Assignment5

Aishat Olatunji

2022-09-23

#DATA ENTRY AND INFORMATIONS  
  
Album\_sales<-read.csv("/Users/aishatolatunji/Downloads/Album Sales 3.csv")  
head(Album\_sales)

## sales adverts airplay attract  
## 1 330 10.256 43 10  
## 2 120 985.685 28 7  
## 3 360 1445.563 35 7  
## 4 270 1188.193 33 7  
## 5 220 574.513 44 5  
## 6 170 568.954 19 5

dim(Album\_sales) # based on the dimensions it shows that we 200 rows in the dataset and 4 attributes which are sales , adverts, airplay and attract

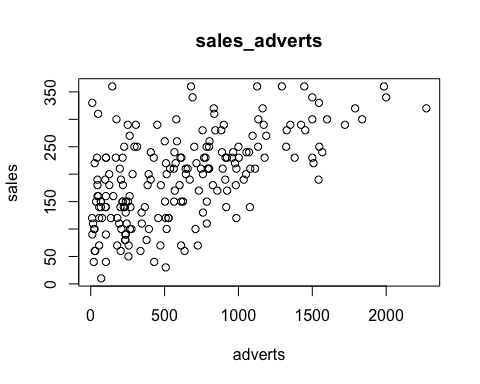
## [1] 200 4

colnames(Album\_sales)

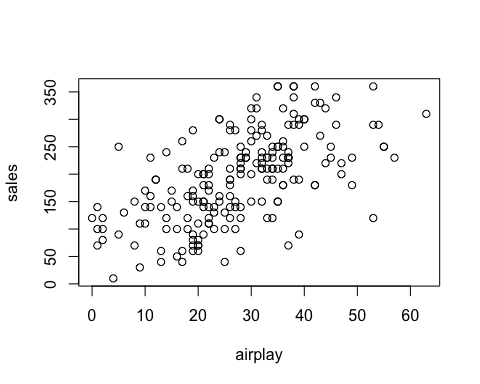
## [1] "sales" "adverts" "airplay" "attract"

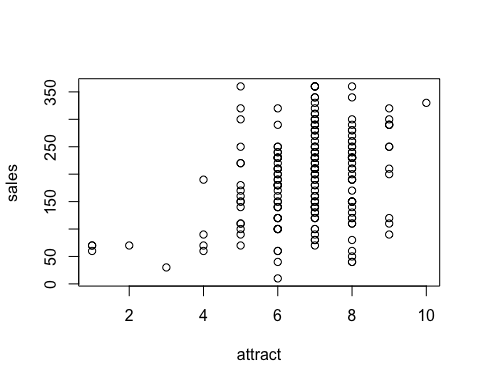
attach(Album\_sales)# the neccesity of attaching my dataset is for this software to identify my attributes names without having to attach it with my dataset name.  
# The dependent variable in this case is sales  
# The independent variable are adverts, airplay and attract

# number one  
  
sales\_adverts = plot(adverts,sales,main="sales\_adverts", xlab = "adverts",ylab = "sales" )



