Untitled

Per Idenfeldt, Oliver Grahn Thuna, Daniel Berg, Gabriel Junhager

9/25/2019

Introduktion

Variabelselektion

Vi börjar med att undersöka data som är icke-kategorisk, annat data undersöks senare.

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

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

```
## Loading required package: carData
```

```
##
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
##
## Attaching package: 'psych'
```

```
## The following object is masked from 'package:car':
##
##     logit
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

Variabler som helt klart är irrelevanta till bränsleförbrukning utesluts också automatiskt, till och exempel standard på krockkudde.

Vektorerna V7 och V8 står för hur många miles man kommer per gallon i stad respektive motorväg. Vi misstänker att vi kommer kunna kombinera dem i en variabel, hur ser de ut om vi plottar dem mot varandra?
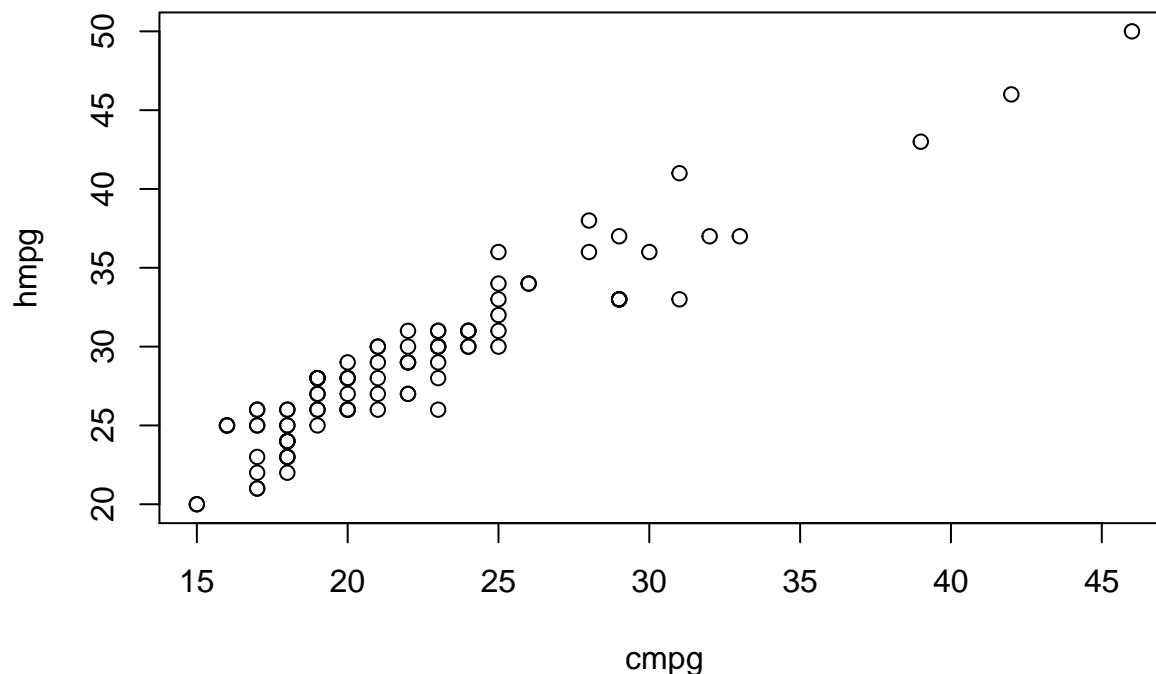
Figur 1: Plot mellan city miles per gallon och highway miles per gallon

Vi ser en klar linjär trend. Korrelationen som visas nedan verkar också relativt hög.

```
## [1] 0.9439358
```

Vi kombinerar helt enkelt dessa variabler istället för att göra en modell åt varje, även fast de kan ha mindre skillnader.

Nu gör vi en korrelationsmatris utav dessa numeriska variabler.

