

Automatic or Manual Transmission for MPG

liuyong

2015-11-22

Data Description

The data come from the datasets of mtcars. It was extracted from 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

The report focus on the parameter of MPG (miles per gallon). Base on the MPG, it gives the analysis which is better between automatic and manual transmission.

Data Preprocess

At first, we have a simple preprocess for the mtcars dataset. Including

```
library(knitr)
library(caret)
data(mtcars)
data <- mtcars
class(data$cyl ) <- as.factor(data$cyl)
class(data$vs) <- as.factor(data$vs)

####split the data base on the automatic and manual transmission
####remove the variable of am
autoTrans <- data[data$am == 0,]
manualTrans <- data[data$am == 1,]
autoTrans <- autoTrans[,-9]
manualTrans <- manualTrans[,-9]

### other preprocess if pca for the data,
autoPCAProcess <- preProcess(autoTrans[2:10],method="pca",thresh=0.9)
autoPCADData <- predict(autoPCAProcess,autoTrans[2:10])
autoPCADData <- cbind(autoPCADData,autoTrans$mpg)
names(autoPCADData)[length(names(autoPCADData)) ] <- "mpg"

manualPCAProcess <- preProcess(manualTrans[2:10],method="pca",thresh=0.9)
manualPCADData <- predict(manualPCAProcess,manualTrans[2:10])
manualPCADData <- cbind(manualPCADData,manualTrans$mpg)
names(manualPCADData)[length(names(manualPCADData))] <- "mpg"
```

Analysis base on the Linear Regression

The basic linear regression for the mtcars:

```
autoFit <- lm(mpg~.,data=autoTrans)
manualFit <- lm(mpg~.,data=manualTrans)
###pca linear regression
```

```
autoPCAFit <- lm(mpg~.,data=autoPCAData)
manualPCAFit <- lm(mpg~.,data=manualPCAData)
```

So, we can get the mean and standard deviation for the MPG base on the basic model and pca model.

```
##           mean_auto sd_auto mean_manual sd_manual
## actual          17.14737 3.833966      24.39231  6.166504
## basic linear    17.14737 3.615471      24.39231  6.078318
## pca linear      17.14737 3.419784      24.39231  5.595815
```

According to the mean of MPG between auto and manual, we hold the manual is better than the auto. Because the manual transmission give the larger miles for every gallon.

But the manual also have a problem: the standard deviation is larger than the auto transmission. It means the manual has a float up or down base on the mean.

We also can give a hypothesis test. null hypothesis: The mean of manual and auto is equal. alternative hypothesis: The mean of auto is smaller than the manual.

```
out <- t.test(autoTrans$mpg,manualTrans$mpg)
out$p.value
```

```
## [1] 0.001373638
```

The p-value is 0.001, so we reject the null hypothesis and accept the alternative hypothesis. The mean of auto is smaller than the manual.

Model Selection

The report gives two linear regression model: basic linear model and pca preprocess linear model. Following is the summary of residuals for the two models.