# number two  
sales\_airplay = plot(airplay,sales)

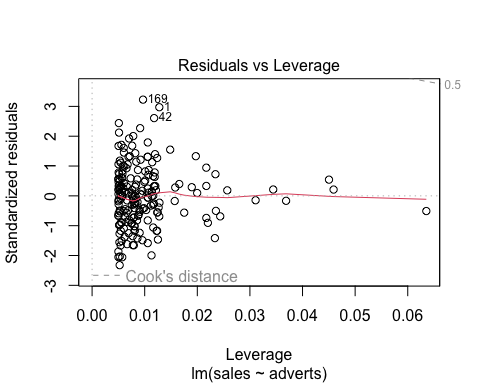
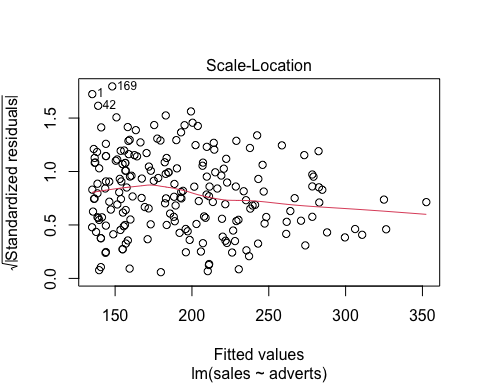
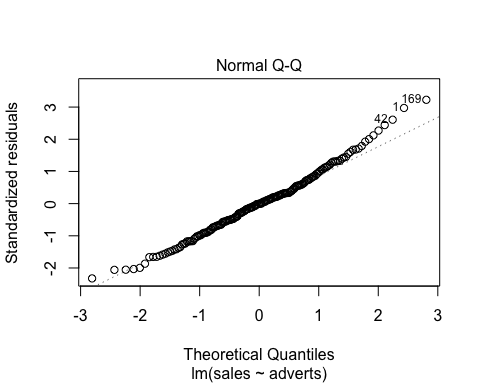
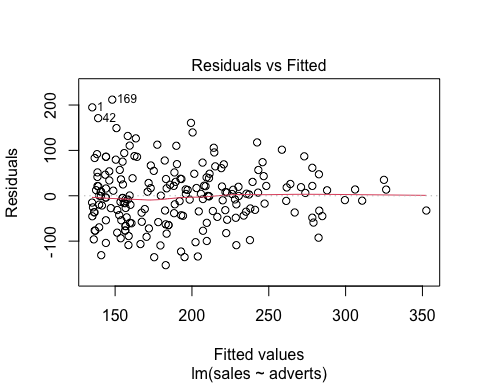
# number three  
sales\_attract = plot(attract,sales)



# number 4  
model1 = lm(sales~adverts, data = Album\_sales)  
summary(model1)

##   
## Call:  
## lm(formula = sales ~ adverts, data = Album\_sales)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -152.949 -43.796 -0.393 37.040 211.866   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.341e+02 7.537e+00 17.799 <2e-16 \*\*\*  
## adverts 9.612e-02 9.632e-03 9.979 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 65.99 on 198 degrees of freedom  
## Multiple R-squared: 0.3346, Adjusted R-squared: 0.3313   
## F-statistic: 99.59 on 1 and 198 DF, p-value: < 2.2e-16

# F-statistic: 99.59 , p-value: < 2.2e-16  
plot(model1)



# number 5  
  
# Ho: model is not good vs H1: model is good  
# Since this p-value is less than .05, we can reject the null hypothesis and conclude that the model is good.

# F statistics interpretation  
# since the fcal(99.59) > ftab (3.89) 0n 95% level of significant, we reject the null hypothesis and conclude that the model is a good model,