```
##                  V4            V5            V6          V11          V12
## V4    1.00000000   0.970601402   0.90675608    0.6209279    0.6454877
## V5    0.97060140   1.000000000   0.98158027    0.5898625    0.5974254
## V6    0.90675608   0.981580272   1.00000000    0.5410688    0.5350120
## V11   0.62092789   0.589862485   0.54106884    1.0000000    0.8740665
## V12   0.64548767   0.597425392   0.53501197    0.8740665    1.0000000
## V13   0.80244412   0.788217578   0.74444475    0.6844695    0.7321197
## V14  -0.04259816  -0.004954931   0.02501478   -0.4392286   -0.5478978
## V15  -0.47039499  -0.426395113  -0.37402421   -0.7383795   -0.8240086
## V17   0.63536902   0.619479981   0.58129439    0.6673762    0.7593062
## V18   0.06123644   0.057860074   0.05321592    0.4014559    0.3727212
## V19   0.55385881   0.503628440   0.44293341    0.6816244    0.7802831
## V20   0.51675786   0.500864163   0.46750079    0.6822698    0.7324842
## V21   0.49287830   0.456027866   0.40841435    0.7825653    0.8671102
## V25   0.66655377   0.647179005   0.60514157    0.8007884    0.8450753
## mpg  -0.61059800  -0.586337687  -0.54317994   -0.6576892   -0.6790722
##                 V13           V14          V15          V17          V18
## V4    0.802444116  -0.042598158   -0.4703950    0.6353690   0.061236438
```

