

Examine how various aspects of car engines affect performance.

2022-11-19

##Placing packages in place and using them in the code:

```
install.packages("tinytex", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'  
## (as 'lib' is unspecified)
```

```
## package 'tinytex' successfully unpacked and MD5 sums checked  
##  
## The downloaded binary packages are in  
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages
```

```
install.packages("dplyr", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'  
## (as 'lib' is unspecified)
```

```
## package 'dplyr' successfully unpacked and MD5 sums checked  
##  
## The downloaded binary packages are in  
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages
```

```
install.packages("tidyr", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'  
## (as 'lib' is unspecified)
```

```
## package 'tidyr' successfully unpacked and MD5 sums checked  
##  
## The downloaded binary packages are in  
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages
```

```
install.packages("magrittr", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'  
## (as 'lib' is unspecified)
```

```
## package 'magrittr' successfully unpacked and MD5 sums checked
```

```

## Warning: cannot remove prior installation of package 'magrittr'

## Warning in file.copy(savedcopy, lib, recursive = TRUE):
## problem copying C:\Users\vkoyya\AppData\Local\R\win-
## library\4.2\OOLOCK\magrittr\libs\x64\magrittr.dll to C:
## \Users\vkoyya\AppData\Local\R\win-library\4.2\magrittr\libs\x64\magrittr.dll:
## Permission denied

## Warning: restored 'magrittr'

##
## The downloaded binary packages are in
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages

install.packages("knitr", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

##
## There is a binary version available but the source version is later:
## binary source needs_compilation
## knitr 1.40 1.41 FALSE

## installing the source package 'knitr'

install.packages("glmnet", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'glmnet' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages

install.packages("leaps", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'leaps' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages

install.packages("gvlma", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

```

```

## package 'gvlma' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages

install.packages("psych", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'psych' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages

install.packages("latticeExtra", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'latticeExtra' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages

install.packages("caret", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/vkoyya/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'caret' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\vkoyya\AppData\Local\Temp\Rtmp6hjUWK\downloaded_packages

library(knitr)
library(magrittr)

## Warning: package 'magrittr' was built under R version 4.2.2

library(leaps)

## Warning: package 'leaps' was built under R version 4.2.2

library(gvlma)
library(glmnet)

## Warning: package 'glmnet' was built under R version 4.2.2

## Loading required package: Matrix

## Loaded glmnet 4.1-4

```

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.2.2
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.2.2
```

```
## Loading required package: ggplot2
```

```
##
```

```
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
```

```
##
```

```
##      %+%, alpha
```

```
## Loading required package: lattice
```

```
library(latticeExtra)
```

```
## Warning: package 'latticeExtra' was built under R version 4.2.2
```

```
##
```

```
## Attaching package: 'latticeExtra'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
##      layer
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.2
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.2
```

```
##
## Attaching package: 'tidyr'

## The following objects are masked from 'package:Matrix':
##
##     expand, pack, unpack

## The following object is masked from 'package:magrittr':
##
##     extract

##read from main input auto-mpg.csv file

masterdf = read.csv("C:/Users/Public/FinalProject/auto-mpg.csv")
str(masterdf)

## 'data.frame':   398 obs. of  9 variables:
##  $ mpg          : num  18 15 18 16 17 15 14 14 15 ...
##  $ cylinder      : int   8  8  8  8  8  8  8  8  8 ...
##  $ displacement: num  307 350 318 304 302 429 454 440 455 390 ...
##  $ horsepower   : chr  "130" "165" "150" "150" ...
##  $ weight        : int  3504 3693 3436 3433 3449 4341 4354 4312 4425 3850 ...
##  $ acceleration: num   12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
##  $ model_year    : int   70  70  70  70  70  70  70  70  70  70 ...
##  $ origin        : int    1  1  1  1  1  1  1  1  1  1 ...
##  $ car.name      : chr  "chevrolet chevelle malibu" "buick skylark 320" "plymouth satellite" "amc rebe

##The dataframe has been modified.

names(masterdf) = c("mpg","cylinder","displacement","horsepower","weight","acceleration","model_year","c
masterdf$horsepower[masterdf$horsepower=="?"] = NA
masterdf$horsepower = as.numeric(masterdf$horsepower)
masterdf$cylinder = as.numeric(masterdf$cylinder)
str(masterdf)

## 'data.frame':   398 obs. of  9 variables:
##  $ mpg          : num  18 15 18 16 17 15 14 14 15 ...
##  $ cylinder      : num   8  8  8  8  8  8  8  8  8 ...
##  $ displacement: num  307 350 318 304 302 429 454 440 455 390 ...
##  $ horsepower   : num   130 165 150 150 140 198 220 215 225 190 ...
##  $ weight        : int  3504 3693 3436 3433 3449 4341 4354 4312 4425 3850 ...
##  $ acceleration: num   12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
##  $ model_year    : int   70  70  70  70  70  70  70  70  70  70 ...
##  $ origin        : int    1  1  1  1  1  1  1  1  1  1 ...
##  $ car_name      : chr  "chevrolet chevelle malibu" "buick skylark 320" "plymouth satellite" "amc rebe

##separating the underlying data from the data that is transformed:

master1df = select(masterdf,mpg,displacement,horsepower,weight,acceleration)
master1df = na.omit(master1df)
kable(summary(master1df),row.names = FALSE)
```

mpg	displacement	horsepower	weight	acceleration
Min. : 9.00	Min. : 68.0	Min. : 46.0	Min. :1613	Min. : 8.00
1st Qu.:17.00	1st Qu.:105.0	1st Qu.: 75.0	1st Qu.:2225	1st Qu.:13.78
Median :22.75	Median :151.0	Median : 93.5	Median :2804	Median :15.50
Mean :23.45	Mean :194.4	Mean :104.5	Mean :2978	Mean :15.54
3rd Qu.:29.00	3rd Qu.:275.8	3rd Qu.:126.0	3rd Qu.:3615	3rd Qu.:17.02
Max. :46.60	Max. :455.0	Max. :230.0	Max. :5140	Max. :24.80

##In the first instance, there were 300 split records, while in the second instance, there were 98 last records:

```
first.instance = master1df[1:300,]
second.instance = na.omit(master1df[301:398,])
```

##Displacement model for first instance:

```
firstinstance.dis = lm(mpg~displacement, data=first.instance)
summary(firstinstance.dis)
```

```
##
## Call:
## lm(formula = mpg ~ displacement, data = first.instance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.9282 -2.0043 -0.5401  1.9737 16.1501
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31.352035   0.435875   71.93  <2e-16 ***
## displacement -0.048913   0.001809  -27.04  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.412 on 298 degrees of freedom
## Multiple R-squared:  0.7104, Adjusted R-squared:  0.7094
## F-statistic: 731.1 on 1 and 298 DF, p-value: < 2.2e-16
```

