(model)
```

Week 9

a) Linear Regression

```r
x <- c(151,174,186,134,178,887,654,423,432,444)
y <- c(63,81,83,63,53,66,58,72,73,88)
relation <- lm(y~x)
print(relation)


x <- c(151,174,186,134,178,887,654,423,432,444)
y <- c(63,81,83,63,53,66,58,72,73,88)
relation <- lm(y~x)
print(summary(relation))


x <- c(151,174,186,134,178,887,654,423,432,444)
y <- c(63,81,83,63,53,66,58,72,73,88)
relation <- lm(y~x)
a <- data.frame(x=170)
result <- predict(relation,a)
print(result)


x <- c(151,174,186,134,178,887,654,423,432,444)
y <- c(63,81,83,63,53,66,58,72,73,88)
relation <- lm(y~x)
png(file = "Linear.png")
plot(y,x,col="blue",main = "height & Weight", abline(lm(x~y)), cex=1.3, pch=16 ,xlab="weight", ylab="height")
dev.off()
```

b) Multiple Regression

```
input <- mtcars[ ,c("mpg","disp","hp","wt")]
print(head(input))


input <- mtcars[ ,c("mpg","disp","hp","wt")]
model <- lm(mpg~disp+hp+wt , data = input)
print(model)


cat("Coeff")


xmpg <- coef(model)[1]
xdisp <- coef(model)[2]
xhp <- coef(model)[3]
xwt <- coef(model)[4]


print(xmpg)
print(xdisp)
print(xhp)
print(xwt)
```

c) Logistic Regression

```
claimants <- read.csv("C:/Users/rajan/Downloads/claimants.csv")


sum(is.na(claimants))

claimants <- na.omit(claimants)


logit <- glm(ATTORNEY ~ factor(CLMSEX) + factor(CLMINSUR) + factor(SEATBELT) + CLMAGE +
LOSS,family = "binomial", data = claimants)

summary(logit)


prob <- predict(logit, type = "response", claimants)


confusion <- table(prob > 0.5, claimants$ATTORNEY)

confusion

accuracy <- sum(diag(confusion)) / sum(confusion)

accuracy


install.packages("ROCR", dependencies = TRUE)

install.packages("pROC", dependencies = TRUE)

library(ROCR)

library(pROC)


rocrpred <- Prediction(prob, claimants$ATTORNEY)

rocrperf <- performance(rocrpred, 'tpr', 'fpr')

plot(rocrperf, colorize = TRUE, text.adj = c(-0.2, 1.7))


auc_value <- auc(claimants$ATTORNEY ~ prob)

auc_value
```

d) Poisson Regression

```
input <- warpbreaks
print(head(input))


output <- glm(formula = breaks ~ wool + tension , data = warpbreaks , family = poisson)


summary(output)
```

Week 10

a) Time Series Analysis with the sample data and visualize the regression graphically.

```
rainfall1 <- c(176,827,483.93,879,947.2,746.2,947.4,383.2,586.43,758.3,753.4,211.1)

rainfall2 <- c(176,827,483.93,879,947.2,746.2,947.4,383.2,586.43,758.3,753.4,211.1)


combined_rainfall <- matrix(c(rainfall1,rainfall2),nrow=12)


rainfall.timeseries <- ts(combined_rainfall,start = c(2012,1), frequency=12)


print(rainfall.timeseries)


png(file = "rainfall.png")


plot(rainfall.timeseries)


dev.off()
```

b) Non Linear Least Square with the sample data and visualize the regression graphically.

```
xvalues <- c(64.1,48.3,94,92,75.2,47,27,21,83,65.2)

yvalues <- c(7.3,2.4,3.2,5.6,6.4,6.4,2.3,6.9,0.4,3.9)

png(file = "xandy.png")

plot(xvalues,yvalues)


model <- nls(yvalues~b1*xvalues^2+b2,start=list(b1=1,b2=3))

new.data <- data.frame(xvalues=seq(min(xvalues),max(xvalues),len=100))


lines(new.data$xvalues,predict(model,newdata = new.data))

dev.off()


print(sum(resid(model)^2))

print(confint(model))
```

c) Decision Tree with the sample data and visualize the regression graphically.

```
install.packages("party")

library(party)


print(head(readingSkills))

library(party)


input.data <- readingSkills[c(1:105),]

png(file = "party.png")


output.tree <- ctree(nativeSpeaker ~ age + shoeSize + score, data=input.data)

plot(output.tree)

dev.off()
```

Week 11

a) Normal Distribution

```r
x <- seq(-10,10,by=0.1)
y <- dnorm(x,mean=3.5,sd=1.5)
png(file = "dnorm.png")
plot(x,y)
dev.off()


x <- seq(-10,10,by=0.2)
y <- pnorm(x,mean=2.5,sd=2)
png(file = "pnorm.png")
plot(x,y)
dev.off()


x <- seq(0,1,by=0.02)
y <- qnorm(x,mean=2,sd=1)
png(file = "qnorm.png")
plot(x,y)
dev.off()


y <- rnorm(50)
png(file="rnorm.png")
hist(y)
dev.off()
```

b) Binomial Distribution

```
x <- seq(0,50,by=1)

y <- dbinom(x,50,0.5)

png(file = "dbinom.png")

plot(x,y)

dev.off()


x <- pbinom(26,51,0.5)

print(x)

x <- qbinom(0.25,51,0.5)

print(x)

x <- rbinom(8,150,0.4)

print(x)
```

Week 12

a) χ2-test

```r
setwd(tempdir())
observed <- c(50,30,20)
expected <- c(40,40,20)

chi_results <- chisq.test(observed,p=expected/sum(expected))
print(chi_results)
dev.new()

barplot(rbind(observed,expected),
    beside=TRUE,
    col=c("skyblue","orange"),
    names.arg=c("A","B","C"),
    main="Chi-squared test",
    ylab="Frequency")

legend("topright",legend = c("obserbed","expected"),fill = c("skyblue","orange"))
```

b) t-test

```r
marks <- c(65,70,75,80,85,90,95,60,55,50)
mu <- 70

result <- t.test(marks,mu=mu)
print(result)

hist(marks,
    main="T-test",
    Xlab="Marks",
    col="lightblue",
    border="black")
```

```
abline(v=mean(marks),col="red",lwd=2)

text(mean(marks),2,paste("Mean:",round(mean(marks),2)),pos=4)
```

b) F-test

```
group1 <- c(12,13,14,15,16,17)

group2 <- c(18,19,20,21,22,23)

result <- var.test(group1,group2)

print(result)


boxplot(group1,group2,

    names=c("group1","group2"),

    main="F-test",

    col=c("lightblue","lightgreen"),

    ylab="values")


abline(h=mean(group1),col="blue",lty=2)

abline(h=mean(group2),col="green",lty=2)
```