Week3 Tutorial exercise

Question 01: Multiple Linear Regression

(a) Upload the data “Advertising.csv” and explore it

```
advertising <-read.csv("Advertising.csv")
attach(advertising)
names(advertising)
dim(advertising)
```

(b) Find the Covariance and Correlation Matrix of Sales, TV, Radio and Newspaper.

```
mod=lm(Sales~TV+Radio+Newspaper)
cov(advertising,method="pearson") #covariance
```

```
> cov(advertising,method="pearson")
                    TV     Radio Newspaper     Sales
TV          7370.94989  69.86249 105.91945 350.39019
Radio         69.86249 220.42774 114.49698  44.63569
Newspaper    105.91945 114.49698 474.30833  25.94139
Sales        350.39019  44.63569  25.94139  27.22185
```

```
cor(advertising,method="pearson") #correlation
```

```
> cor(advertising,method="pearson")
                  TV      Radio  Newspaper     Sales
TV        1.00000000 0.05480866 0.05664787 0.7822244
Radio     0.05480866 1.00000000 0.35410375 0.5762226
Newspaper 0.05664787 0.35410375 1.00000000 0.2282990
Sales     0.78222442 0.57622257 0.22829903 1.0000000
```

(c) Construct the multiple linear regression model and find the least square estimates of the model parameters.

```
mod=lm(Sales~TV+Radio+Newspaper)
```

```
> summary(mod)

Call:
lm(formula = Sales ~ TV + Radio + Newspaper)

Residuals:
    Min      1Q  Median      3Q     Max
-8.8277 -0.8908  0.2418  1.1893  2.8292

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.938889   0.311908   9.422   <2e-16 ***
TV           0.045765   0.001395  32.809   <2e-16 ***
Radio        0.188530   0.008611  21.893   <2e-16 ***
Newspaper   -0.001037   0.005871  -0.177     0.86
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.686 on 196 degrees of freedom
Multiple R-squared:  0.8972,    Adjusted R-squared:  0.8956
F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

(d) Test the significance of the parameters and find the resulting model to model Sales in terms of advertising modes, TV, Radio and Newspaper.
mod=lm(Sales~TV+Radio+Newspaper)

