Assignment 4

Alka Santosh Naik

2023-09-26

1. Conduct a simple regression analysis using your sample data with the dependent variable being “price” and the independent variable being “odometer”.

#Alka Santosh Naik  
rm(list=ls())  
library(rio)  
carlist=import("6304 Module 4 Assignment Data.xlsx",sheet="Craigs List Cars")  
colnames(carlist)=tolower(make.names(colnames(carlist)))  
set.seed(16999752)  
sample.data=carlist[sample(1:nrow(carlist),200),]  
attach(sample.data)  
sample.out=lm(price~odometer,data=sample.data)

1. Give verbal interpretations of all beta coefficients in your regression model. Make certain the language you use is understandable to a reasonably competent lay person shopping for a car on Craig’s List.

summary(sample.out)

##   
## Call:  
## lm(formula = price ~ odometer, data = sample.data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14961 -5561 -1744 4265 37502   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.434e+04 1.359e+03 17.906 < 2e-16 \*\*\*  
## odometer -9.477e-02 1.081e-02 -8.764 8.5e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9193 on 198 degrees of freedom  
## Multiple R-squared: 0.2795, Adjusted R-squared: 0.2759   
## F-statistic: 76.81 on 1 and 198 DF, p-value: 8.505e-16

Observation: From the beta coefficients in the regression model, we can interpret that for every 1000 increase in the odometer reading the price decreases by 94.77.

Slop: -9.477e-02, Intercept: 2.434e+04

1. Evaluate and interpret both the p value and the confidence interval on the “odometer” coefficient in your regression model.

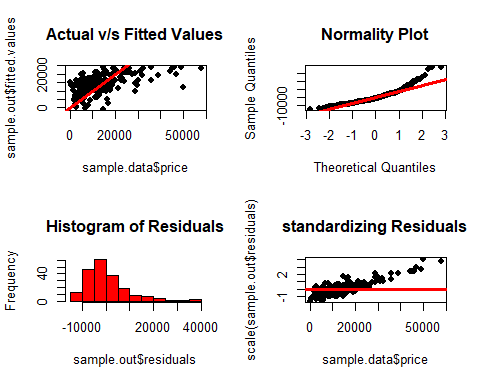
confint(sample.out)

## 2.5 % 97.5 %  
## (Intercept) 21658.6712318 2.701973e+04  
## odometer -0.1160966 -7.344789e-02

Observation: The p value of the odometer is approaching zero, so we can reject the null hypothesis and accept the alternating hypothesis that slop is not equal to zero which indicates there is some linear relationship. From the confidence interval we can determine the width, which is high leading to wide spread and less precision.

1. Run appropriate diagnostics on your regression model to determine if it is in conformity with the LINE assumptions of regression.

# Checking the Assumptions of Regression  
par(mfrow=c(2,2))  
#Linearity  
plot(sample.data$price,sample.out$fitted.values,  
 pch=19,main="Actual v/s Fitted Values")  
abline(0,1,col="red",lwd=3)  
  
#Normality  
qqnorm(sample.out$residuals,pch=19,main="Normality Plot")  
qqline(sample.out$residuals,col="red",lwd=3)  
hist(sample.out$residuals,col="red",  
 main="Histogram of Residuals")  
  
#Equality of Variances  
plot(sample.data$price,scale(sample.out$residuals),  
 pch=19,  
 main="standardizing Residuals")  
abline(0,0,col="red",lwd=3)



par(mfrow=c(1,1))  
moments::skewness(sample.out$residuals)

## [1] 1.478706

moments::kurtosis(sample.out$residuals)

## [1] 6.020394

Observation:

Linearity: From graph “Actual v/s Fitted Values” we can say the data follows weak linear relationship.

Normality: From the “Normality Plot” graph we can say initially say the data follows weak normality but taking into consideration the graph “Histogram of Residuals” we can see the data is right skewed. skewness value 1.47 which is not equal to zero and kurtosis value 6.02 which is not equal to 3 leads state the data is not normally distributed.

Equality of Variances: From the graph “standardizing Residuals” we do not observe any repeating patterns in the plotted data hence, we can say that data follows Equality of Variances.

1. Ms. Trayla Parks is considering offering her Toyota for sale on Craig’s List. The Toyota currently has an odometer reading of 78,521 miles. Use your regression model to predict the price of the vehicle on Craig’s List. Determine and verbally interpret the appropriate confidence interval on this prediction. If Trayla kept her vehicle one more year until the odometer showed 98,000 by how much would your model predict the price of her car would change?

#prediction for odometer reading 78521  
newdata=data.frame(odometer=78521)  
predict(sample.out,newdata,interval="predict")

## fit lwr upr  
## 1 16897.59 -1289.317 35084.5

predict(sample.out,newdata,interval="confidence")

## fit lwr upr  
## 1 16897.59 15446.62 18348.56

Observation: The predicted price for odometer reading 78521 is 16897.59, here we are 95% confident that the predicted value can take any value from the range -1289.317 to 35084.5(at odometer reading 78521). Width of the confidence interval is high leading to wide spread and less precision.

#prediction for odometer reading 98000  
newdata2=data.frame(odometer=98000)  
predict(sample.out,newdata2,interval="predict")

## fit lwr upr  
## 1 15051.52 -3124.602 33227.64

predict(sample.out,newdata2,interval="confidence")

## fit lwr upr  
## 1 15051.52 13742.64 16360.4

#difference in the price prediction for odometer reading 78521 and 98000  
16897.59-15051.52

## [1] 1846.07

Observation: If Trayla kept her vehicle one more year until the odometer showed 98,000 the price of her vehicle would decrease from 16897.59 to 15051.52, that is a 1846.07 decrease in the price.