Fuel Consumption with Automatic versus Manual Transmissions

T. Bennett

August 23, 2015

Executive Summary

We estimate that a car with a manual transmission gets 2.7(95% CI x to y) more miles per gallon than a car with an automatic transmission after adjustment for vehicle weight, number of carburetors, and quarter mile time.

Data Processing

1) Loading packages

```
library(dplyr)
library(ggplot2)
library(stringr)
library(xtable)
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.2.2
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.2.2
```

```
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 3.2.2
```

2) Loading the raw data

```
## load and inspect the data
data(mtcars)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```r
glimpse(mtcars)
```

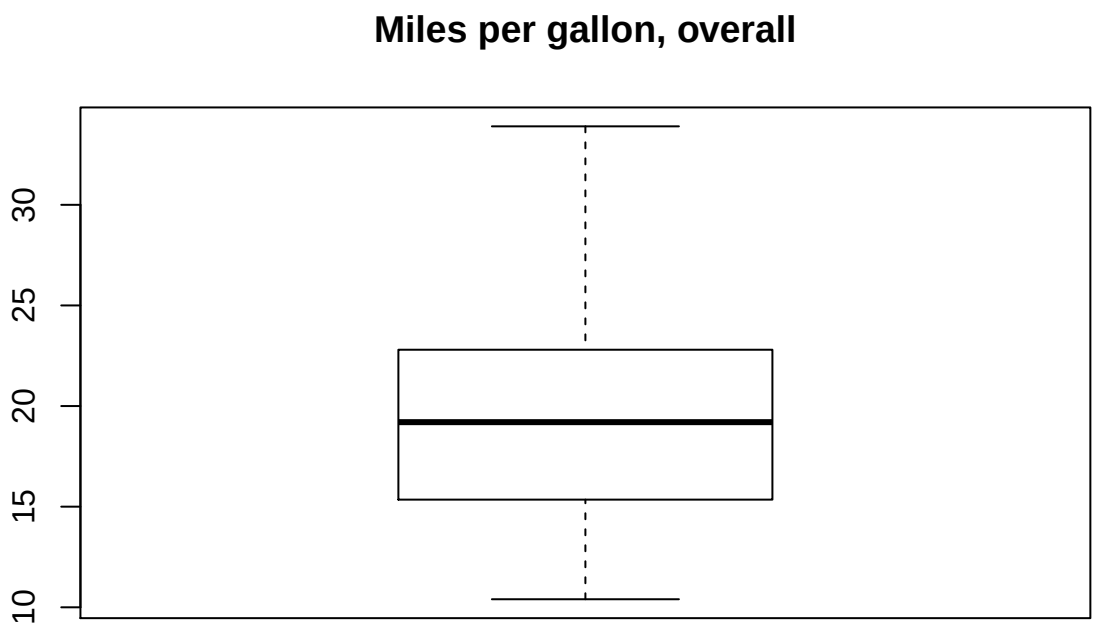
```
## Observations: 32
## Variables:
## $ mpg  (dbl) 21.0, 21.0, 22.8, 21.4, 18.7, 18.1, 14.3, 24.4, 22.8, 19....
## $ cyl  (dbl) 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 8, 4, 4, ...
## $ disp (dbl) 160.0, 160.0, 108.0, 258.0, 360.0, 225.0, 360.0, 146.7, 1...
## $ hp   (dbl) 110, 110, 93, 110, 175, 105, 245, 62, 95, 123, 123, 180, ...
## $ drat (dbl) 3.90, 3.90, 3.85, 3.08, 3.15, 2.76, 3.21, 3.69, 3.92, 3.9...
## $ wt   (dbl) 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.190, 3...
## $ qsec (dbl) 16.46, 17.02, 18.61, 19.44, 17.02, 20.22, 15.84, 20.00, 2...
## $ vs   (dbl) 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, ...
## $ am   (dbl) 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, ...
## $ gear (dbl) 4, 4, 4, 3, 3, 3, 3, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 4, 4, ...
## $ carb (dbl) 4, 4, 1, 1, 2, 1, 4, 2, 2, 4, 4, 3, 3, 3, 4, 4, 4, 1, 2, ...
```

This analysis is based on the `mtcars` dataset included in the `datasets` package in base R. It contains data on the 1973-74 models of 32 different automobiles including fuel consumption and 10 aspects of automobile design. Total 32 rows and 11 fields, no missing data.

Exploratory Data Analyses

Miles per gallon is the primary outcome.

```r
with(mtcars, boxplot(mpg, main = "Miles per gallon, overall"))
```



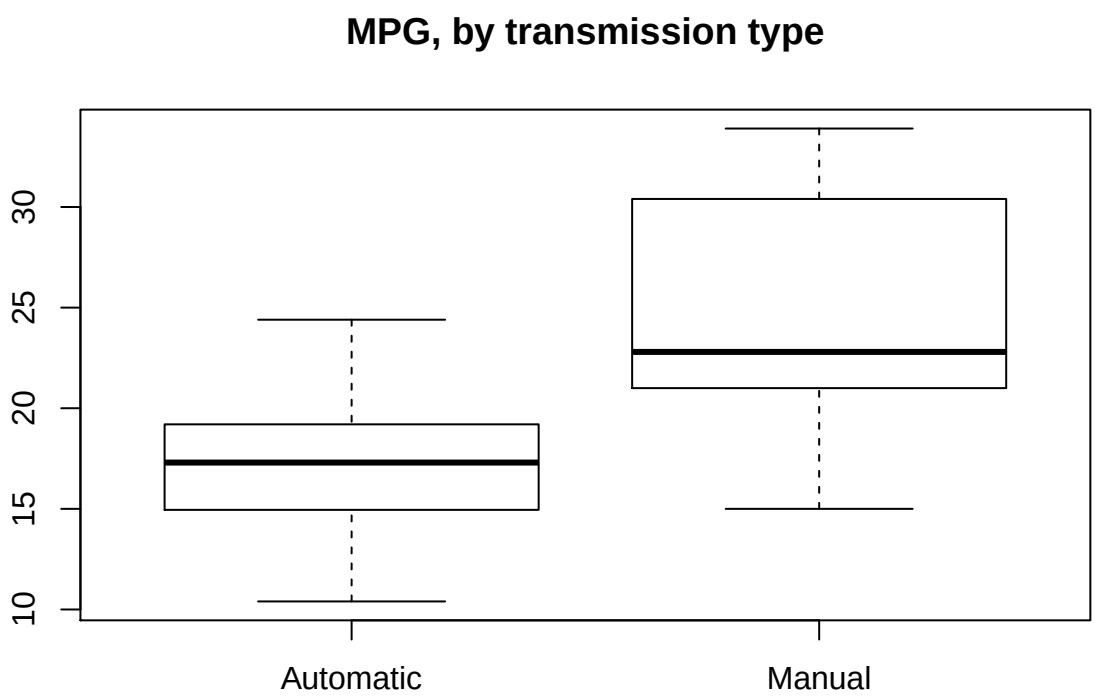
```
mtcars <- mtcars %>%
    mutate(trtype = factor(am, labels = c("Automatic", "Manual"))) %>%
    select(-am)

tab <- xtable(table(mtcars$trtype))
tab
```

% latex table generated in R 3.2.1 by xtable 1.7-4 package % Sat Aug 22 11:22:42 2015

	V1
Automatic	19
Manual	13

Transmission type is the primary predictor of interest. 13 of 32(0.40625) have a manual transmission and 19 of 32(0.59375) have a manual transmission.

```
boxplot(mpg ~ trtype, main = "MPG, by transmission type", data = mtcars)
```



A bivariate plot does suggest that manual transmissions are associated with better gas mileage.