```
## V5    0.788217578 -0.004954931 -0.4263951   0.6194800   0.057860074
## V6    0.744444746  0.025014782 -0.3740242   0.5812944   0.053215917
## V11   0.684469459 -0.439228560 -0.7383795   0.6673762   0.401455913
## V12   0.732119730 -0.547897805 -0.8240086   0.7593062   0.372721168
## V13   1.000000000  0.036688212 -0.6003139   0.7117903   0.009263668
## V14   0.036688212  1.000000000  0.4947642  -0.3333452  -0.467137627
## V15  -0.600313870  0.494764211  1.0000000  -0.6097098  -0.334975577
## V17   0.711790317 -0.333345218 -0.6097098   1.0000000   0.472095108
## V18   0.009263668 -0.467137627 -0.3349756   0.4720951   1.000000000
## V19   0.550864666 -0.441249316 -0.6902333   0.6904612   0.485294130
## V20   0.486854213 -0.467812289 -0.6368238   0.7576745   0.694054395
## V21   0.644413421 -0.539721132 -0.7804604   0.7987190   0.489978637
## V25   0.738797516 -0.427931473 -0.7352642   0.8940181   0.553272980
## mpg  -0.655795351  0.343757989  0.6523034  -0.8113934  -0.447278629
##              V19        V20        V21        V25        mpg
## V4    0.5538588  0.5167579  0.4928783   0.6665538 -0.6105980
## V5    0.5036284  0.5008642  0.4560279   0.6471790 -0.5863377
## V6    0.4429334  0.4675008  0.4084144   0.6051416 -0.5431799
## V11   0.6816244  0.6822698  0.7825653   0.8007884 -0.6576892
## V12   0.7802831  0.7324842  0.8671102   0.8450753 -0.6790722
## V13   0.5508647  0.4868542  0.6444134   0.7387975 -0.6557954
## V14  -0.4412493 -0.4678123 -0.5397211  -0.4279315  0.3437580
## V15  -0.6902333 -0.6368238 -0.7804604  -0.7352642  0.6523034
## V17   0.6904612  0.7576745  0.7987190   0.8940181 -0.8113934
## V18   0.4852941  0.6940544  0.4899786   0.5532730 -0.4472786
## V19   1.0000000  0.8236504  0.8221479   0.8062743 -0.6148637
## V20   0.8236504  1.0000000  0.8072134   0.8718953 -0.6511107
## V21   0.8221479  0.8072134  1.0000000   0.8749605 -0.6912520
## V25   0.8062743  0.8718953  0.8749605   1.0000000 -0.8391626
## mpg  -0.6148637 -0.6511107 -0.6912520  -0.8391626  1.0000000
```

Figur 2: Korrelationsmatris på data som endast är numerisk och relevant

Vi säger arbiträrt att vi vill testa alla variabler som fick $|r| < 0.5$, genom att plotta dem mot mpg.
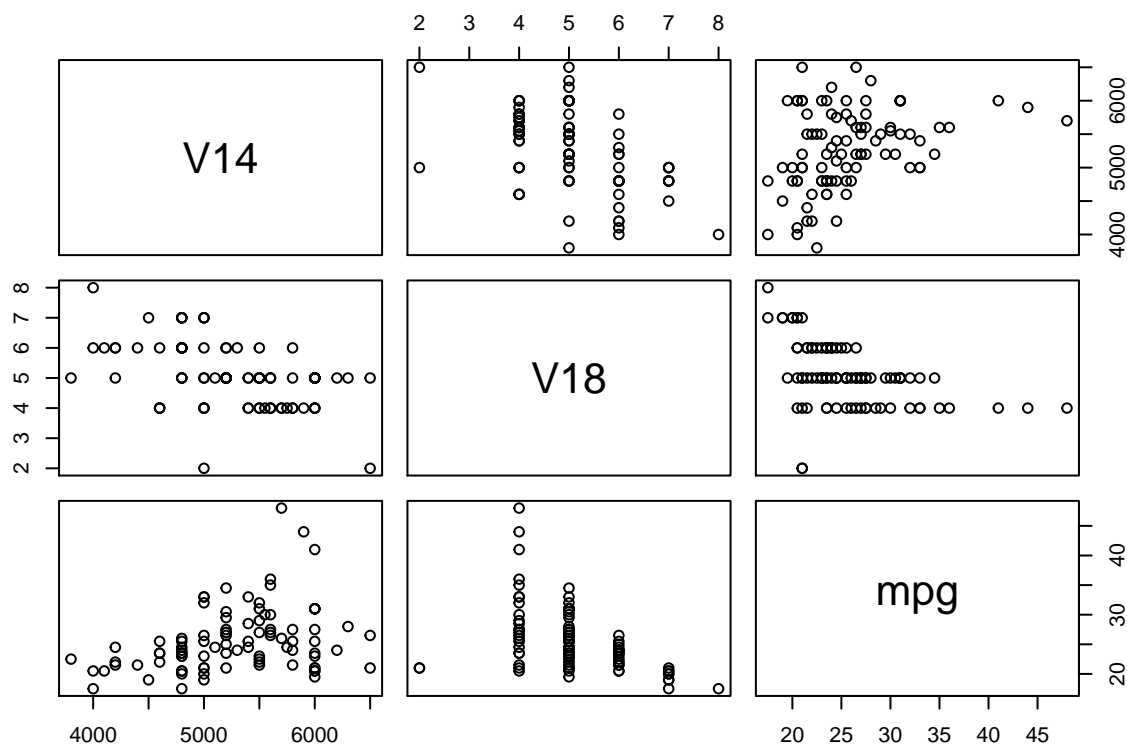
Figur 3: Plotten av de variablerna som har dålig korrelation med cmpg

Av denna figur kan vi inte riktigt avgöra om variablerna bör vara med i modellen eller ej, så vi har kvar dem och utför yttligare tester.

Variabelselektion - forward och backward

```
## Start:  AIC=314.61
## mpg ~ 1
##
##                 Df Sum of Sq      RSS     AIC
## + weight         1    1888.13   793.13 203.34
## + fueltankcap    1    1765.23   916.03 216.73
## + width          1    1281.18  1400.07 256.19
## + enginesize     1    1236.43  1444.83 259.11
## + cylinders      1    1159.79  1521.47 263.92
## + horsepower     1    1153.12  1528.14 264.33
## + enginerev      1    1140.87  1540.38 265.07
## + wheelbase      1    1136.71  1544.55 265.32
## + Uturn          1    1098.31  1582.94 267.60
## + lencar         1    1013.67  1667.59 272.45
## + minprice       1     999.65  1681.60 273.23
## + midprice       1     921.79  1759.46 277.44
## + maxprice       1     791.09  1890.17 284.10
## + passcap        1     536.41  2144.85 295.85
## + rpm            1     316.84  2364.41 304.92
## + domestic       1     136.66  2544.60 311.75
```