```
ffirstinstance.dis = summary(firstinstance.dis)
ffirstinstance.dis$r.squared
```

```
## [1] 0.7104182
```

```
ffirstinstance.dis$adj.r.squared
```

```
## [1] 0.7094464
```

```
coef(ffirstinstance.dis)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31.35203522 0.435875376   71.92890 2.258211e-190
## displacement -0.04891259 0.001809011  -27.03831 3.483733e-82
```

```
coef(firstinstance.dis)
```

```
## (Intercept) displacement  
## 31.35203522 -0.04891259
```

```
##Horsepower model for first instance:
```

```
firstinstance.hrp = lm(mpg-horsepower, data=first.instance)  
summary(firstinstance.hrp)
```

```
##  
## Call:  
## lm(formula = mpg ~ horsepower, data = first.instance)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -10.8442  -2.7816  -0.3376   2.4948  14.2360   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) 34.903508   0.648037   53.86  <2e-16 ***  
## horsepower  -0.125824   0.005455  -23.07  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 3.8 on 298 degrees of freedom  
## Multiple R-squared:  0.641, Adjusted R-squared:  0.6397   
## F-statistic: 532 on 1 and 298 DF, p-value: < 2.2e-16
```

```
ffirstinstance.hrp = summary(firstinstance.hrp)  
ffirstinstance.hrp$r.squared
```

```
## [1] 0.6409527
```

```
ffirstinstance.hrp$adj.r.squared
```

```
## [1] 0.6397479
```

```
coef(ffirstinstance.hrp)
```

```
##              Estimate Std. Error  t value      Pr(>|t|)      
## (Intercept) 34.9035083 0.648036714  53.86039 1.252684e-155  
## horsepower  -0.1258239 0.005455289 -23.06457 3.004974e-68
```

```
coef(firstinstance.hrp)
```

```
## (Intercept) horsepower  
## 34.9035083 -0.1258239
```

```
##Acceleration model for first instance:
```

```
firstinstance.acc = lm(mpg~acceleration, data=first.instance)
summary(firstinstance.acc)
```

```
##
## Call:
## lm(formula = mpg ~ acceleration, data = first.instance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.202  -4.126  -1.012   3.268  16.154
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.0012     1.8352   2.725  0.00681 **
## acceleration    1.0379     0.1183   8.770 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.654 on 298 degrees of freedom
## Multiple R-squared:  0.2052, Adjusted R-squared:  0.2025
## F-statistic: 76.91 on 1 and 298 DF,  p-value: < 2.2e-16
```

```
ffirstinstance.acc = summary(firstinstance.acc)
ffirstinstance.acc$r.squared
```

```
## [1] 0.2051531
```

```
ffirstinstance.acc$adj.r.squared
```

```
## [1] 0.2024858
```

```
coef(ffirstinstance.acc)
```

```
##              Estimate Std. Error t value    Pr(>|t|)
## (Intercept)  5.001162   1.8351855  2.725153 6.807164e-03
## acceleration 1.037865   0.1183411  8.770118 1.397098e-16
```

```
coef(firstinstance.acc)
```

```
## (Intercept) acceleration
##      5.001162      1.037865
```

```
##Weight model for first instance:
```

```
firstinstance.wght = lm(mpg~weight, data=first.instance)
summary(firstinstance.wght)
```

```
##
## Call:
```



```
## lm(formula = mpg ~ weight, data = first.instance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.2011 -1.9157 -0.0812  1.7341 15.0246
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.5619792  0.6461532   62.77  <2e-16 ***
## weight      -0.0062905  0.0001984  -31.71  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.032 on 298 degrees of freedom
## Multiple R-squared:  0.7714, Adjusted R-squared:  0.7706
## F-statistic: 1005 on 1 and 298 DF, p-value: < 2.2e-16
```

```
ffirstinstance.wght = summary(firstinstance.wght)
ffirstinstance.wght$r.squared
```

```
## [1] 0.7713783
```

```
ffirstinstance.wght$adj.r.squared
```

```
## [1] 0.7706111
```

```
coef(ffirstinstance.wght)
```

```
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) 40.561979247 0.6461531581  62.77456 7.613401e-174
## weight      -0.006290453 0.0001983804 -31.70904 1.693958e-97
```

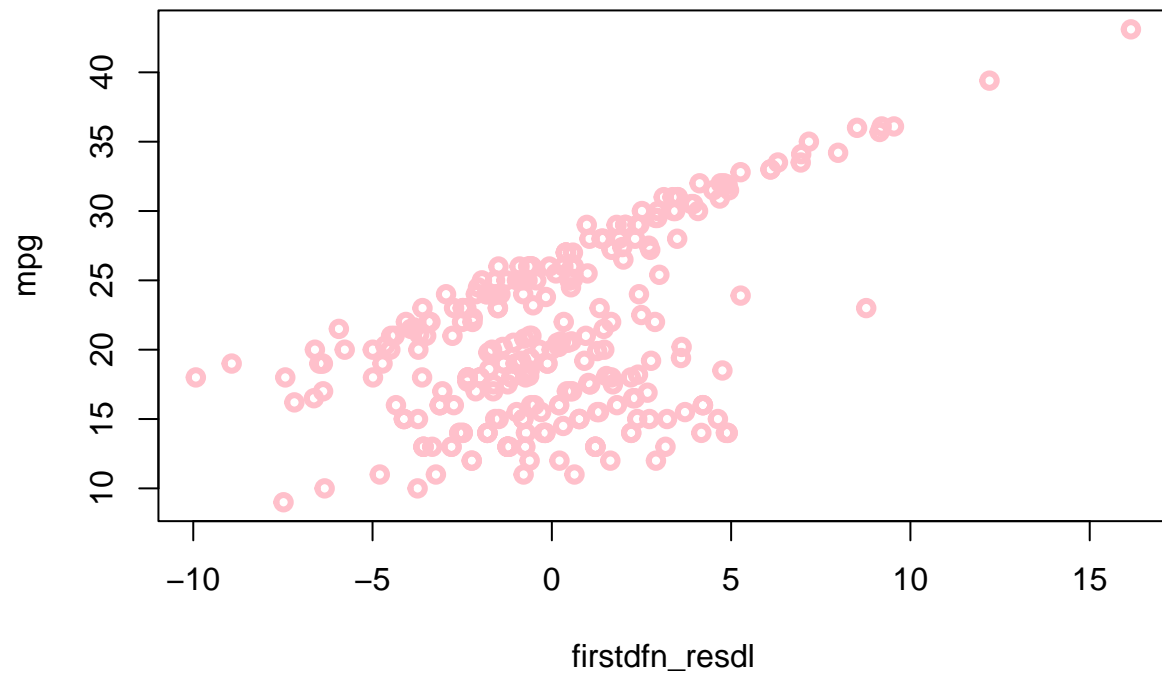
```
coef(firstinstance.wght)
```