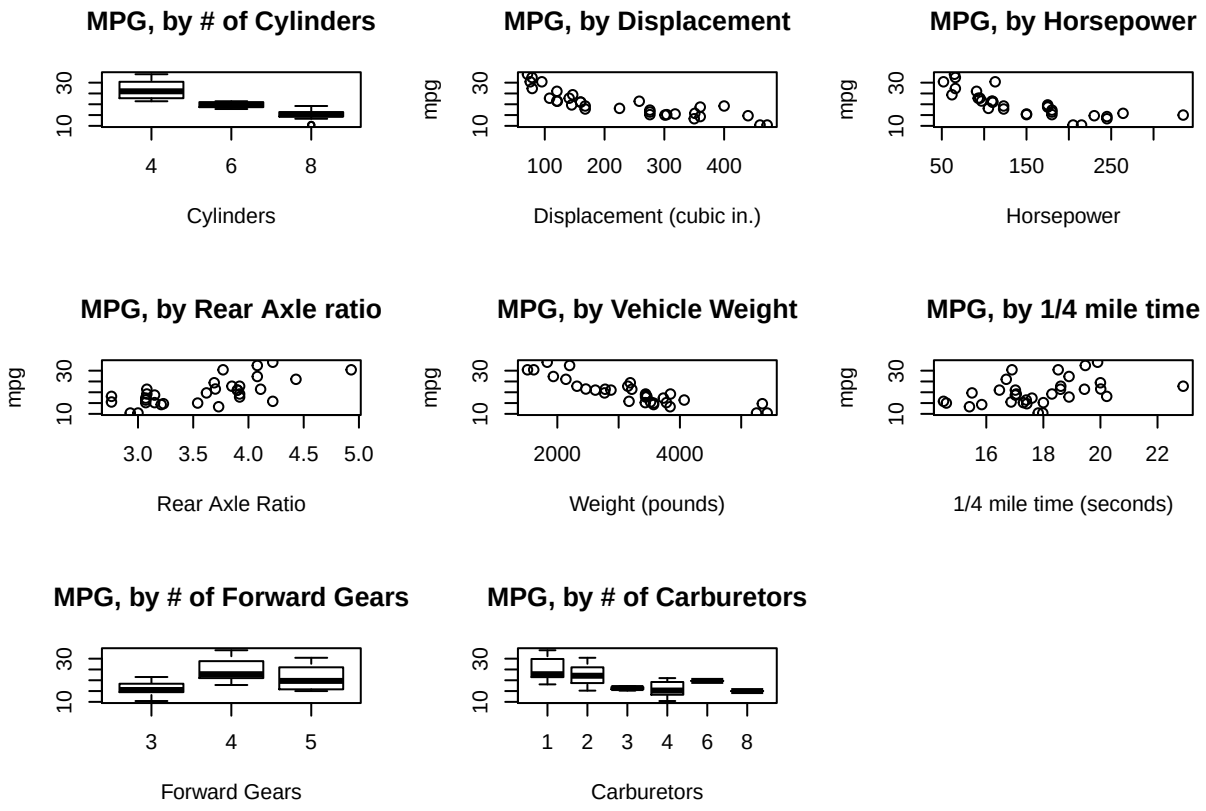
```
res <- t.test(mpg ~ trtype, var.equal=FALSE, data = mtcars)
```

A t-test of `mpg` using unequal variances bears that out: cars with automatic transmissions get 17.1473684 miles per gallon and cars with manual transmissions get 24.3923077 miles per gallon, t-test $p = 0.0013736$.

Potential Confounders

Several other variables might confound the Transmission Type - Mileage relationship. Based on my working knowledge of automobile engines, any of the other 9 variables in the `mtcars` dataset might be hypothesized to be important. The below bivariate plots of the 9 variables against `mpg` suggest that any of them could reasonably be a predictor.

```
par(mfrow = c(3,3))
boxplot(mpg ~ cyl, main = "MPG, by # of Cylinders", data = mtcars, xlab = "Cylinders")
with(mtcars, plot(disp, mpg, main = "MPG, by Displacement", xlab = "Displacement (cubic in.)"))
with(mtcars, plot(hp, mpg, main = "MPG, by Horsepower", xlab = "Horsepower"))
with(mtcars, plot(drat, mpg, main = "MPG, by Rear Axle ratio", xlab = "Rear Axle Ratio"))
with(mtcars, plot(wt*1000, mpg, main = "MPG, by Vehicle Weight", xlab = "Weight (pounds)"))
with(mtcars, plot(qsec, mpg, main = "MPG, by 1/4 mile time", xlab = "1/4 mile time (seconds)"))
boxplot(mpg ~ gear, main = "MPG, by # of Forward Gears", data = mtcars, xlab = "Forward Gears")
boxplot(mpg ~ carb, main = "MPG, by # of Carburetors", data = mtcars, xlab = "Carburetors")
par(mfrow = c(1,1))
```

