Acceleration

Min. 1st Qu. Median Mean 3rd Qu. Max.

8.00 13.82 15.50 15.57 17.17 24.80

> sd(cars\_num$acceleration)

[1] 2.757689

Initial Model\_0

> model<-lm(mpg~.,data=train)

> summary(model)

Call:

lm(formula = mpg ~ ., data = train)

Residuals:

Min 1Q Median 3Q Max

-7.7037 -1.6934 0.0002 1.4610 10.9744

Coefficients: (3 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) 33.1593810 2.5012372 13.257 < 2e-16 \*\*\*

displacement 0.0222577 0.0082516 2.697 0.00745 \*\*

horsepower 0.0018702 0.0072344 0.259 0.79622

acceleration 0.1876877 0.0816542 2.299 0.02234 \*

weight -0.0068233 0.0006447 -10.583 < 2e-16 \*\*\*

origin.1 -2.6641278 0.5993093 -4.445 1.31e-05 \*\*\*

origin.2 -0.3705687 0.6014590 -0.616 0.53837

origin.3 NA NA NA NA

model\_year.2003 9.2127271 1.0287479 8.955 < 2e-16 \*\*\*

model\_year.2004 8.1743719 0.9868678 8.283 6.75e-15 \*\*\*

model\_year.2005 10.7941269 1.0424496 10.355 < 2e-16 \*\*\*

model\_year.2006 6.8290339 0.9494733 7.192 7.00e-12 \*\*\*

model\_year.2007 4.7595240 0.9154362 5.199 4.10e-07 \*\*\*

model\_year.2008 4.4565680 0.9669937 4.609 6.40e-06 \*\*\*

model\_year.2009 2.4906622 0.9450018 2.636 0.00891 \*\*

model\_year.2010 2.4659620 0.9411837 2.620 0.00932 \*\*

model\_year.2011 2.8955452 0.9762143 2.966 0.00330 \*\*

model\_year.2012 0.4181033 0.8750693 0.478 0.63320

model\_year.2013 1.2182957 0.9959875 1.223 0.22238

model\_year.2014 1.8938593 1.0150178 1.866 0.06321 .

model\_year.2015 NA NA NA NA

cylinders.3 -3.6562297 2.3568450 -1.551 0.12206

cylinders.4 1.9875948 1.4298959 1.390 0.16573

cylinders.5 6.3978685 3.2778796 1.952 0.05205 .

cylinders.6 -0.5672377 0.8821621 -0.643 0.52079

cylinders.8 NA NA NA NA

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.845 on 256 degrees of freedom

Multiple R-squared: 0.8802, Adjusted R-squared: 0.8699

F-statistic: 85.49 on 22 and 256 DF, p-value: < 2.2e-16

Model\_1

Call:

lm(formula = mpg ~ displacement + acceleration + weight + origin.1 +

model\_year.2003 + model\_year.2004 + model\_year.2005 + model\_year.2006 +

model\_year.2007 + model\_year.2008 + model\_year.2009 + model\_year.2010 +

model\_year.2011 + model\_year.2014 + cylinders.3 + cylinders.6,

data = train)

Residuals:

Min 1Q Median 3Q Max

-8.619 -1.716 -0.146 1.549 10.722

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 36.1292710 1.5544814 23.242 < 2e-16 \*\*\*

displacement 0.0136575 0.0060863 2.244 0.025670 \*

acceleration 0.1962544 0.0810289 2.422 0.016113 \*

weight -0.0067403 0.0006296 -10.705 < 2e-16 \*\*\*

origin.1 -2.4397393 0.5169966 -4.719 3.86e-06 \*\*\*

model\_year.2003 8.8762986 0.8162189 10.875 < 2e-16 \*\*\*

model\_year.2004 7.6664722 0.7590456 10.100 < 2e-16 \*\*\*

model\_year.2005 10.5798719 0.8008326 13.211 < 2e-16 \*\*\*

model\_year.2006 6.2337149 0.7300727 8.538 1.12e-15 \*\*\*

model\_year.2007 4.1793503 0.6741262 6.200 2.19e-09 \*\*\*

model\_year.2008 3.8906623 0.7335303 5.304 2.41e-07 \*\*\*

model\_year.2009 1.8612625 0.7143738 2.605 0.009700 \*\*

model\_year.2010 1.9926826 0.7110919 2.802 0.005453 \*\*

model\_year.2011 2.2860453 0.7450926 3.068 0.002380 \*\*

model\_year.2014 1.5032456 0.8044760 1.869 0.062794 .

cylinders.3 -5.4587910 1.7524543 -3.115 0.002044 \*\*

cylinders.6 -1.7076024 0.4596033 -3.715 0.000248 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.846 on 262 degrees of freedom

Multiple R-squared: 0.8772, Adjusted R-squared: 0.8697

F-statistic: 117 on 16 and 262 DF, p-value: < 2.2e-16

>

> vif(model\_1)

displacement acceleration weight origin.1 model\_year.2003 model\_year.2004

13.740131 1.721951 9.950223 2.135160 1.456202 1.381199

model\_year.2005 model\_year.2006 model\_year.2007 model\_year.2008 model\_year.2009 model\_year.2010

1.401819 1.333428 1.413174 1.289902 1.329537 1.369255

model\_year.2011 model\_year.2014 cylinders.3 cylinders.6

1.330886 1.275434 1.125137 1.242894

Final Model

|  |
| --- |
| > model\_5<-lm(formula = mpg ~ weight +  + model\_year.2003 + model\_year.2005 + model\_year.2006 +  + cylinders.6,  + data = train)  > summary(model\_5)  Call:  lm(formula = mpg ~ weight + model\_year.2003 + model\_year.2005 +  model\_year.2006 + cylinders.6, data = train)  Residuals:  Min 1Q Median 3Q Max  -10.9908 -2.2409 -0.0889 2.0404 13.6727  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 43.0655339 0.8322913 51.743 < 2e-16 \*\*\*  weight -0.0068706 0.0002591 -26.522 < 2e-16 \*\*\*  model\_year.2003 5.4187995 0.8655506 6.261 1.48e-09 \*\*\*  model\_year.2005 8.2876229 0.8734230 9.489 < 2e-16 \*\*\*  model\_year.2006 3.5469252 0.7957273 4.457 1.21e-05 \*\*\*  cylinders.6 -1.4539851 0.5254038 -2.767 0.00604 \*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 3.551 on 273 degrees of freedom  Multiple R-squared: 0.8009, Adjusted R-squared: 0.7972  F-statistic: 219.6 on 5 and 273 DF, p-value: < 2.2e-16  >  > VIF(model\_5)  [1] 5.022119  > vif(model\_5)  weight model\_year.2003 model\_year.2005 model\_year.2006 cylinders.6  1.081940 1.051936 1.071158 1.017564 1.043398 |
|  |
| |  | | --- | | > | |

Final Model MAPE

> ##Training MAPE

> MAPE(train$mpg,pred1\_5)

[1] 0.1220894

> ##Testing MAPE

> MAPE(test$mpg,pred2\_5)

[1] 0.128095

Correlation

> cor(test$mpg,pred2\_5)

[1] 0.8809019

> cor(test$mpg,pred2\_5)^2

[1] 0.7759882

Diagnostics

> shapiro.test(model\_5$residuals)

Shapiro-Wilk normality test

data: model\_5$residuals

W = 0.97504, p-value = 8.543e-05