```
## (Intercept)      weight
## 40.561979247 -0.006290453
```

```
##Displacement model residual for first instance:
```

```
firstins.dist = lm(mpg~displacement, data=first.instance)
firstdfn_resdl = firstins.dist$residuals
plot(first.instance$mpg~firstdfn_resdl ,lwd=3, col="pink",main="Displacement Model residual for first instance")
```

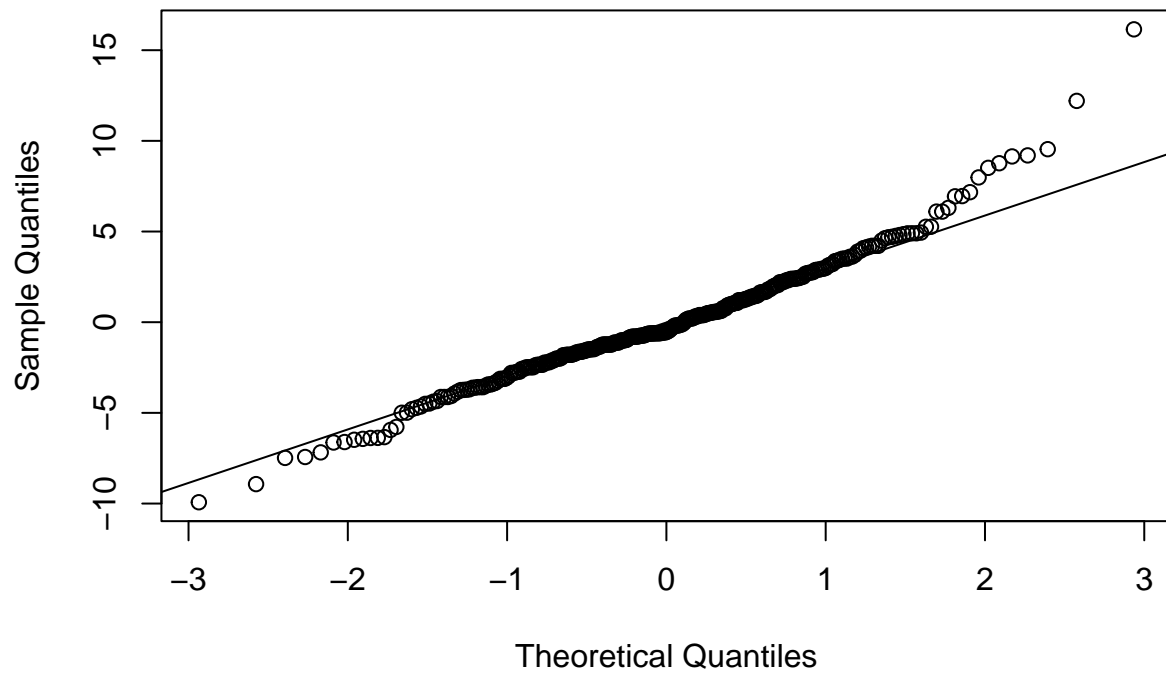
Displacement Model residual for first instance



##Finding linear model residual of displacement for first instance:

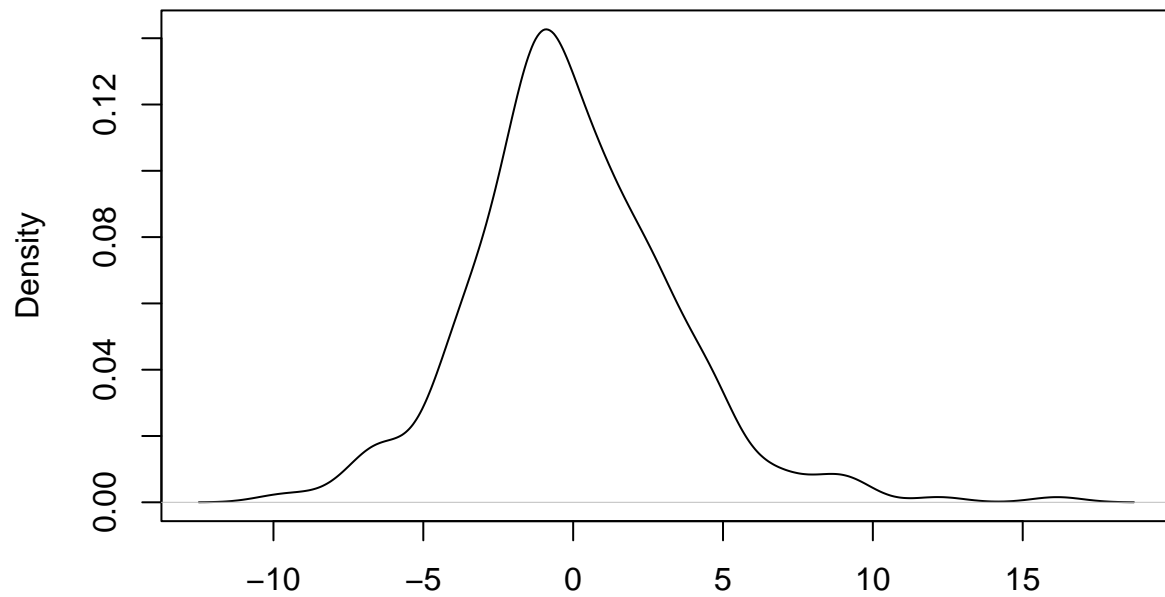
```
qqnorm(resid(firstinstance.dis))  
qqline(resid(firstinstance.dis))
```

Normal Q-Q Plot