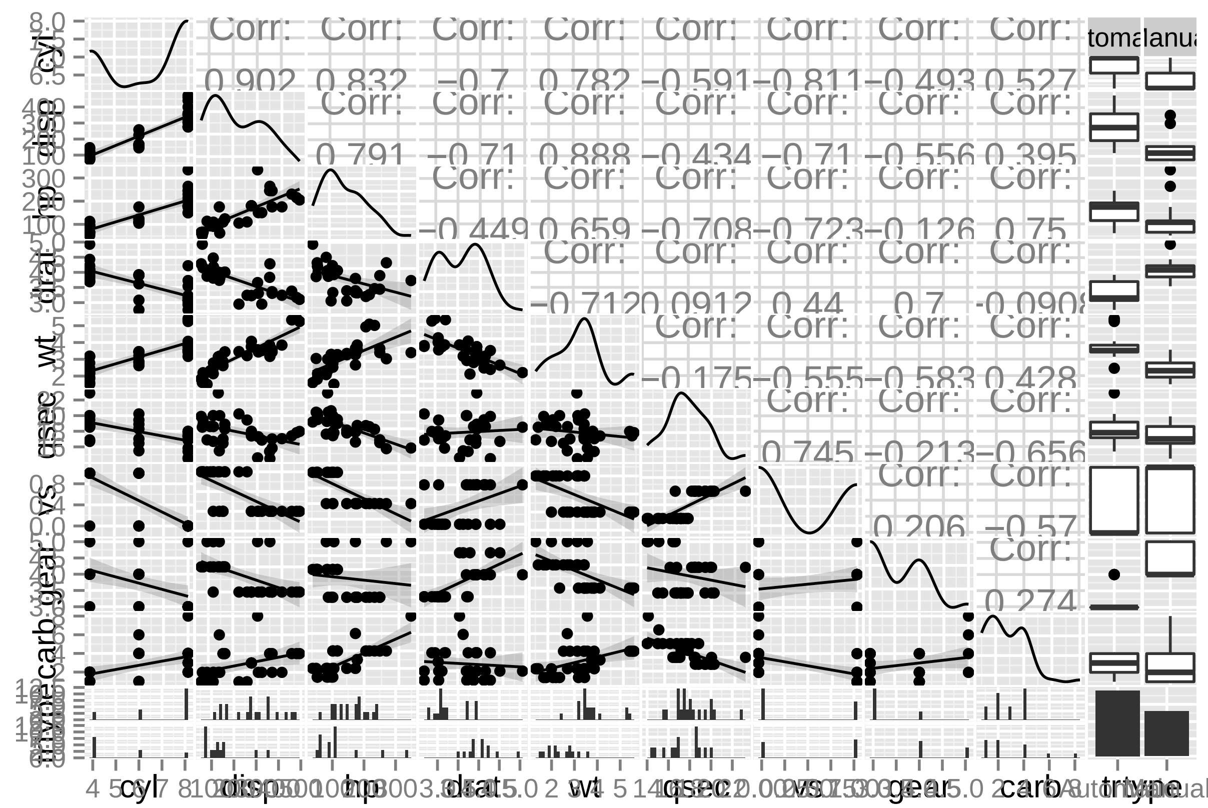


To avoid including two covariates that are either highly correlated themselves or highly correlated with the primary predictor of interest, I created the `pairs` plot below.

```
p <- ggpairs(mtcars, columns = 2:11, lower = list(continuous = "smooth"), params = c(method = "loess"))
p
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



Number of cylinders, weight, and displacement are highly correlated (which makes sense, I suppose). I decided to include displacement and weight in the model selection process because they are continuous variables that reflects physical properties of the car. Number of cylinders is also reasonably highly correlated with transmission type, our primary predictor of interest. Horsepower is also highly correlated with number of cylinders, but not quite as highly correlated with the other variables, so I will consider it during model selection.

Linear Regression including all potential predictors (Model 1)

```
##
## Call:
## lm(formula = mpg ~ I(1 * (trtype == "Manual")) + disp + drat +
##     wt + qsec + vs + gear + carb + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4286 -1.5908 -0.0412  1.2120  4.5961
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 10.96007   13.53030   0.810   0.4266
## I(1 * (trtype == "Manual"))  2.57743    1.94035   1.328   0.1977
## disp                         0.01283    0.01682   0.763   0.4538
## drat                         0.83520    1.53625   0.544   0.5921
## wt                          -3.69251    1.83954  -2.007   0.0572 .
## qsec                         0.84244    0.68678   1.227   0.2329
## vs                           0.38975    1.94800   0.200   0.8433
## gear                         0.71155    1.36562   0.521   0.6075
## carb                        -0.21958    0.78856  -0.278   0.7833
## hp                          -0.02191    0.02091  -1.048   0.3062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.59 on 22 degrees of freedom
## Multiple R-squared:  0.8689, Adjusted R-squared:  0.8153
## F-statistic: 16.21 on 9 and 22 DF,  p-value: 9.031e-08
```

In this full model, we estimate that cars with manual transmissions get 2.6(95% CI x to y) more miles per gallon than cars with automatic transmissions after adjustment for 8 other variables. However, this comparison is not statistically significant. Weight is the variable that appears to potentially be an important predictor of fuel consumption, as each 1000 pounds of additional weight is associated with 3.7(95% CI x to y) fewer miles per gallon, $p = []$. This model fits the data reasonably well, as the adjusted R^2 is [0.82].

Assessing variance inflation

```
vif(fit)
```

```
## I(1 * (trtype == "Manual"))                         disp
##                    4.332286                   20.088643
##                        drat                           wt
##                    3.118062                   14.971795
##                        qsec                           vs
##                    6.960353                    4.454935
##                        gear                         carb
##                    4.691536                    7.497054
##                          hp
##                    9.499795
```

Our concern about the correlation between weight, horsepower, and displacement appears well-founded, as those variables have the highest variance inflation factors. Horsepower and displacement were poor predictors of fuel consumption in the full model, therefore we will remove those variables and re-fit the model.

Model without horsepower or displacement (Model 2)

```
fit2 <- lm(mpg ~ I(1 * (trtype == 'Manual')) + drat + wt + qsec + vs + gear + carb, data = mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ I(1 * (trtype == "Manual")) + drat + wt +
##     qsec + vs + gear + carb, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9187 -1.1587 -0.1858  1.3021  4.3141
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    8.4612    10.5891   0.799  0.43210
## I(1 * (trtype == "Manual"))    2.5377     1.8883   1.344  0.19155
## drat                           1.0565     1.4897   0.709  0.48504
## wt                            -2.9502     1.0543  -2.798  0.00997 **
## qsec                           0.8955     0.5198   1.723  0.09782 .
## vs                            -0.1033     1.8548  -0.056  0.95605
## gear                           0.6730     1.3400   0.502  0.62006
## carb                          -0.7573     0.5530  -1.370  0.18350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.542 on 24 degrees of freedom
## Multiple R-squared:  0.8623, Adjusted R-squared:  0.8221
## F-statistic: 21.46 on 7 and 24 DF,  p-value: 6.989e-09
```

```
vif(fit2)
```

```
## I(1 * (trtype == "Manual"))                         drat
##                    4.258479                    3.043073
##                          wt                         qsec
##                    5.104823                    4.139107
##                          vs                         gear
##                    4.191818                    4.688164
##                        carb
##                    3.826243
```

Model 2 still fits the data reasonably well (adjusted R^2 is the same, [0.82]), but the variance inflation factors for the variables remaining in the model are now similar. The estimated effect of a manual transmission has changed only marginally (2.5 mpg higher instead of 2.6) and continues to be not statistically significant. The weight effect is slightly smaller (3.0 instead of 3.7), but is now statistically significant. This model is a candidate for a final model.

Highly parsimonious model (Model 3)

Because vehicle weight was the only significant predictor in either previous model, I considered a model containing only transmission type and vehicle weight.