```
## <none>                         2681.26 314.61
## + rearseatroom 24     971.56 1709.70 320.77
##
## Step:  AIC=203.34
## mpg ~ weight
##
##               Df Sum of Sq      RSS     AIC
## + wheelbase    1     72.55   720.58 196.41
## + fueltankcap  1     49.98   743.16 199.28
## + lencar       1     29.20   763.93 201.85
## + width        1     21.13   772.00 202.82
## <none>                        793.13 203.34
## + minprice     1     12.67   780.46 203.84
## + midprice     1      8.63   784.50 204.32
## + enginesize   1      8.49   784.64 204.34
## + horsepower   1      7.58   785.56 204.44
## + enginerev    1      7.27   785.86 204.48
## + domestic     1      5.68   787.45 204.67
## + maxprice     1      5.29   787.84 204.71
## + cylinders    1      1.53   791.60 205.16
## + Uturn        1      1.13   792.00 205.20
## + passcap      1      1.12   792.01 205.20
## + rpm          1      0.77   792.36 205.25
## + rearseatroom 24   195.86   597.28 224.96
## - weight       1   1888.13  2681.26 314.61
##
## Step:  AIC=196.41
## mpg ~ weight + wheelbase
##
##               Df Sum of Sq      RSS     AIC
## + fueltankcap  1     39.10   681.48 193.22
## + passcap      1     15.47   705.11 196.40
## + domestic     1     15.47   705.12 196.40
## <none>                        720.58 196.41
## + width        1      9.35   711.23 197.20
## + enginesize   1      9.33   711.25 197.20
## + enginerev    1      6.70   713.89 197.55
## + minprice     1      4.37   716.21 197.85
## + lencar       1      4.17   716.42 197.88
## + cylinders    1      2.90   717.69 198.04
## + midprice     1      2.29   718.29 198.12
## + horsepower   1      2.21   718.37 198.13
## + maxprice     1      1.00   719.59 198.28
## + rpm          1      0.93   719.65 198.29
## + Uturn        1      0.03   720.55 198.41
## - wheelbase    1     72.55   793.13 203.34
## + rearseatroom 24   173.01   547.58 218.88
## - weight       1    823.97  1544.55 265.32
##
## Step:  AIC=193.22
## mpg ~ weight + wheelbase + fueltankcap
##
##               Df Sum of Sq     RSS     AIC
## + domestic     1    23.951  657.53 191.90
```

```
## + passcap        1    15.986 665.49 193.02
## <none>                       681.48 193.22
## + width          1    13.544 667.93 193.36
## + enginerev      1    13.160 668.32 193.41
## + enginesize     1     9.888 671.59 193.87
## + horsepower     1     5.694 675.78 194.44
## + rpm            1     2.623 678.85 194.87
## + lencar         1     2.373 679.10 194.90
## + minprice       1     2.118 679.36 194.94
## + midprice       1     0.743 680.74 195.12
## + Uturn          1     0.411 681.07 195.17
## + cylinders      1     0.290 681.19 195.19
## + maxprice       1     0.144 681.33 195.21
## - fueltankcap    1    39.105 720.58 196.41
## - wheelbase      1    61.679 743.16 199.28
## + rearseatroom  24   194.481 487.00 209.98
## - weight         1   226.236 907.71 217.88
##
## Step:  AIC=191.9
## mpg ~ weight + wheelbase + fueltankcap + domestic
##
##               Df Sum of Sq    RSS    AIC
## + width          1    54.176 603.35 185.90
## + enginesize     1    34.028 623.50 188.96
## <none>                       657.53 191.90
## + passcap        1    13.145 644.38 192.02
## + lencar         1     9.095 648.43 192.60
## + minprice       1     7.991 649.54 192.76
## + midprice       1     5.718 651.81 193.09
## + horsepower     1     4.912 652.62 193.20
## - domestic       1    23.951 681.48 193.22
## + Uturn          1     4.652 652.88 193.24
## + maxprice       1     3.601 653.93 193.39
## + enginerev      1     2.829 654.70 193.50
## + cylinders      1     2.802 654.73 193.50
## + rpm            1     0.346 657.18 193.85
## - fueltankcap    1    47.588 705.12 196.40
## - wheelbase      1    72.966 730.49 199.68
## + rearseatroom  24   191.243 466.28 207.93
## - weight         1   213.916 871.44 216.09
##
## Step:  AIC=185.9
## mpg ~ weight + wheelbase + fueltankcap + domestic + width
##
##               Df Sum of Sq    RSS    AIC
## + enginerev      1    16.051 587.30 185.39
## <none>                       603.35 185.90
## + enginesize     1    11.482 591.87 186.11
## + passcap        1     6.013 597.34 186.97
## + minprice       1     3.784 599.57 187.32
## + horsepower     1     1.797 601.55 187.62
## + midprice       1     1.399 601.95 187.69
## + lencar         1     0.900 602.45 187.76
## + rpm            1     0.515 602.84 187.82
```