The summary of residuals of auto PCA model:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.0030 -1.5540 -0.1555  0.0000  1.1630  3.4050
```

The summary of residuals of auto basic model:

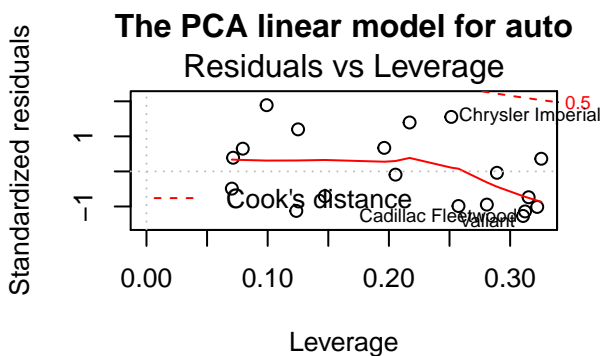
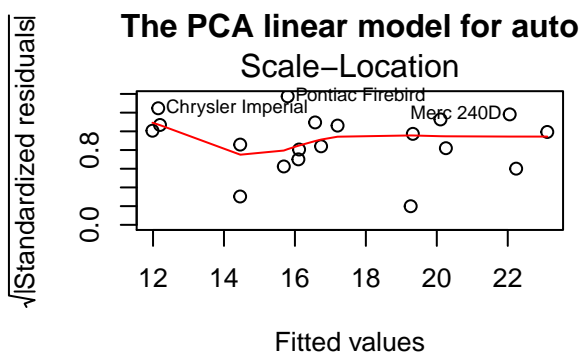
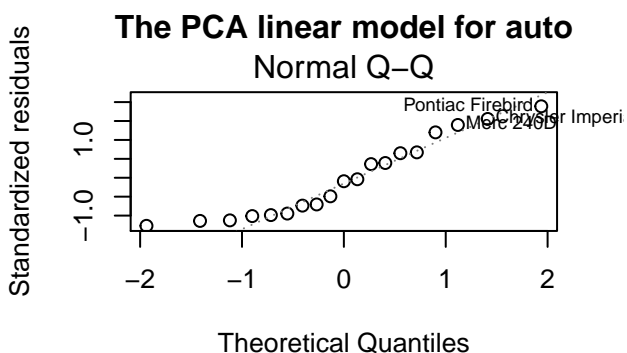
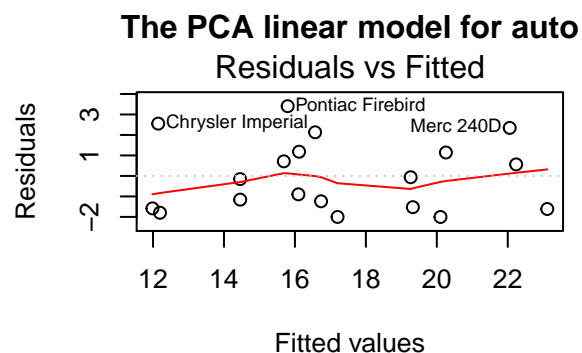
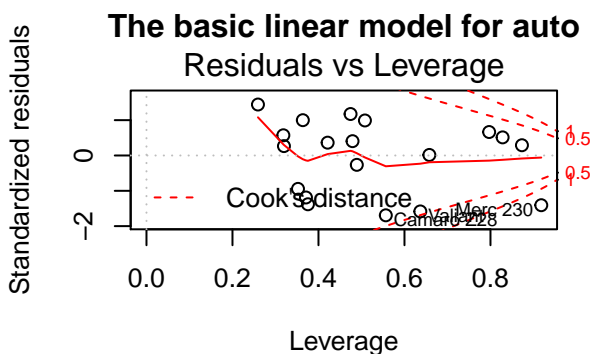
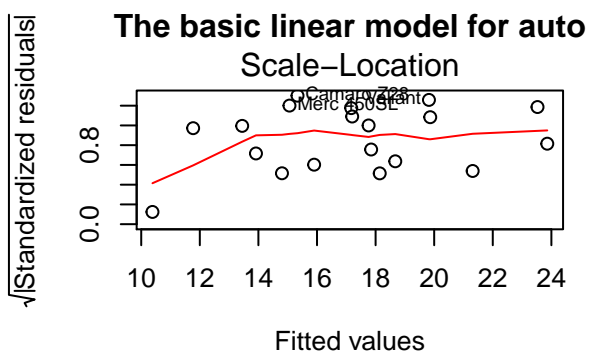
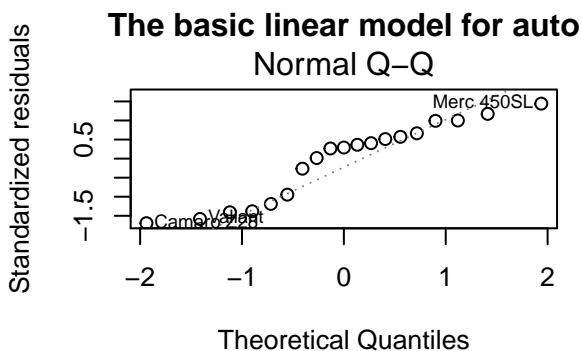
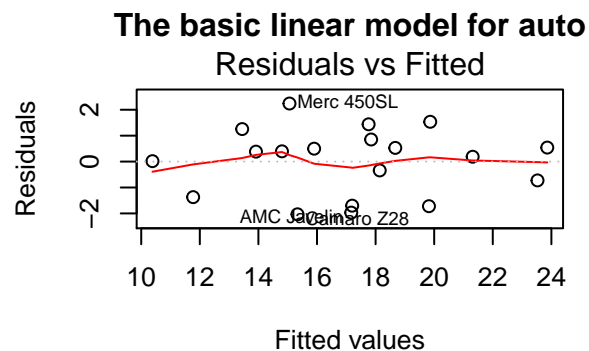
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.0350 -1.0490  0.3835  0.0000  0.6959  2.2390
```

The conclusion is same for the manual transimission. So we hold that the PCA model is much better, because it gives less residuals. The attachments gives the plot of pca predict model.