```
fit3 <- lm(mpg ~ I(1 * (trtype == 'Manual')) + wt, data = mtcars)
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ I(1 * (trtype == "Manual")) + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 37.32155    3.05464  12.218 5.84e-13 ***
## I(1 * (trtype == "Manual")) -0.02362    1.54565  -0.015    0.988
## wt                          -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

```
lrtest(fit2,fit3)
```

```
## Likelihood ratio test
##
## Model 1: mpg ~ I(1 * (trtype == "Manual")) + drat + wt + qsec + vs + gear +
##     carb
## Model 2: mpg ~ I(1 * (trtype == "Manual")) + wt
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   9 -70.661
## 2   4 -80.015 -5 18.708   0.002178 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This model is likely too parsimonious, as the adjusted R^2 decreased to [0.74]. As you might expect, a nested likelihood ratio (LR) test comparing the highly parsimonious model to the model without horsepower or displacement was significant, $p = 0.002$. Interestingly, transmission type has no association with fuel consumption in this model.

Adding variables back to the parsimonious model (Model 4)

Of the remaining variables, quarter mile time (perhaps a proxy for several vehicle and tuning properties) and number of carburetors were the closest to significance in the previous models. Therefore, we fit a model that added them back to the parsimious model.

```r
fit4 <- lm(mpg ~ I(1 * (trtype == 'Manual')) + wt + carb + qsec, data = mtcars)
summary(fit4)
```

```
##
## Call:
## lm(formula = mpg ~ I(1 * (trtype == "Manual")) + wt + carb +
##     qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1184 -1.5414 -0.1392  1.2917  4.3604
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   12.8972     7.4725   1.726 0.095784 .
## I(1 * (trtype == "Manual"))    3.5114     1.4875   2.361 0.025721 *
## wt                            -3.4343     0.8200  -4.188 0.000269 ***
## carb                          -0.4886     0.4212  -1.160 0.256212
## qsec                           1.0191     0.3378   3.017 0.005507 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.444 on 27 degrees of freedom
## Multiple R-squared:  0.8568, Adjusted R-squared:  0.8356
## F-statistic: 40.39 on 4 and 27 DF,  p-value: 5.064e-11
```

```r
lrtest(fit3,fit4)
```

```
## Likelihood ratio test
##
## Model 1: mpg ~ I(1 * (trtype == "Manual")) + wt
## Model 2: mpg ~ I(1 * (trtype == "Manual")) + wt + carb + qsec
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   4 -80.015
## 2   6 -71.282  2 17.466  0.0001612 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lrtest(fit4,fit2)
```

```
## Likelihood ratio test
##
## Model 1: mpg ~ I(1 * (trtype == "Manual")) + wt + carb + qsec
## Model 2: mpg ~ I(1 * (trtype == "Manual")) + drat + wt + qsec + vs + gear +
##     carb
##   #Df  LogLik Df Chisq Pr(>Chisq)
## 1   6 -71.282
## 2   9 -70.661  3 1.242     0.7429
```