```
## + maxprice      1      0.281 603.07 187.86
## + Uturn         1      0.146 603.20 187.88
## + cylinders     1      0.026 603.32 187.90
## - width         1     54.176 657.53 191.90
## - wheelbase     1     56.342 659.69 192.20
## - domestic      1     64.583 667.93 193.36
## - fueltankcap   1     67.667 671.02 193.79
## + rearseatroom 24    188.067 415.28 199.16
## - weight        1    267.780 871.13 218.06
##
## Step:  AIC=185.39
## mpg ~ weight + wheelbase + fueltankcap + domestic + width + enginerev
##
##                 Df Sum of Sq    RSS    AIC
## + enginesize    1     30.289 557.01 182.47
## <none>                       587.30 185.39
## + passcap       1      9.802 577.50 185.83
## - enginerev     1     16.051 603.35 185.90
## + horsepower    1      4.188 583.11 186.73
## + minprice      1      1.937 585.36 187.09
## + lencar        1      1.741 585.56 187.12
## + Uturn         1      0.940 586.36 187.24
## + cylinders     1      0.622 586.68 187.29
## + midprice      1      0.603 586.70 187.30
## + rpm           1      0.156 587.14 187.37
## + maxprice      1      0.075 587.23 187.38
## - domestic      1     45.535 632.83 190.34
## - wheelbase     1     46.438 633.74 190.47
## - width         1     67.398 654.70 193.50
## - fueltankcap   1     77.129 664.43 194.87
## + rearseatroom 24    172.441 414.86 201.07
## - weight        1    198.315 785.61 210.45
##
## Step:  AIC=182.47
## mpg ~ weight + wheelbase + fueltankcap + domestic + width + enginerev +
##     enginesize
##
##                 Df Sum of Sq    RSS    AIC
## + minprice      1     17.136 539.88 181.56
## <none>                       557.01 182.47
## + midprice      1      9.223 547.79 182.92
## + cylinders     1      6.657 550.35 183.35
## + maxprice      1      4.122 552.89 183.78
## + passcap       1      3.827 553.18 183.83
## + rpm           1      2.345 554.67 184.08
## + Uturn         1      0.405 556.61 184.40
## + lencar        1      0.179 556.83 184.44
## + horsepower    1      0.047 556.96 184.46
## - enginesize    1     30.289 587.30 185.39
## - enginerev     1     34.857 591.87 186.11
## - width         1     41.378 598.39 187.13
## - wheelbase     1     53.433 610.44 188.99
## - domestic      1     54.878 611.89 189.21
## - fueltankcap   1     85.224 642.24 193.71
```

