Hw3stat632

Jessica Grover

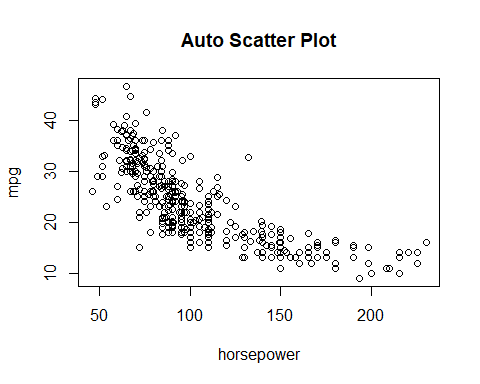
3/5/2022

#Problem 1

library(ISLR)

#a. Make a scatter plot with mpg on the y-axis, and horsepower on the x-axis.

mpg=Auto$mpg  
horsepower=Auto$horsepower  
plot(horsepower,mpg,main="Auto Scatter Plot")

 #b. Use the lm() function to estimate a second degree (quadratic) polynomial regression model. That is, fit the model Y = β0 + β1x + β2x2 + e, where Y = mpg and x = horsepower. Use the summary() function to print the results.

equ=lm(mpg~horsepower+I(horsepower^2))  
summary(equ)

##   
## Call:  
## lm(formula = mpg ~ horsepower + I(horsepower^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.7135 -2.5943 -0.0859 2.2868 15.8961   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.9000997 1.8004268 31.60 <2e-16 \*\*\*  
## horsepower -0.4661896 0.0311246 -14.98 <2e-16 \*\*\*  
## I(horsepower^2) 0.0012305 0.0001221 10.08 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.374 on 389 degrees of freedom  
## Multiple R-squared: 0.6876, Adjusted R-squared: 0.686   
## F-statistic: 428 on 2 and 389 DF, p-value: < 2.2e-16

From the output, the equation for polynomial regression is: mpg=56.9000997-0.4661896*horsepower+0.0012305*horsepower^2

#c. Use the fitted regression model to make a prediction and 95% prediction interval for the mpg of a vehicle that has horsepower = 150.

newdata=data.frame(horsepower=150)  
predict(equ,newdata,interval = "prediction")

## fit lwr upr  
## 1 14.65872 6.027273 23.29016

horsepower=150 —— given then, mpg=56.9000997-0.4661896*150+0.0012305*150^2=14.65872

The 95% prediction interval of mpg is (6.027273,23.29016)

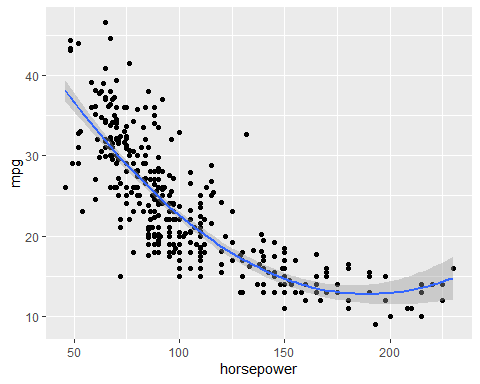
#d. Add the fitted second degree polynomial regression curve to the scatter plot of mpg versus horsepower. You may use either the base-R or ggplot2 approach.

library(ggplot2)

##   
## Attaching package: 'ggplot2'

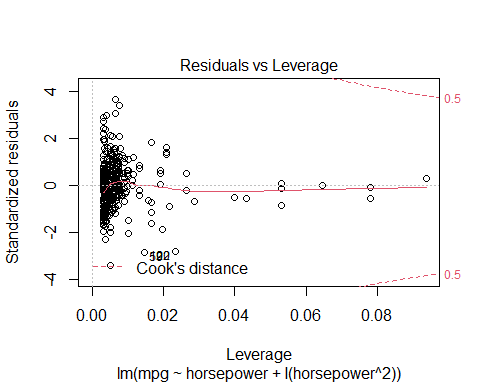
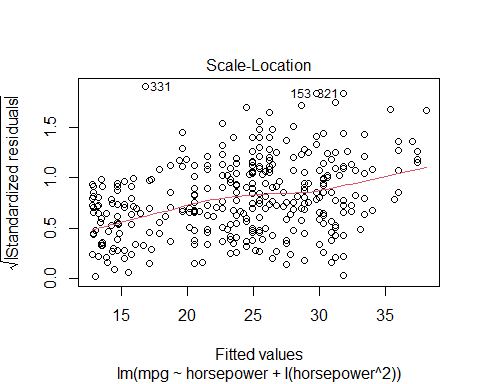
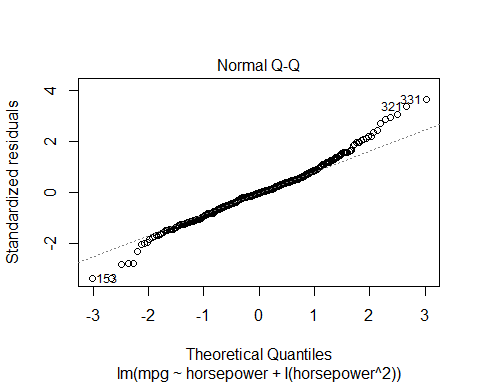
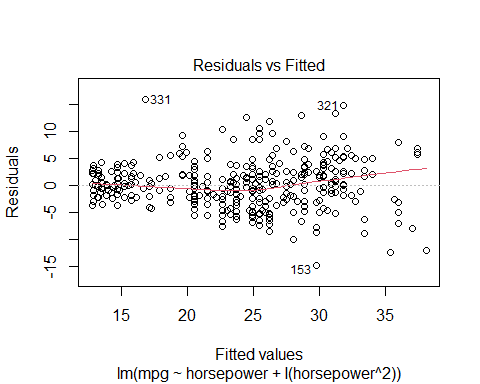
## The following object is masked \_by\_ '.GlobalEnv':  
##   
## mpg