```r
vif(fit4)
```

```
## I(1 * (trtype == "Manual"))                      wt
##                    2.859625                3.341631
##                        carb                    qsec
##                    2.402649                1.890918
```

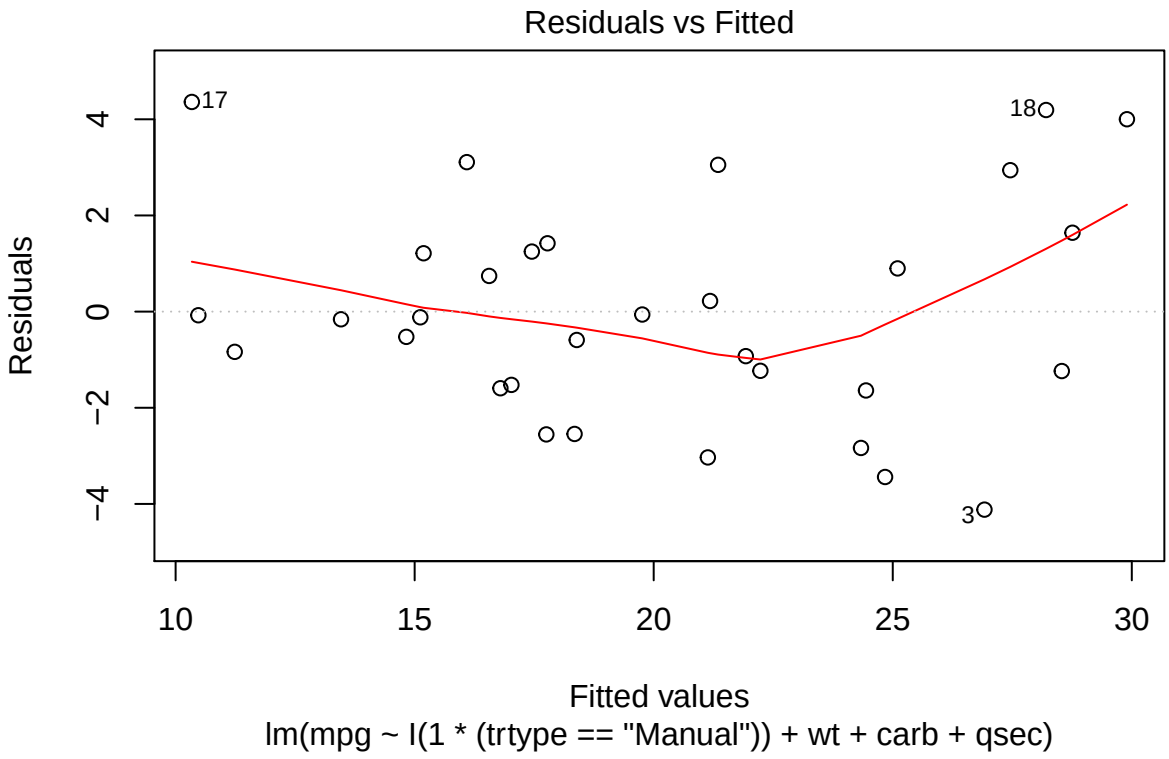
Model 4 has the best adjusted R^2 of any considered thus far ([0.84)]). It also has no difference by LR test from Model 2, which contained several more variables. The variance inflation factors of Model 4 are the lowest of any model evaluated thus far. Model 4 will the basis for our final model.

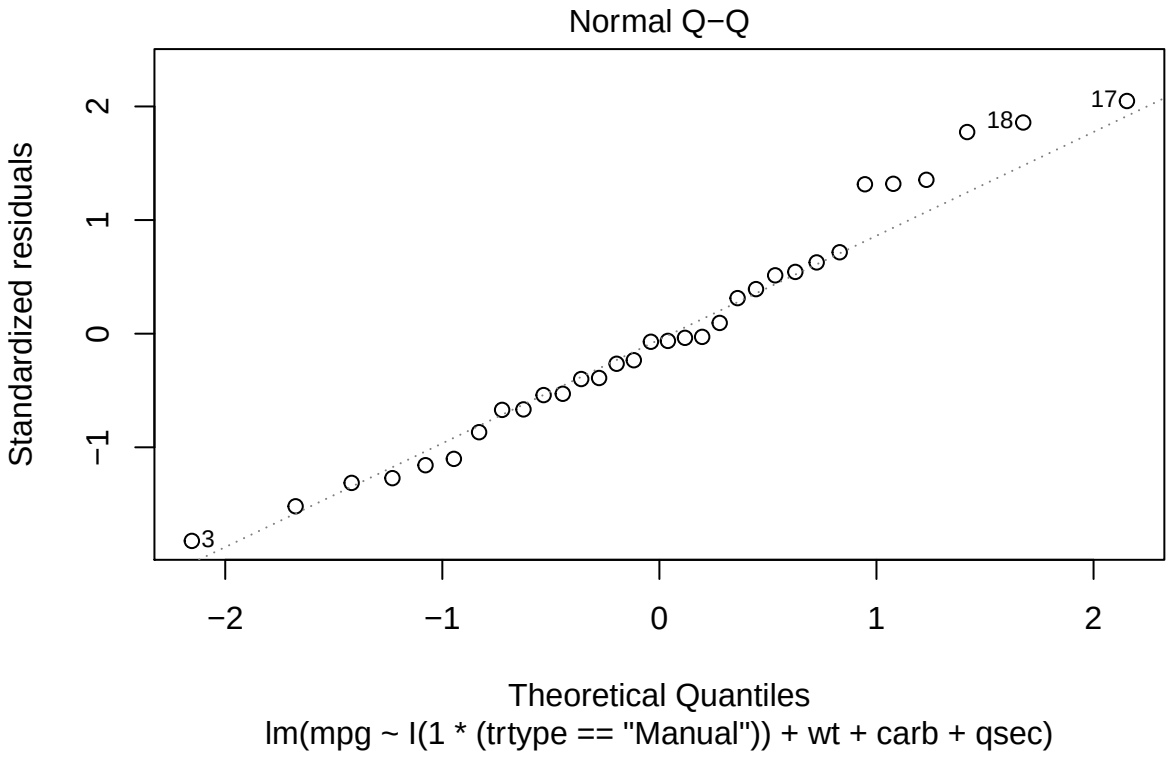
Estimates from Model 4

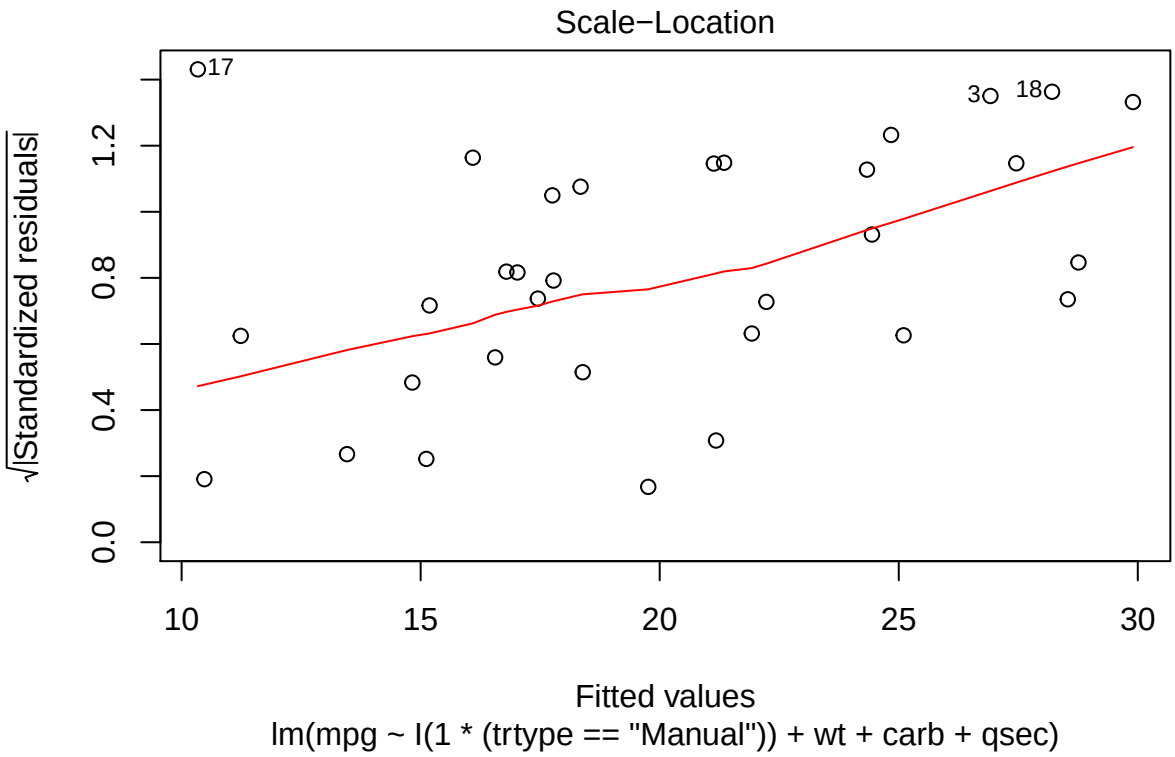
We estimate that a car with a manual transmission gets 3.5(95% CI x to y) more miles per gallon than a car with an automatic transmission after adjustment for vehicle weight, number of carburetors, and quarter mile time.

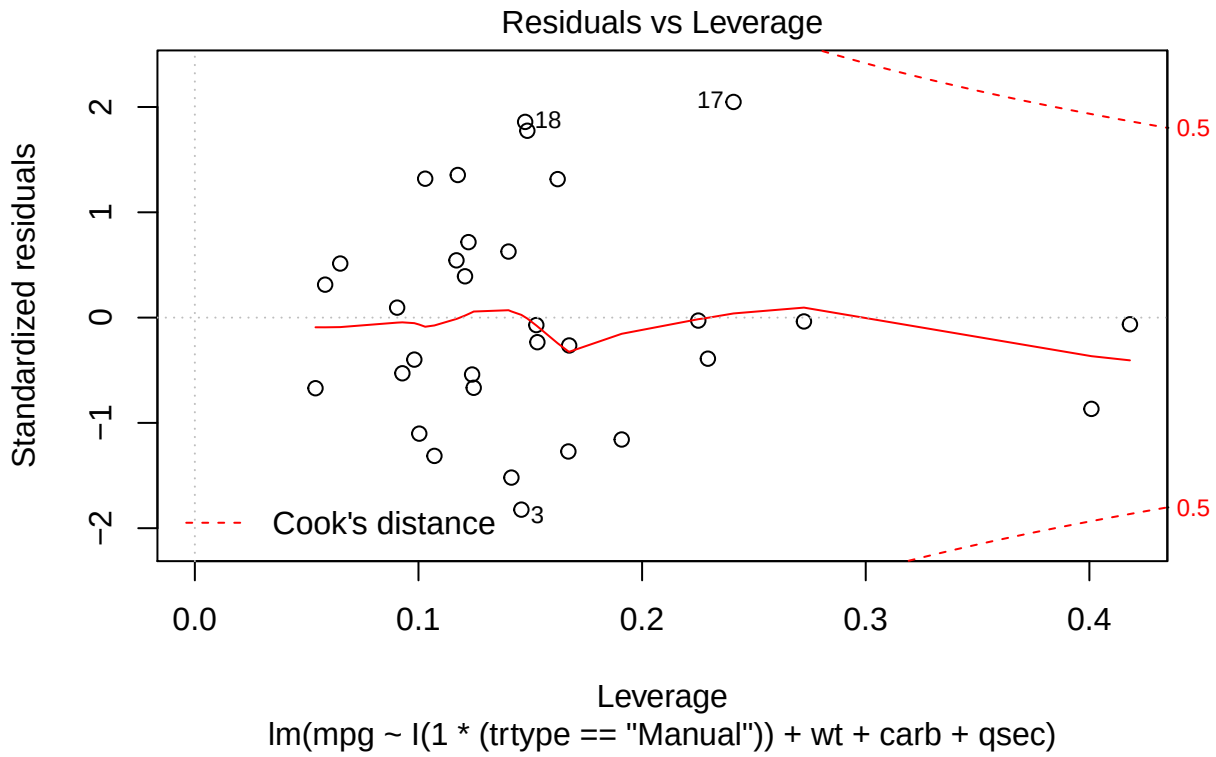
Evaluation of Residuals