```
> confint(mod)
                 2.5 %     97.5 %
(Intercept)  2.32376228 3.55401646
TV           0.04301371 0.04851558
Radio        0.17154745 0.20551259
Newspaper   -0.01261595 0.01054097
```

TV and Radio p-values <0.05 but NewPapaer >0.05
so linear relationship between Sales and TV, and Radio are significant
Sales and NewsPaper are not linealy related

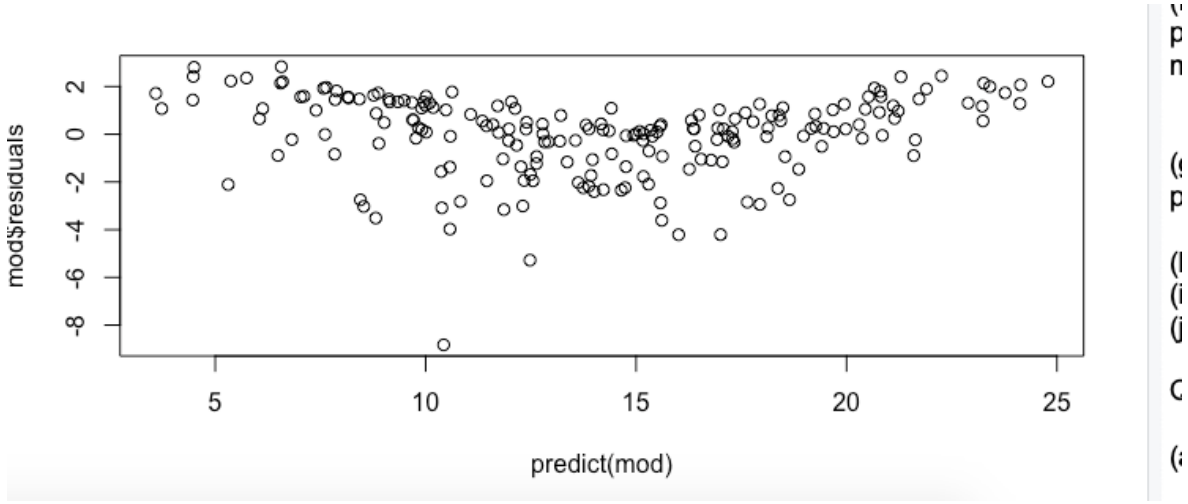
```
> cor(Sales,Newspaper)
[1] 0.228299
> cor(Sales,Radio)
[1] 0.5762226
> cor(Sales,TV)
[1] 0.7822244
>
```

(e) Assess the overall accuracy of the model.
anova(mod)

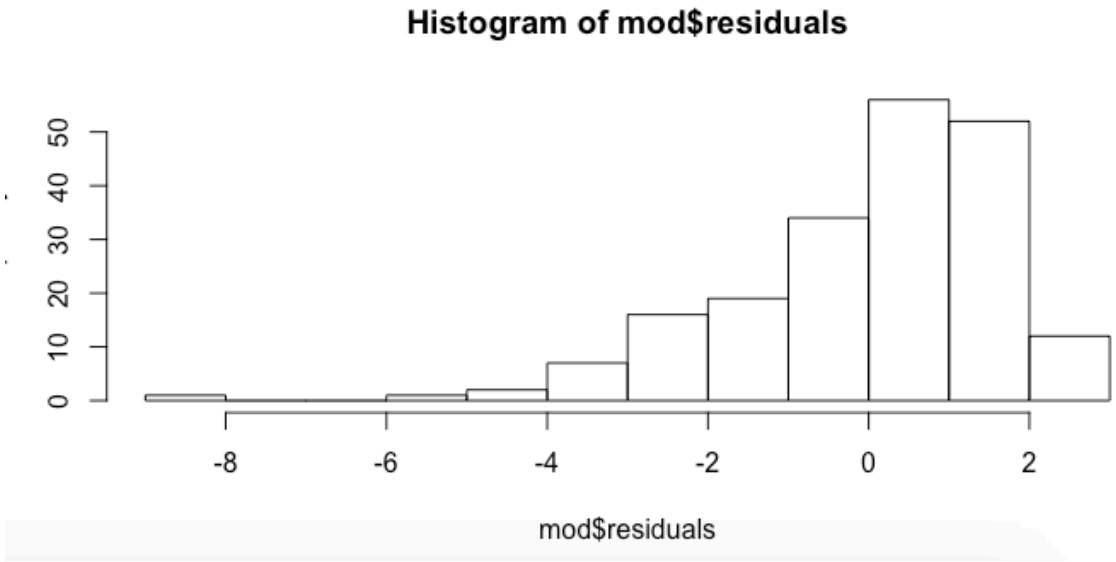
```
> anova(mod)
Analysis of Variance Table

Response: Sales
           Df Sum Sq Mean Sq   F value Pr(>F)
TV          1 3314.6  3314.6 1166.7308 <2e-16 ***
Radio       1 1545.6  1545.6  544.0501 <2e-16 ***
Newspaper   1    0.1     0.1    0.0312 0.8599
Residuals 196  556.8     2.8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(f) Calculate the predicted values and residuals
predict(mod)
mod$residuals