```
plot(density(resid(firstinstance.dis)))
```

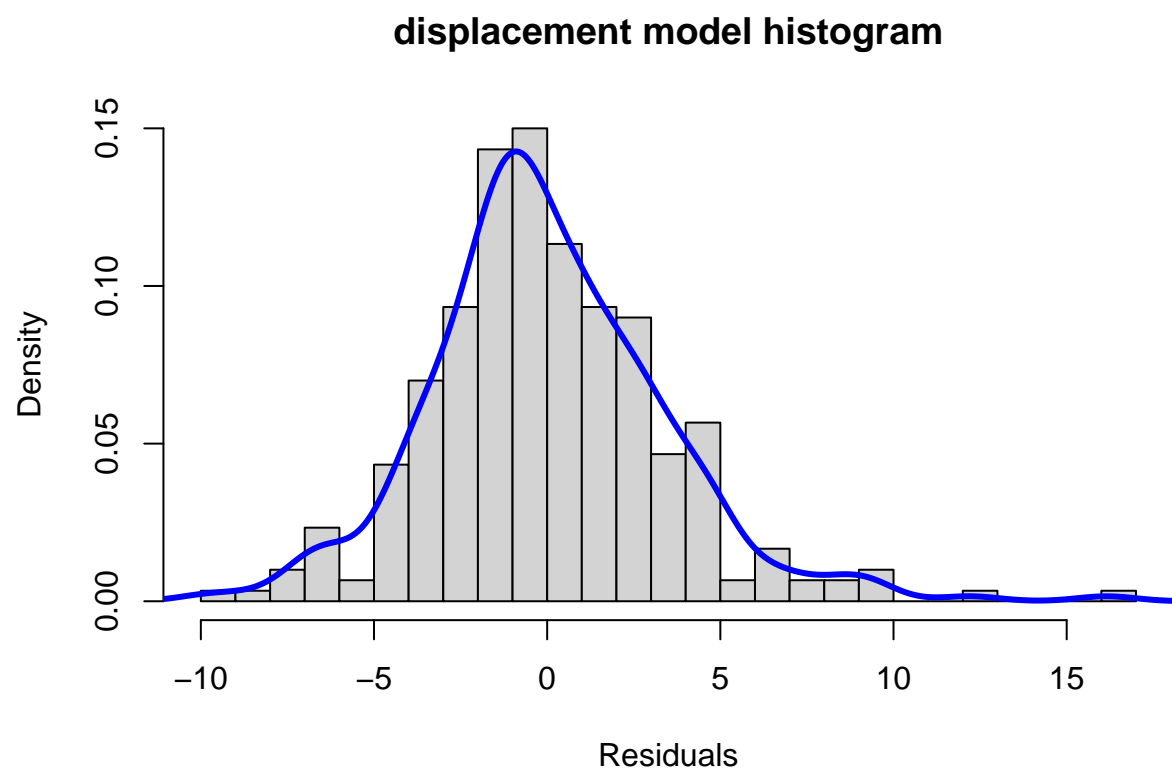
density.default(x = resid(firstinstance.dis))



N = 300 Bandwidth = 0.8539

##Finding Histogram of model displacement for first instance:

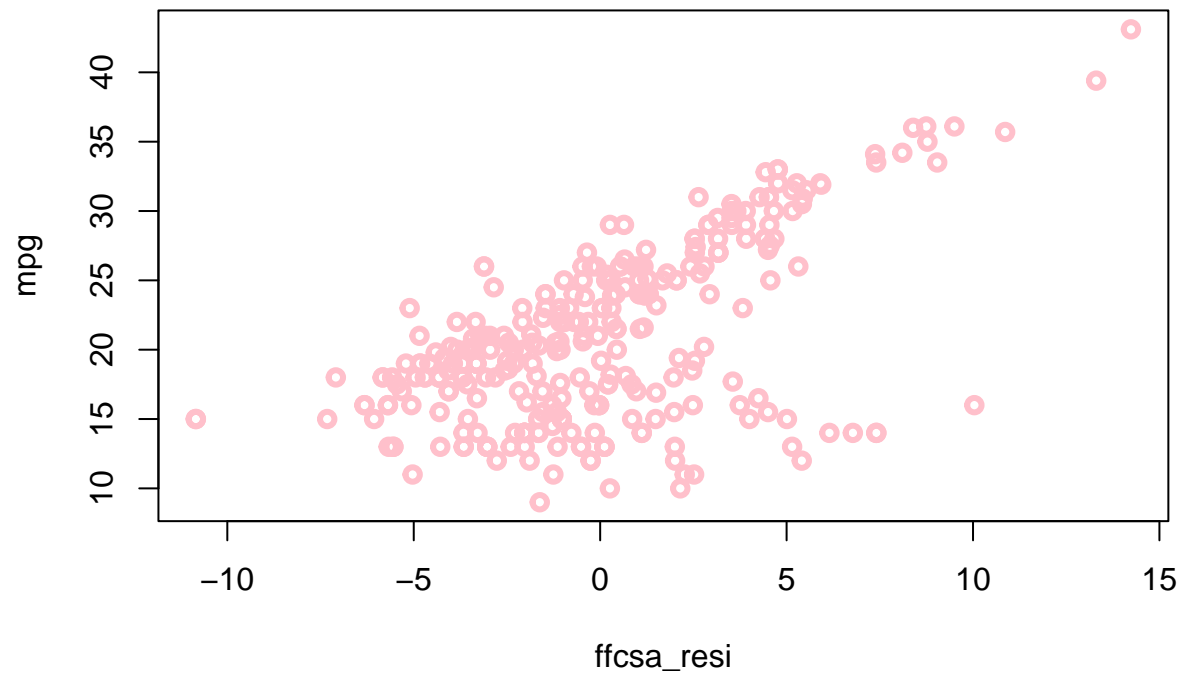
```
hist(firstdfn_resdl ,prob=T,breaks=20,main="displacement model histogram",xlab="Residuals")
lines(density(firstdfn_resdl ),col="blue",lwd=3)
```



##Model of Horsepower Residual for first instance:

```
ffcsa_resi = firstinstance.hrp$residuals  
plot(first.instance$mpg~ffcsa_resi ,lwd=3, col="pink",main="Horsepower Residual for first instance",ylab="Density")
```

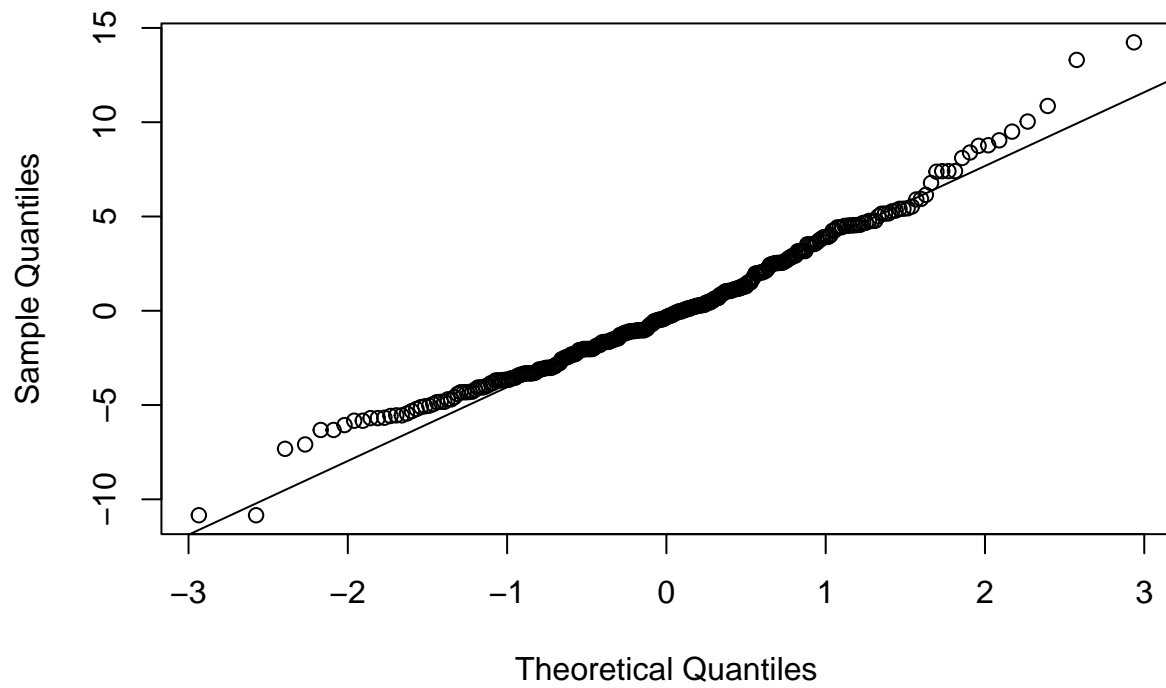
Horsepower Residual for first instance



##linear horsepower models for first instance:

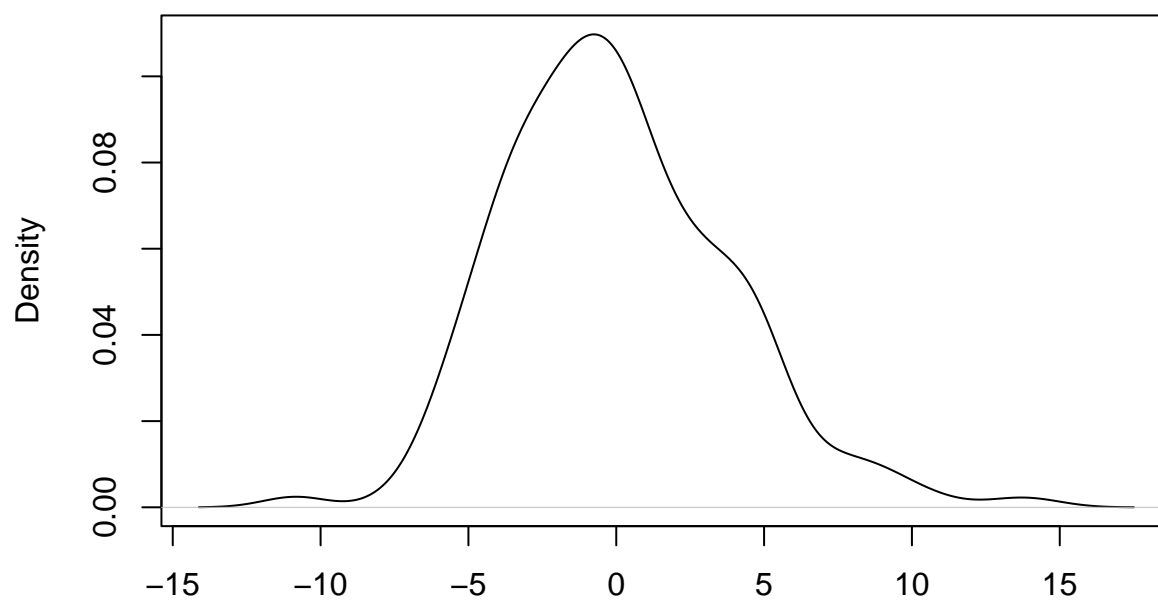
```
qqnorm(resid(firstinstance.hrp))  
qqline(resid(firstinstance.hrp))
```

Normal Q-Q Plot