```
residplot <- plot(fit4)
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ I(1 * (trtype == "Manual")) + wt + carb + qsec)
17
18
3
10
15
20
25
30
4
2
0
−2
−4


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ I(1 * (trtype == "Manual")) + wt + carb + qsec)
17
18
3
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
17
3
18
10
15
20
25
30
Fitted values
lm(mpg ~ I(1 * (trtype == "Manual")) + wt + carb + qsec)




```
residplot
```

```
## NULL
```

The residual plot suggests that our final model fits the data reasonably well. The Residuals vs Fitted panel shows that, as might be expected in such a small data set, the model does not fit as well where the data are sparse, i.e. for very low or very high fuel consumption cars. Points 3, 17, and 18 have residuals that are 2 standard deviations from the regression line and have at least medium leverage. Point 17, in particular (see the Scale-Location plot), might be quite influential to the fit.

Diagnostics

```
round(dfbetas(fit4), 3) ###
```

```
##    (Intercept) I(1 * (trtype == "Manual"))     wt   carb   qsec
## 1       -0.010                      -0.051  0.019 -0.046  0.011
## 2        0.037                      -0.073 -0.024 -0.034 -0.026
## 3        0.050                      -0.519 -0.313  0.465  0.010
## 4        0.004                      -0.009 -0.002 -0.011  0.000
## 5        0.149                      -0.095 -0.034 -0.089 -0.135
## 6        0.129                       0.010 -0.100  0.139 -0.162
## 7       -0.066                       0.071  0.051 -0.020  0.053
## 8       -0.130                      -0.201 -0.155  0.147  0.238
```

```
## 9        0.515                      0.073  0.101 -0.359 -0.633
## 10      -0.024                     -0.162 -0.147  0.189  0.082
## 11       0.032                      0.064  0.061 -0.093 -0.058
## 12       0.032                     -0.024  0.036 -0.033 -0.040
## 13       0.026                     -0.041 -0.013  0.003 -0.018
## 14      -0.015                      0.071  0.015 -0.021  0.001
## 15       0.104                     -0.099 -0.168  0.021 -0.055
## 16       0.010                     -0.011 -0.018  0.004 -0.005
## 17      -0.449                      0.592  1.029 -0.266  0.131
## 18      -0.252                      0.518  0.255 -0.300  0.235
## 19       0.013                      0.003 -0.132  0.046  0.042
## 20      -0.253                      0.324  0.006 -0.097  0.322
## 21      -0.098                      0.391  0.374 -0.078 -0.083
## 22      -0.189                      0.106  0.024  0.127  0.180
## 23      -0.262                      0.186  0.067  0.157  0.231
## 24      -0.022                      0.017  0.009  0.000  0.020
## 25       0.296                     -0.085  0.113 -0.326 -0.325
## 26       0.006                     -0.090 -0.015  0.079 -0.014
## 27       0.075                      0.036 -0.007 -0.066 -0.075
## 28       0.351                     -0.106 -0.329 -0.058 -0.250
## 29      -0.256                     -0.232 -0.167  0.197  0.359
## 30       0.000                      0.000  0.005 -0.010 -0.001
## 31       0.011                     -0.005  0.007 -0.039 -0.009
## 32       0.253                     -0.530 -0.367  0.221 -0.164
```

```
round(hatvalues(fit4), 3)
```

```
##     1     2     3     4     5     6     7     8     9    10    11    12
## 0.093 0.098 0.146 0.090 0.117 0.107 0.153 0.103 0.401 0.140 0.167 0.065
##    13    14    15    16    17    18    19    20    21    22    23    24
## 0.058 0.054 0.229 0.272 0.241 0.148 0.122 0.149 0.167 0.125 0.100 0.153
##    25    26    27    28    29    30    31    32
## 0.118 0.124 0.121 0.162 0.191 0.225 0.418 0.142
```

```
round(cooks.distance(fit4), 3)
```

```
##     1     2     3     4     5     6     7     8     9    10    11    12
## 0.006 0.003 0.114 0.000 0.008 0.041 0.002 0.040 0.101 0.013 0.003 0.004
##    13    14    15    16    17    18    19    20    21    22    23    24
## 0.001 0.005 0.009 0.000 0.266 0.120 0.014 0.110 0.065 0.013 0.027 0.000
##    25    26    27    28    29    30    31    32
## 0.049 0.008 0.004 0.067 0.063 0.000 0.001 0.076
```

```
round(dffits(fit4), 3)
```

```
##      1      2      3      4      5      6      7      8      9     10
## -0.167 -0.130 -0.790  0.029  0.195 -0.461 -0.098  0.453 -0.706  0.250
##     11     12     13     14     15     16     17     18     19     20
## -0.117  0.134  0.077 -0.158 -0.209 -0.022  1.232  0.813  0.265  0.774
##     21     22     23     24     25     26     27     28     29     30
## -0.576 -0.249 -0.370 -0.030  0.502 -0.201  0.143  0.587 -0.566 -0.015
##     31     32
## -0.053 -0.633
```

Point 17 does appear to be quite influential, as it has by far the largest Cook's Distance and the largest `dfbeta` and `dffit` value of any point. We will evaluate the model with that point removed.