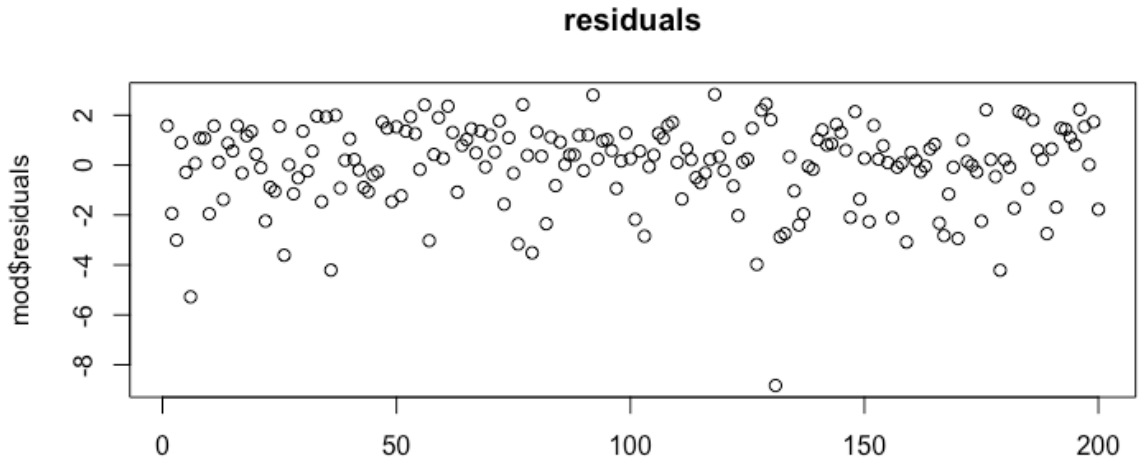
(g) Plot the residuals against the predicted values
plot(predict(mod),mod$residuals)



(h) Plot the histogram of the residuals
plot(hist(mod$residuals))



(i) Comment on the residual plots
plot(mod$residuals,main='residuals')



(j) Use the multivariate model for prediction

predict(mod,as.data.frame(cbind (TV=50,Radio=50,Newspaper=50)))

```
> predict(mod,as.data.frame(cbind (TV=50,Radio=50,Newspaper=50)))
       1
14.60175
```

Question 02: Non Linear Models: Use Advertising data set

(a) Add the Interaction Term TV*Radio and test the significance of the interaction term

model4=lm(Sales~TV+Radio+TV*Radio)
confint(model4)

Output:

```
                  2.5 %      97.5 %
(Intercept) 6.2613828568 7.239057549
TV          0.0161346865 0.022067461
Radio       0.0112978842 0.046422796
TV:Radio    0.0009831143 0.001189875
```

```
> confint(model4)
                    2.5 %      97.5 %
(Intercept) 6.2613828568 7.239057549
TV          0.0161346865 0.022067461
Radio       0.0112978842 0.046422796
TV:Radio    0.0009831143 0.001189875
> cor(Sales,TV*Radio)
[1] 0.963932
```

(b) Give the resulting model after considering this interaction term.

summary(model4)

```
> summary(model4)

Call:
lm(formula = Sales ~ TV + Radio + TV * Radio)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3366 -0.4028  0.1831  0.5948  1.5246

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.750e+00  2.479e-01  27.233   <2e-16 ***
TV          1.910e-02  1.504e-03  12.699   <2e-16 ***
Radio       2.886e-02  8.905e-03   3.241   0.0014 **
TV:Radio    1.086e-03  5.242e-05  20.727   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9435 on 196 degrees of freedom
Multiple R-squared:  0.9678,    Adjusted R-squared:  0.9673
```

(c) Construct the Polynomial Regression Model of order 3 and test the model significance

model.ploy=lm(Sales~poly(TV,3))

anova(model.ploy)

```
> model.ploy=lm(Sales~poly(TV,3))
> anova(model.ploy)
Analysis of Variance Table

Response: Sales
             Df Sum Sq Mean Sq F value    Pr(>F)
poly(TV, 3)   3 3369.5 1123.16  107.51 < 2.2e-16 ***
Residuals   196 2047.7   10.45
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(d) Give the resulting selected model

```
> summary(model.ploy)

Call:
lm(formula = Sales ~ poly(TV, 3))

Residuals:
    Min      1Q  Median      3Q     Max
-7.9734 -1.8900 -0.0897  2.0189  7.3765

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.0225     0.2286  61.353   <2e-16 ***
poly(TV, 3)1  57.5727     3.2322  17.812   <2e-16 ***
poly(TV, 3)2  -6.2288     3.2322  -1.927   0.0554 .
poly(TV, 3)3   4.0074     3.2322   1.240   0.2165
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.232 on 196 degrees of freedom
Multiple R-squared:  0.622,     Adjusted R-squared:  0.6162
F-statistic: 107.5 on 3 and 196 DF,  p-value: < 2.2e-16
```