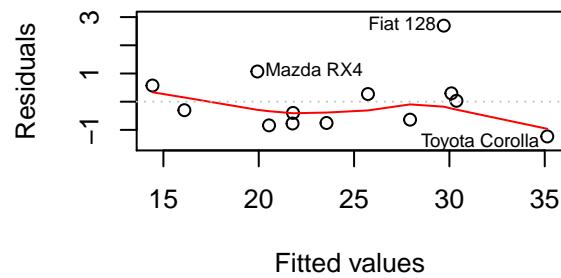
summary

As described in the end, we hold the manual transmission is much better in the MPG parameter . The MPG mean of auto transmission is 17 miles for a gallon, however, the manual transmission is 24 miles. The hypothesis test accept the the manual have much further mileages. Although it has a large standard deviation, we also accept the manual has a better performance in the MPG parameter.

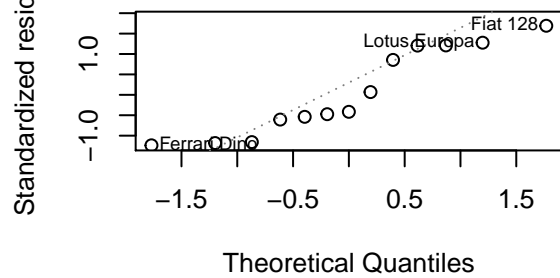
Appendix



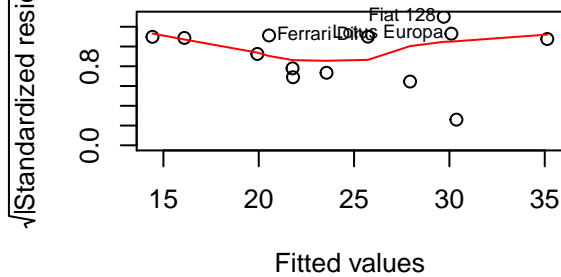
The basic linear model for manual
Residuals vs Fitted



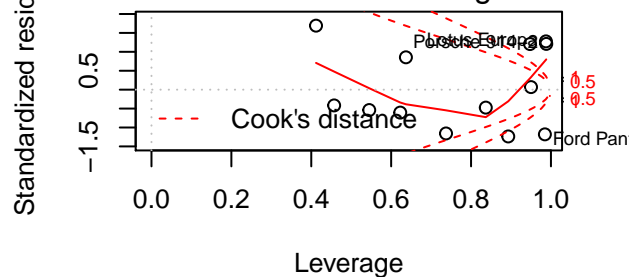
The basic linear model for manual
Normal Q-Q



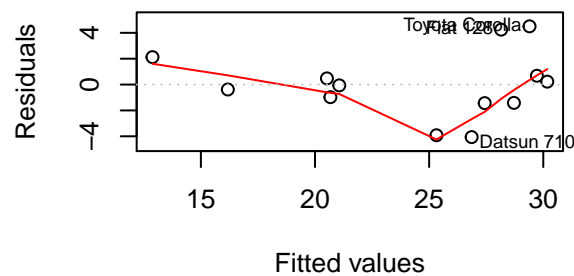
The basic linear model for manual
Scale-Location



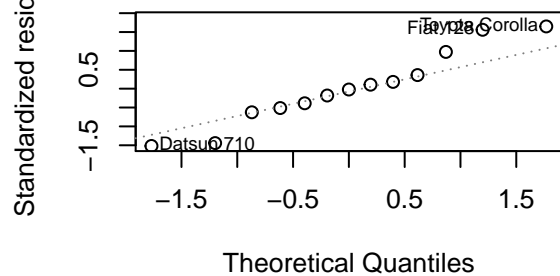
The basic linear model for manual
Residuals vs Leverage



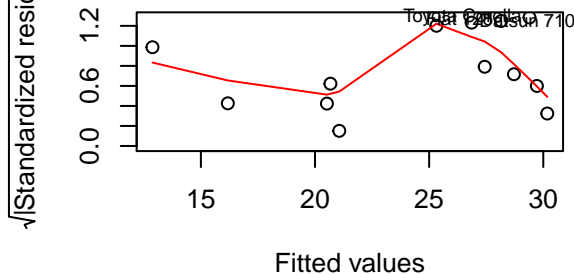
The PCA linear model for manual
Residuals vs Fitted



The PCA linear model for manual
Normal Q-Q



The PCA linear model for manual
Scale-Location



The PCA linear model for manual
Residuals vs Leverage