Evaluating fit with point 17 removed

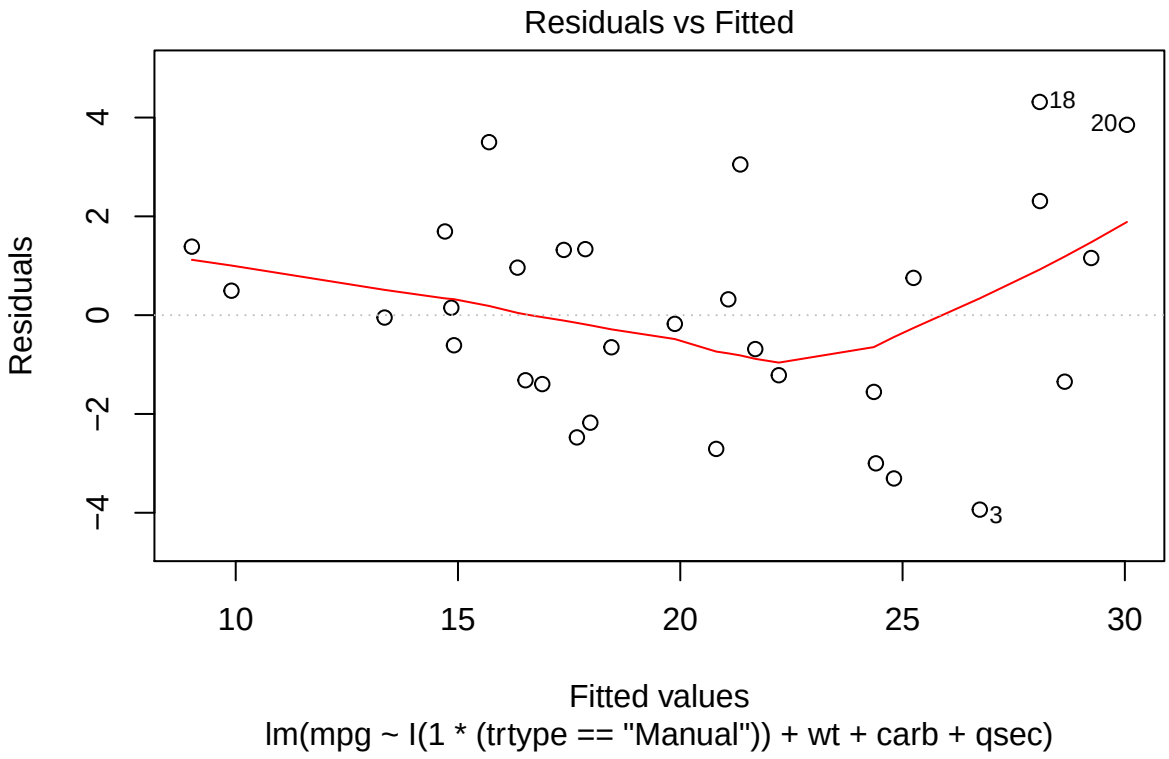
```
mtcarssub <- mtcars[-17,]
fit5 <- lm(mpg ~ I(1 * (trtype == 'Manual')) + wt + carb + qsec, data = mtcarssub)
summary(fit5)
```

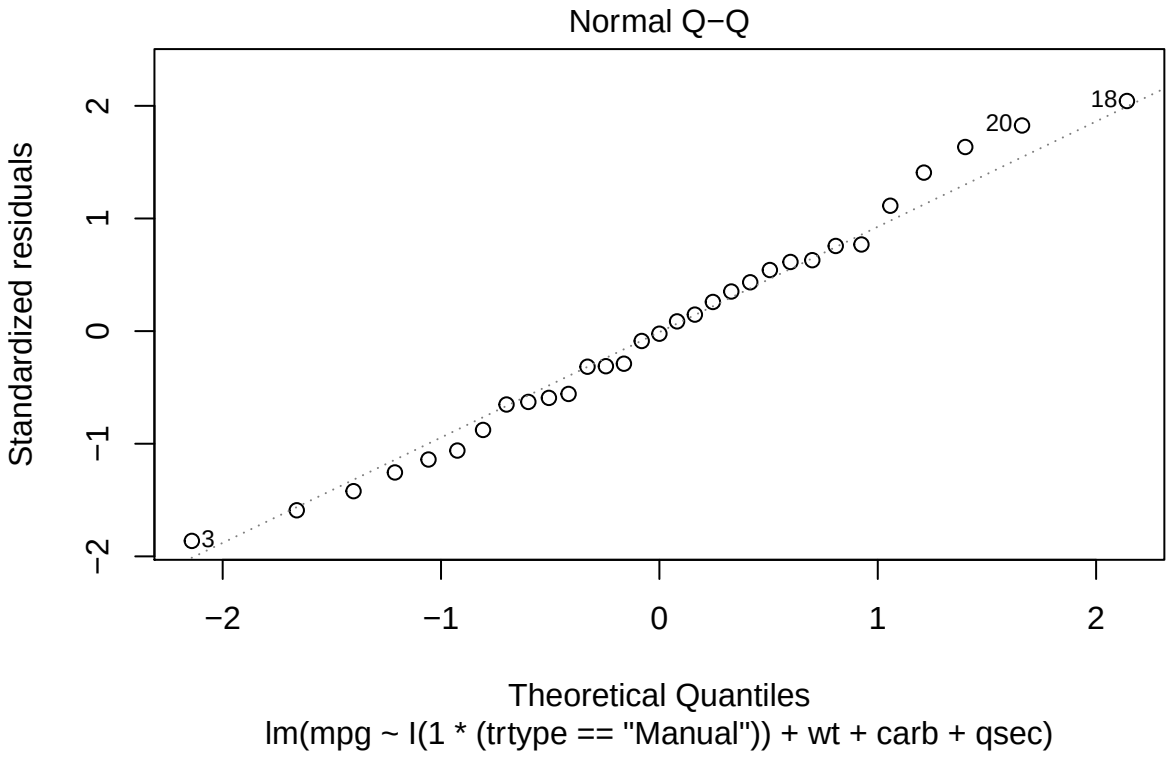
```
##
## Call:
## lm(formula = mpg ~ I(1 * (trtype == "Manual")) + wt + carb +
##     qsec, data = mtcarssub)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9367 -1.3710 -0.0486  1.3280  4.3155
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  16.0378     7.1444   2.245  0.03351 *
## I(1 * (trtype == "Manual"))   2.6871     1.4432   1.862  0.07396 .
## wt                           -4.2245     0.8488  -4.977 3.57e-05 ***
## carb                         -0.3836     0.3974  -0.965  0.34328
## qsec                          0.9778     0.3169   3.085  0.00478 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.289 on 26 degrees of freedom
## Multiple R-squared:  0.8757, Adjusted R-squared:  0.8566
## F-statistic: 45.81 on 4 and 26 DF,  p-value: 2.087e-11
```