```
plot(density(resid(firstinstance.hrp)))
```

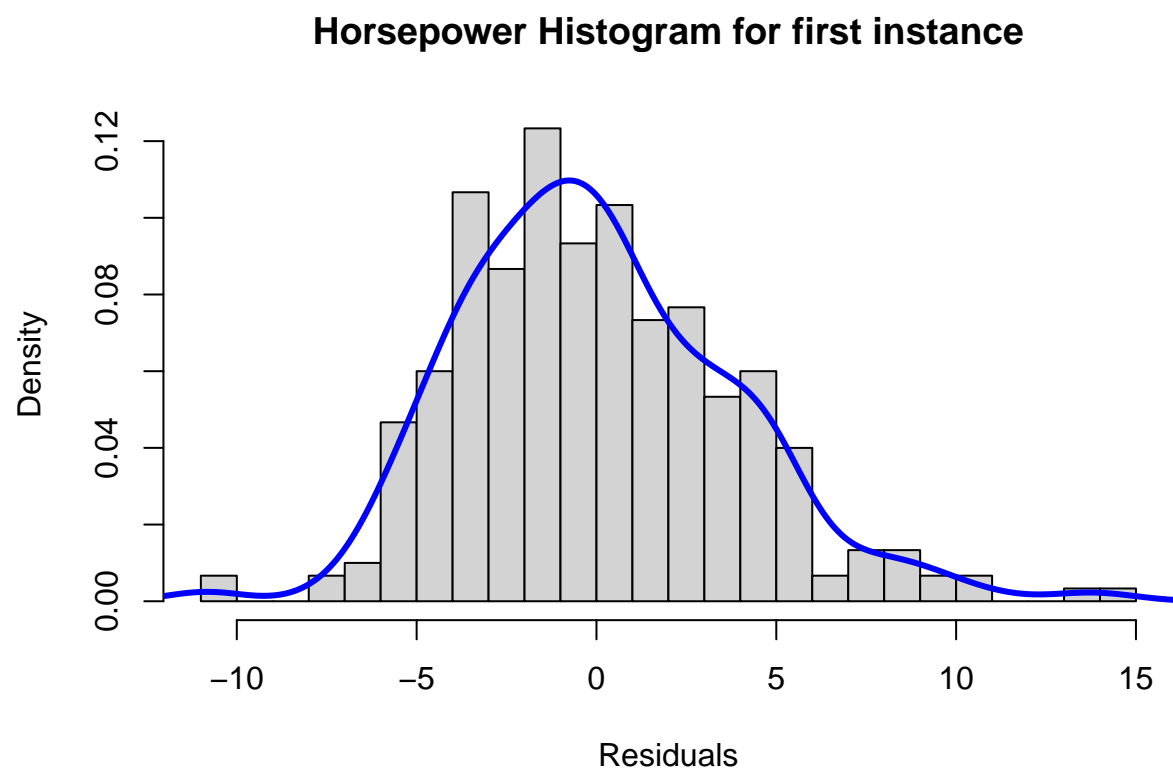
density.default(x = resid(firstinstance.hrp))



N = 300 Bandwidth = 1.091

##horsepower models histogram for first instance:

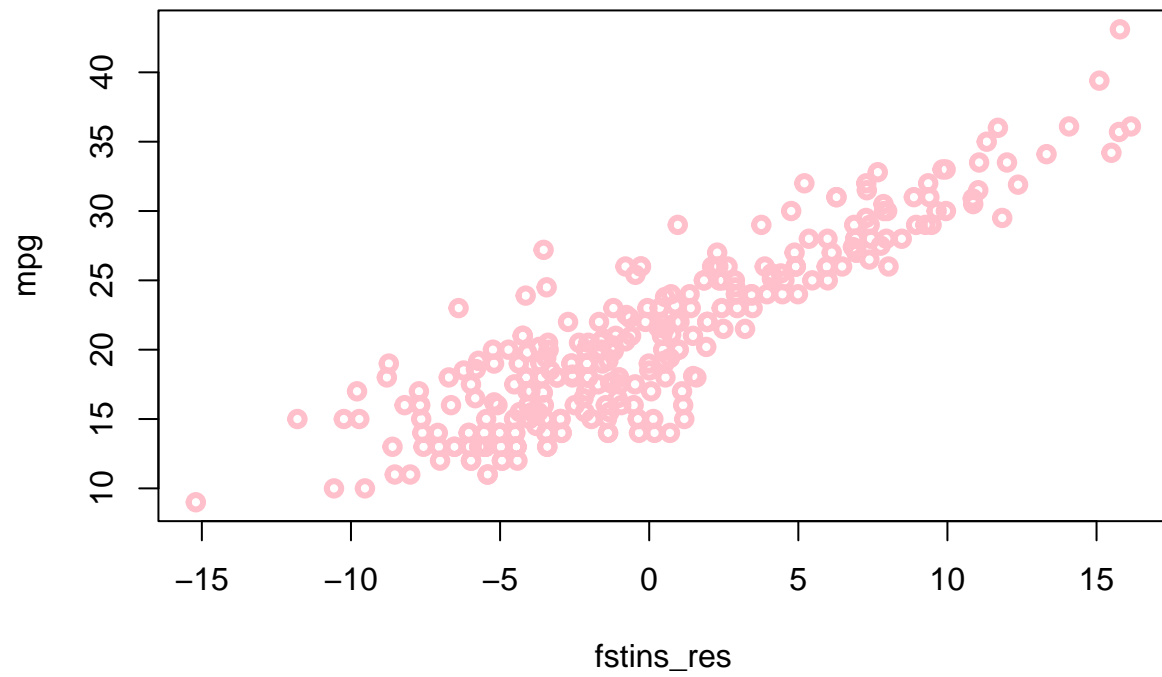
```
hist(ffcsa_resi ,prob=T,breaks=20,main="Horsepower Histogram for first instance",xlab="Residuals")  
lines(density(ffcsa_resi ),col="blue",lwd=3)
```

##acceleration residual model for first instance:

```
fstins_res = firstinstance.acc$residuals  
plot(first.instance$mpg~fstins_res ,lwd=3, col="pink",main="Acceleration Residual for first instance",y
```

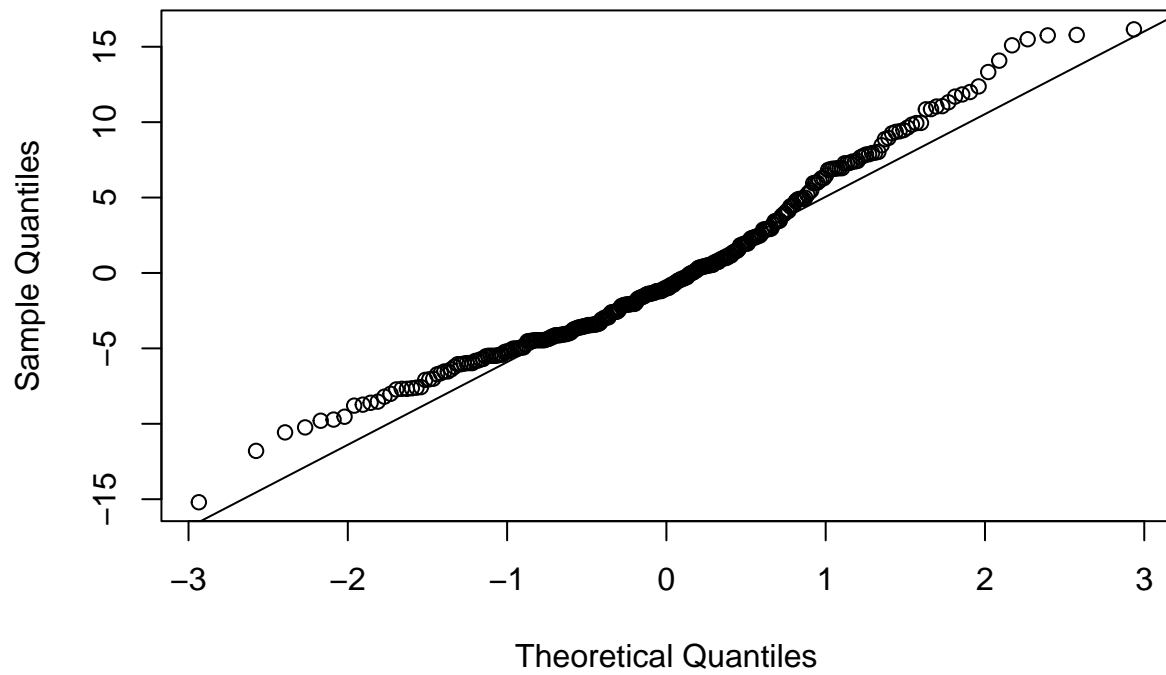
Acceleration Residual for first instance



##linear acceleration residual model for first instance:

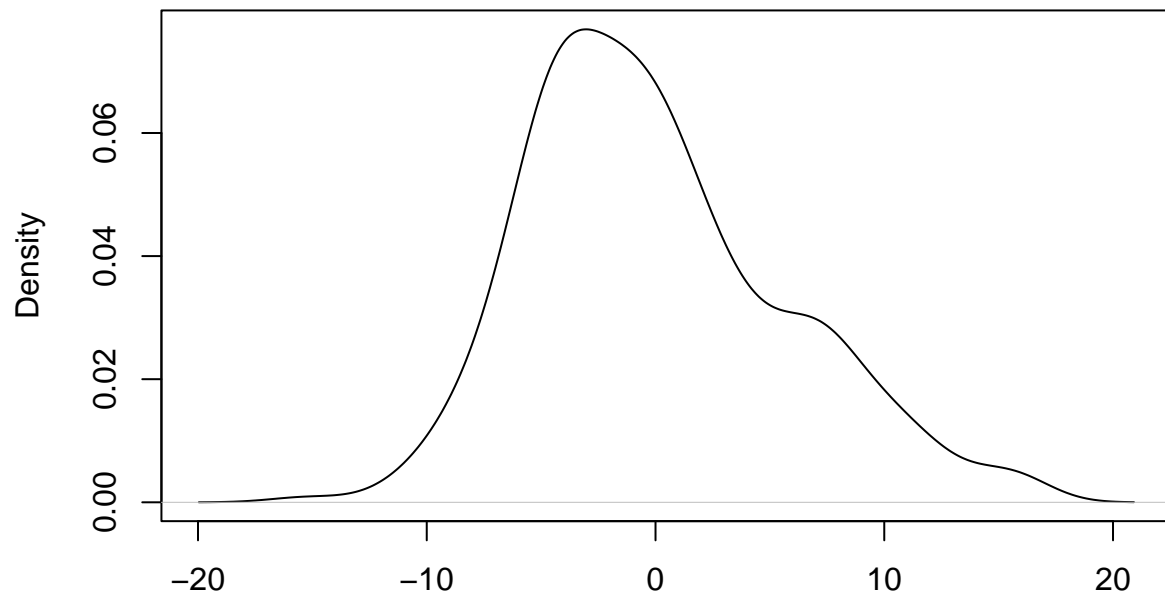
```
qqnorm(resid(firstinstance.acc))  
qqline(resid(firstinstance.acc))
```

Normal Q-Q Plot