```
## + rearseatroom 24    180.715 376.30 193.99
## - weight        1    223.916 780.93 211.89
##
## Step:  AIC=181.56
## mpg ~ weight + wheelbase + fueltankcap + domestic + width + enginerev +
##     enginesize + minprice
##
##                 Df Sum of Sq    RSS    AIC
## + passcap        1    18.738 521.14 180.28
## <none>                        539.88 181.56
## + rpm           1    11.013 528.86 181.65
## + horsepower    1    10.456 529.42 181.74
## - minprice      1    17.136 557.01 182.47
## + midprice      1     5.366 534.51 182.63
## + maxprice      1     5.203 534.67 182.66
## + cylinders     1     4.659 535.22 182.76
## + lencar        1     1.788 538.09 183.25
## + Uturn         1     0.144 539.73 183.54
## - width         1    24.852 564.73 183.75
## - enginerev     1    37.176 577.05 185.76
## - enginesize    1    45.487 585.36 187.09
## - wheelbase     1    49.994 589.87 187.80
## - domestic      1    68.053 607.93 190.60
## - fueltankcap   1    78.376 618.25 192.17
## + rearseatroom 24   169.543 370.33 194.51
## - weight        1   184.528 724.40 206.91
##
## Step:  AIC=180.28
## mpg ~ weight + wheelbase + fueltankcap + domestic + width + enginerev +
##     enginesize + minprice + passcap
##
##                 Df Sum of Sq    RSS    AIC
## <none>                        521.14 180.28
## - width         1    16.238 537.38 181.13
## + rpm           1     6.268 514.87 181.15
## + midprice      1     4.057 517.08 181.55
## - passcap       1    18.738 539.88 181.56
## + maxprice      1     3.938 517.20 181.57
## + horsepower    1     2.314 518.82 181.86
## + cylinders     1     1.726 519.41 181.97
## + lencar        1     1.424 519.71 182.02
## + Uturn         1     0.007 521.13 182.28
## - minprice      1    32.047 553.18 183.83
## - enginerev     1    41.747 562.89 185.44
## - enginesize    1    44.773 565.91 185.94
## - domestic      1    59.977 581.11 188.41
## - wheelbase     1    68.730 589.87 189.80
## - fueltankcap   1    72.448 593.59 190.38
## + rearseatroom 24   157.754 363.38 194.75
## - weight        1   153.514 674.65 202.29
##
## Call:
## lm(formula = mpg ~ weight + wheelbase + fueltankcap + domestic +
```
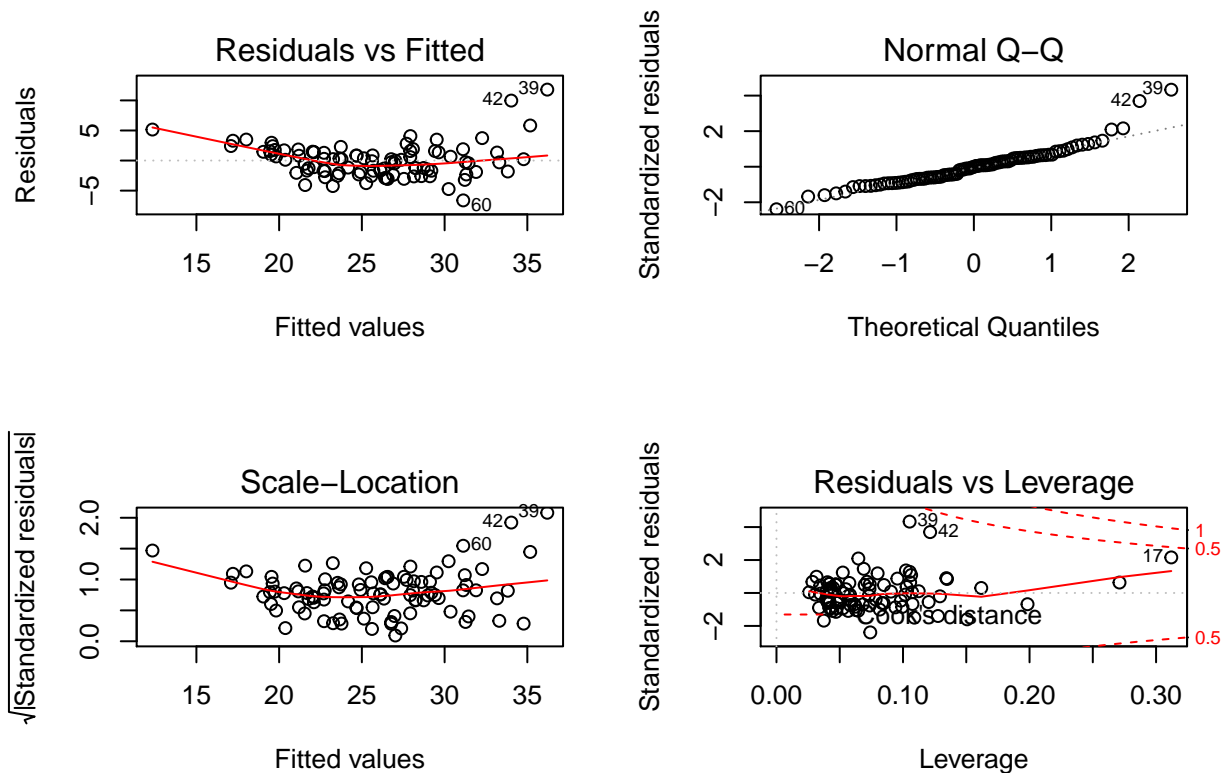
```
##       width + enginerev + enginesize + minprice + passcap)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -4.5669 -1.4095  0.0993  1.3611  8.6676
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.661431  12.139692    0.219  0.82701
## weight      -0.008281   0.001675   -4.945 3.92e-06 ***
## wheelbase    0.318107   0.096147    3.309  0.00139 **
## fueltankcap -0.641508   0.188855   -3.397  0.00105 **
## domestic    -2.164667   0.700386   -3.091  0.00272 **
## width        0.319103   0.198430    1.608  0.11160
## enginerev    0.002587   0.001003    2.579  0.01169 *
## enginesize   1.814765   0.679593    2.670  0.00911 **
## minprice    -0.121207   0.053650   -2.259  0.02649 *
## passcap     -0.724709   0.419511   -1.728  0.08780 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.506 on 83 degrees of freedom
## Multiple R-squared:  0.8056, Adjusted R-squared:  0.7846
## F-statistic: 38.23 on 9 and 83 DF,  p-value: < 2.2e-16