p <- ggplot(Auto, aes(x=horsepower,y=mpg)) + geom\_point()  
p + stat\_smooth(method = "lm", formula = y ~ x + I(x^2), size = 1)

 The scattered plot is not scattered,it does not satisfy the normality condition.This is the second degree curve for mpg vs horsepower.

#e. Make a plot of the residuals versus fitted values, and a QQ plot of the standardized residuals. Comment on whether or not there are any violations of the assumptions for regression modeling.

par=(mfrow=c(2,2))  
plot(equ)

 It is clear that many points are not close to the line so the assumption is not linear. #Problem 2

library(ISLR)  
head(Carseats)

## Sales CompPrice Income Advertising Population Price ShelveLoc Age Education  
## 1 9.50 138 73 11 276 120 Bad 42 17  
## 2 11.22 111 48 16 260 83 Good 65 10  
## 3 10.06 113 35 10 269 80 Medium 59 12  
## 4 7.40 117 100 4 466 97 Medium 55 14  
## 5 4.15 141 64 3 340 128 Bad 38 13  
## 6 10.81 124 113 13 501 72 Bad 78 16  
## Urban US  
## 1 Yes Yes  
## 2 Yes Yes  
## 3 Yes Yes  
## 4 Yes Yes  
## 5 Yes No  
## 6 No Yes

#a.Fit a multiple linear regression model to predict Sales using Price, Urban, and US.

library(ISLR)  
df<-Carseats  
model1<-lm(Sales~Price+Urban+US,data=df)  
summary(model1)

##   
## Call:  
## lm(formula = Sales ~ Price + Urban + US, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.9206 -1.6220 -0.0564 1.5786 7.0581   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.043469 0.651012 20.036 < 2e-16 \*\*\*  
## Price -0.054459 0.005242 -10.389 < 2e-16 \*\*\*  
## UrbanYes -0.021916 0.271650 -0.081 0.936   
## USYes 1.200573 0.259042 4.635 4.86e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.472 on 396 degrees of freedom  
## Multiple R-squared: 0.2393, Adjusted R-squared: 0.2335   
## F-statistic: 41.52 on 3 and 396 DF, p-value: < 2.2e-16

#b. Provide an interpretation of each coefficient in the model. Note that some of the variables are qualitative. The interpretations are as follows: 1. The sale of the store is not affected if it is in urban area or not. 2.When price increases by $1000 and other predictors are held constant, sales decrease by 54.459 unit sales. In otherwords, when price increases by $1000, the number of Carseats sold decrease by 54,459. 3.The sales of a store in US is 1200 more carseats(approx) as compared to other stores abroad.

#c.Write our the equation for the fitted model. The equation is: Sales=13.043469-0.054459*Price-0.021916*Urban+1.200573\*US

#d. For which of the predictors can you reject the null hypothesis H0 : βj = 0? We can reject the null hypothesis for the predictor “Urban”. The p-value is not statistically significant with a value of 0.936.

#e. On the basis of the your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.

model2=lm(Sales~Price+US,data=Carseats)  
summary(model2)

##   
## Call:  
## lm(formula = Sales ~ Price + US, data = Carseats)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.9269 -1.6286 -0.0574 1.5766 7.0515   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.03079 0.63098 20.652 < 2e-16 \*\*\*  
## Price -0.05448 0.00523 -10.416 < 2e-16 \*\*\*  
## USYes 1.19964 0.25846 4.641 4.71e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.469 on 397 degrees of freedom  
## Multiple R-squared: 0.2393, Adjusted R-squared: 0.2354   
## F-statistic: 62.43 on 2 and 397 DF, p-value: < 2.2e-16

#f. How well do the models in (a) and (e) fit the data? Based on their respective R-square values, these two models are mediocre (only 24% change in response explained).

#g. Using the model from (e), obtain 95% confidence intervals for the coefficients.

confint(model2)

## 2.5 % 97.5 %  
## (Intercept) 11.79032020 14.27126531  
## Price -0.06475984 -0.04419543  
## USYes 0.69151957 1.70776632