```
plot(density(resid(firstinstance.acc)))
```

density.default(x = resid(firstinstance.acc))

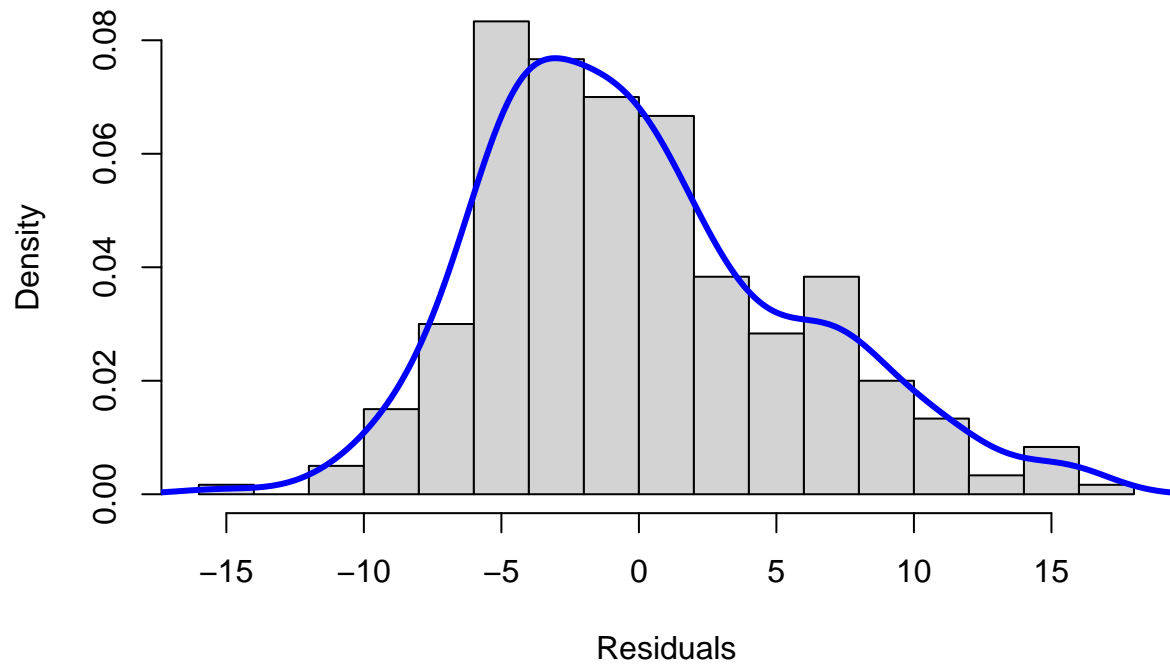


N = 300 Bandwidth = 1.587

##model of the acceleration histogram for first instance:

```
hist(fstins_res ,prob=T,breaks=20,main="Acceleration Histogram model for first instance",xlab="Residual.  
lines(density(fstins_res ),col="blue",lwd=3)
```

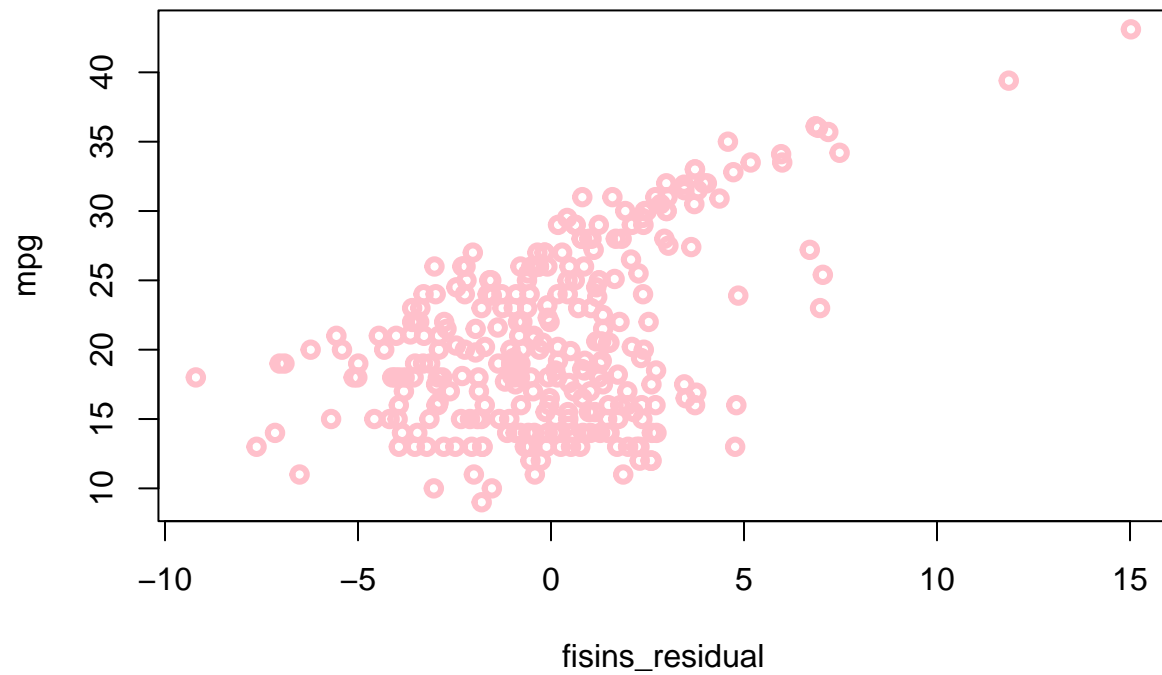
Acceleration Histogram model for first instance



Weight residual model for first instance:

```
fisins_residual = firstinstance.wght$residuals  
plot(first.instance$mpg~fisins_residual ,lwd=3, col="pink",main="weight residual model for first instance")
```

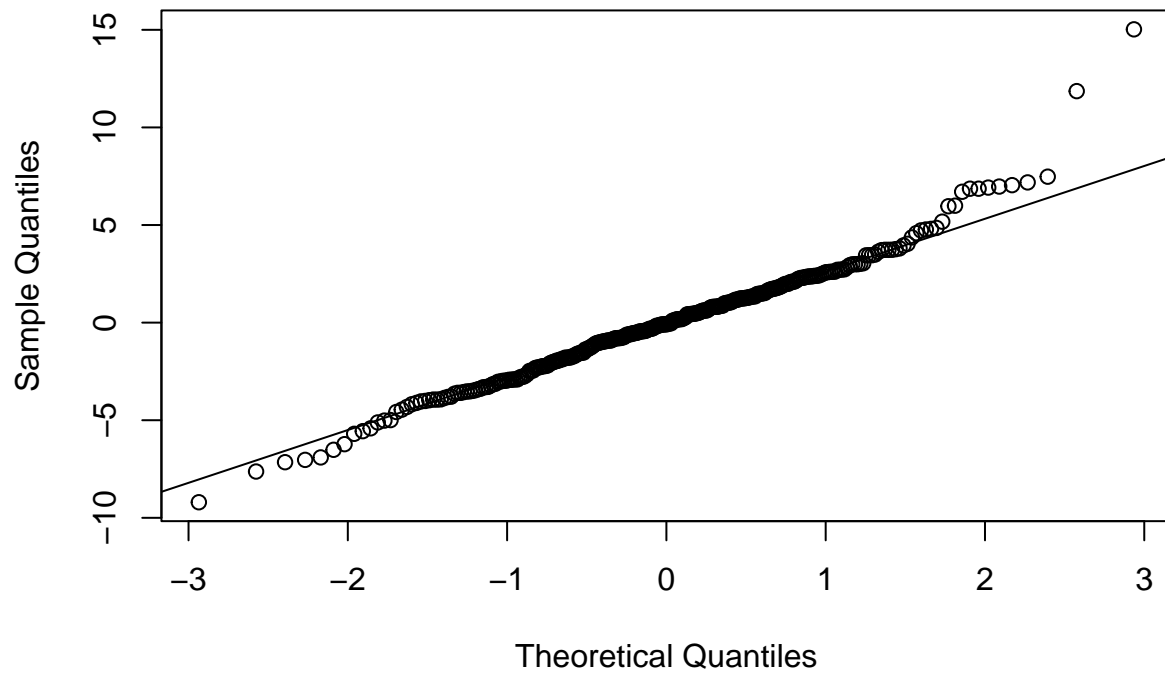
weight residual model for first instance



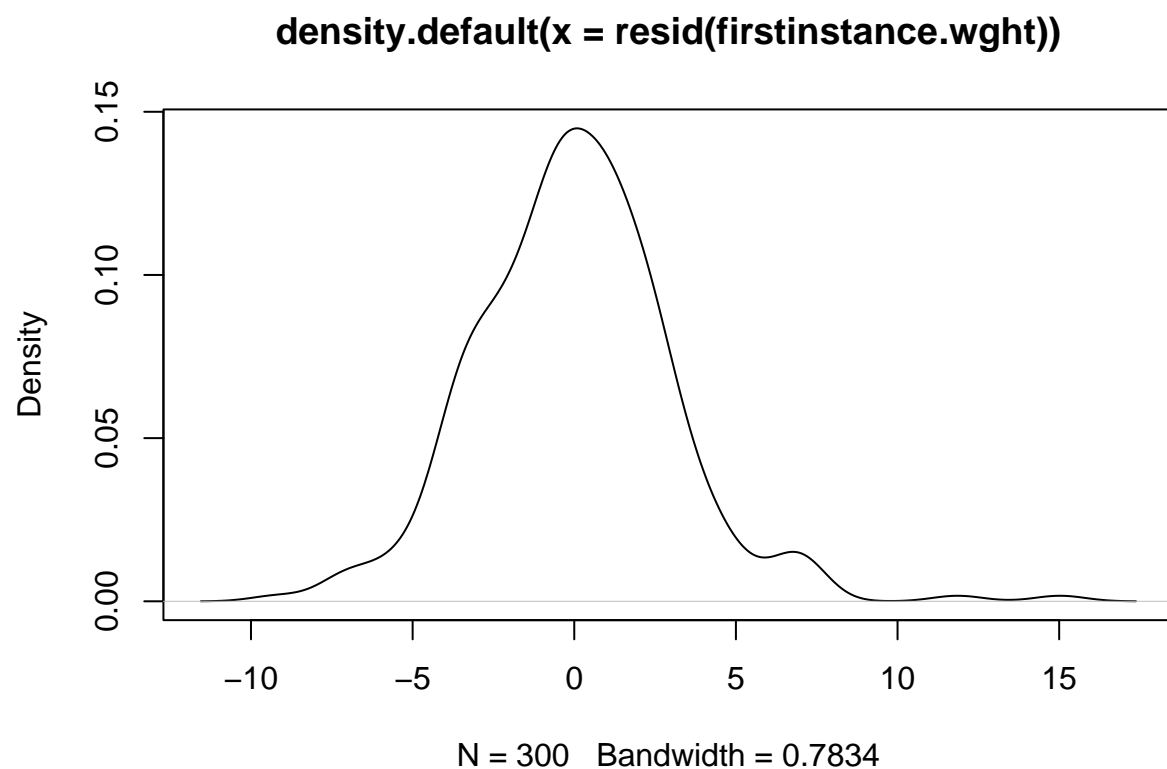
##linear Weight Models for first instance:

```
qqnorm(resid(firstinstance.wght))  
qqline(resid(firstinstance.wght))
```

Normal Q-Q Plot



```
plot(density(resid(firstinstance.wght)))
```



##histogram model weight for first instance:

```
hist(fisins_residual ,prob=T,breaks=20,main="Histogram model weight for first instance",xlab="Residuals",col="blue",lwd=3)
lines(density(fisins_residual ),col="blue",lwd=3)
```


Histogram model weight for first instance