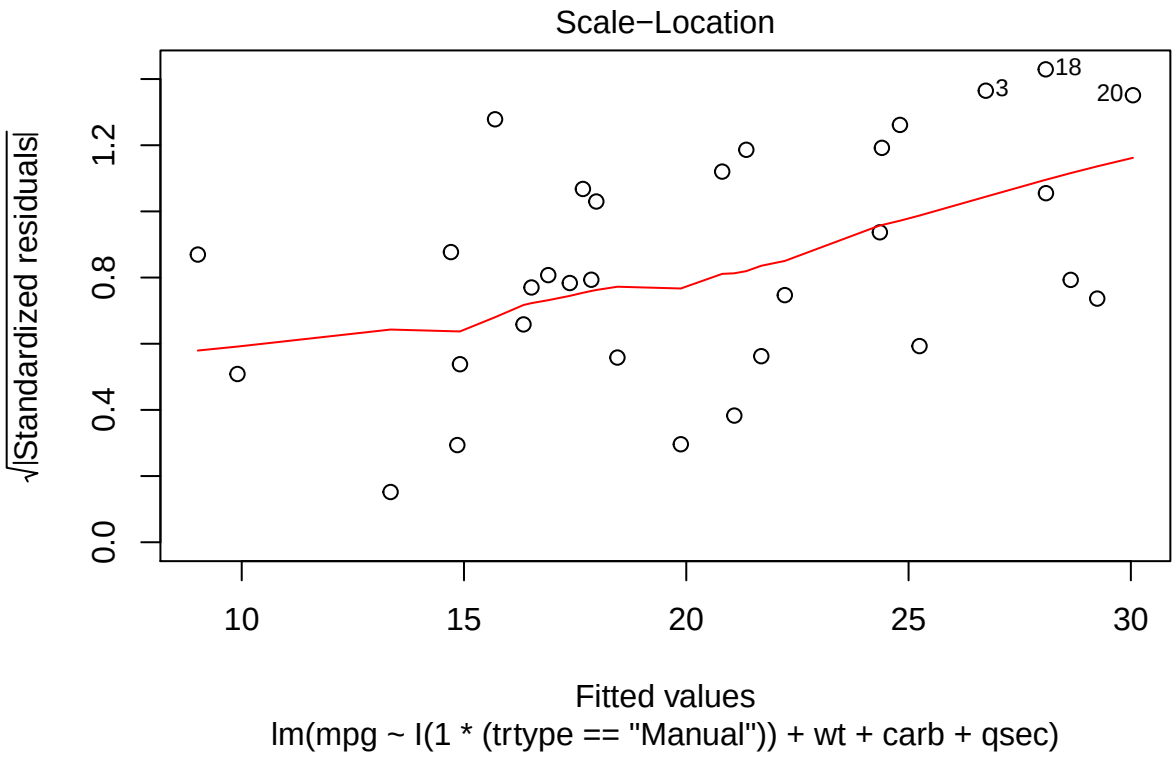
```
vif(fit5)
```

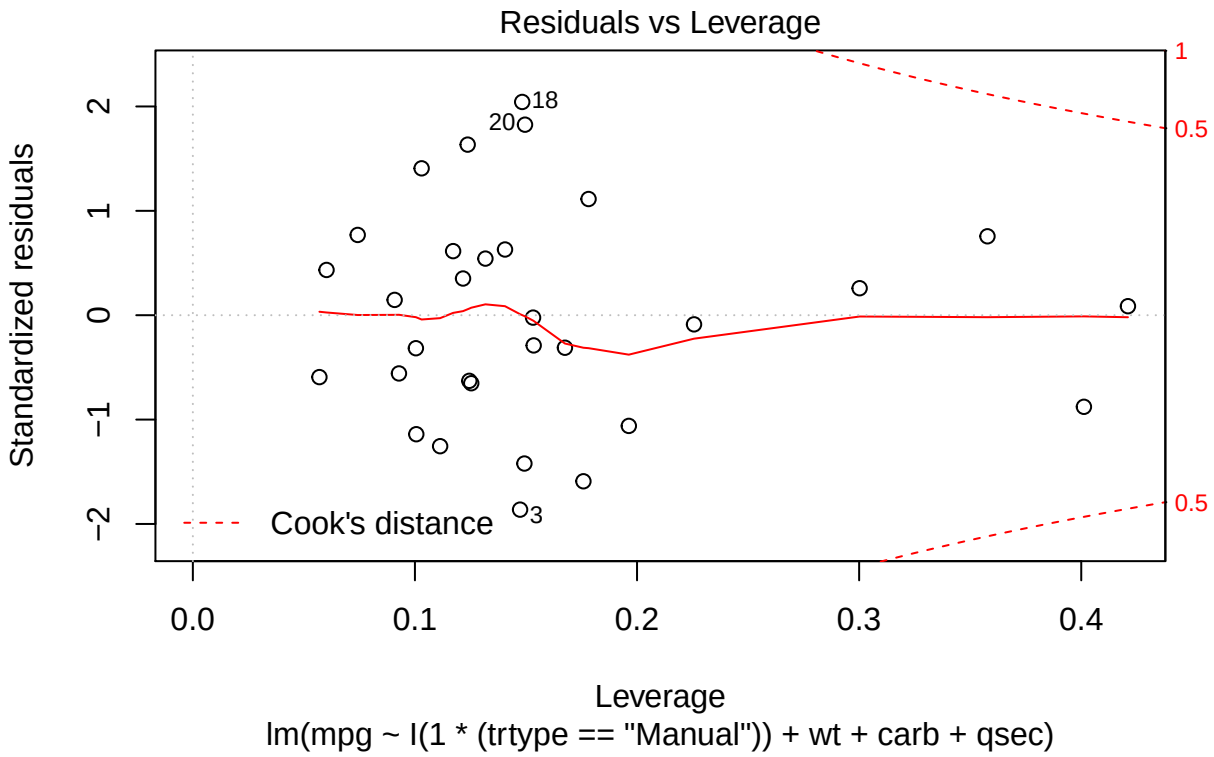
```
## I(1 * (trtype == "Manual"))                          wt
##                    3.001281                    3.438842
##                        carb                        qsec
##                    2.394347                    1.894040
```

```
plot(fit5)
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ I(1 * (trtype == "Manual")) + wt + carb + qsec)
18
20
3
10
15
20
25
30
−4
−2
0
2
4


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ I(1 * (trtype == "Manual")) + wt + carb + qsec)
18
20
3
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
10
15
20
25
30
18
3
20
Fitted values
lm(mpg ~ I(1 * (trtype == "Manual")) + wt + carb + qsec)



The residual plots suggest a better fit point 17 removed. This will be our final model.

Estimates from Final Model

We estimate that a car with a manual transmission gets 2.7(95% CI x to y) more miles per gallon than a car with an automatic transmission after adjustment for vehicle weight, number of carburetors, and quarter mile time.