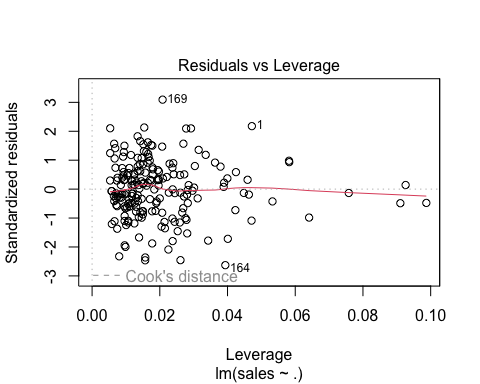
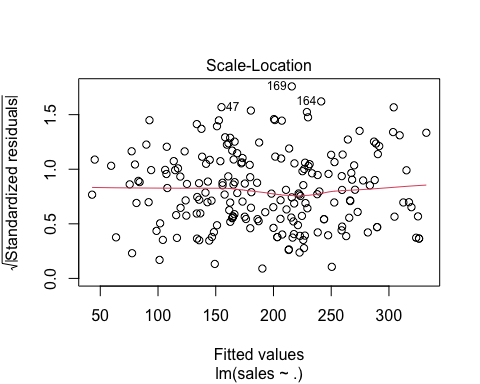
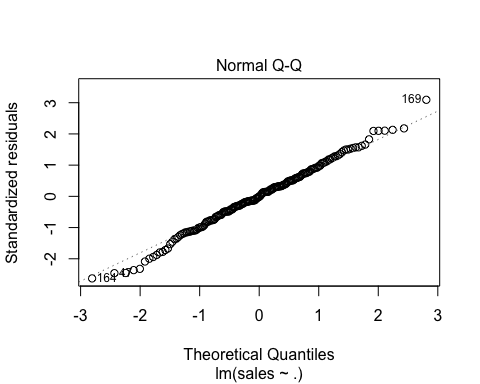
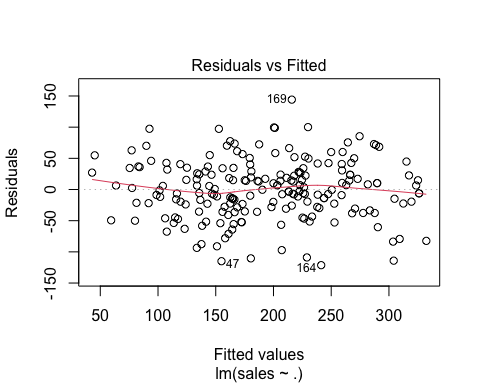
# number six  
  
# intercept value = 1.341e+02  
# adverts value = 9.612e-02   
  
# Abulm\_sales = 1.341e+02 + 9.612e-02\*adverts

# number 7  
  
# Abulm\_sales = 1.341e+02 + 9.612e-02\*adverts  
# Abulm\_sales = 1.341e+02 + 9.612e-02\*(135,000)  
# Abulm\_sales = 134.1 + 0.09612\*(135,000) = 13110.3 or approximately 13110 records will be sold.  
   
  
# approximately 13110.3 records will be sold or approximately 13110 records will be sold. if we spent $135 000 on advertising the latest album “Dear Agony” by Breaking Benjamin

# number 8  
  
model2 = lm(sales~., data = Album\_sales)  
summary(model2)

##   
## Call:  
## lm(formula = sales ~ ., data = Album\_sales)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -121.324 -28.336 -0.451 28.967 144.132   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -26.612958 17.350001 -1.534 0.127   
## adverts 0.084885 0.006923 12.261 < 2e-16 \*\*\*  
## airplay 3.367425 0.277771 12.123 < 2e-16 \*\*\*  
## attract 11.086335 2.437849 4.548 9.49e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 47.09 on 196 degrees of freedom  
## Multiple R-squared: 0.6647, Adjusted R-squared: 0.6595   
## F-statistic: 129.5 on 3 and 196 DF, p-value: < 2.2e-16

plot(model2)



#F-statistic: 129.5 on 3 and 196 DF, p-value: < 2.2e-16  
  
# interpretations  
  
# Ho: model is not good vs H1: model is good  
# Since this p-value(2.2e-16) is less than .05, we can reject the null hypothesis. And conclude that the model is good between sales and the attributes adverts, attract and airplay and it also shows that the model is slightly a better model combinning all the attributes together.

# F statistics interpretation)  
# since the f-cal (129.5 ) > ftab(f distribution) (3.92) 0n 95% level of significant, we reject the null hypothesis and conclude that the model is a good model, which basically means there is relationship or weak relationship between sales and all the attributes.

# number 9  
  
# Based on the r\_squred of both model 1 and model 2,  
# The r\_squared for model 1 is extremely small(0.3346) which it means it's not as good of a model compare to model 2 therefore model 2 is said to be a better model than one because it tends to minimize the error better based on the r\_squared which is (0.6647) and definitely a better model .