##      weight   wheelbase fueltankcap    domestic       width    enginerev
##   14.300130    6.299636    5.620167    1.814558    8.239058    3.636914
##  enginesize    minprice     passcap
##    7.282365    3.226134    2.783643
```

Vi ser att bredden på bilen har en väldigt låg signifikans, och högt VIF-värde, så vi testar att ta bort den variabeln ur vår modell.

```
##
## Call:
## lm(formula = mpg ~ wheelbase + fueltankcap + passcap + enginerev +
##     minprice + domestic, data = cars)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -6.6268 -1.8208 -0.0243  1.4521 11.7952
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 25.2338745  8.6092177    2.931  0.00433 **
## wheelbase    0.1952111  0.0951716    2.051  0.04330 *
## fueltankcap -0.9542848  0.1613814   -5.913 6.61e-08 ***
## passcap     -1.2217060  0.4657073   -2.623  0.01030 *
## enginerev    0.0023672  0.0009267    2.554  0.01240 *
## minprice    -0.1597589  0.0527812   -3.027  0.00326 **
## domestic    -0.9068639  0.7235670   -1.253  0.21348
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.879 on 86 degrees of freedom
```

```
## Multiple R-squared:  0.7341, Adjusted R-squared:  0.7155
## F-statistic: 39.56 on 6 and 86 DF,  p-value: < 2.2e-16
```



```
##    wheelbase fueltankcap     passcap   enginerev    minprice    domestic
##     4.674244    3.107832    2.597820    2.349101    2.364555    1.466595
```
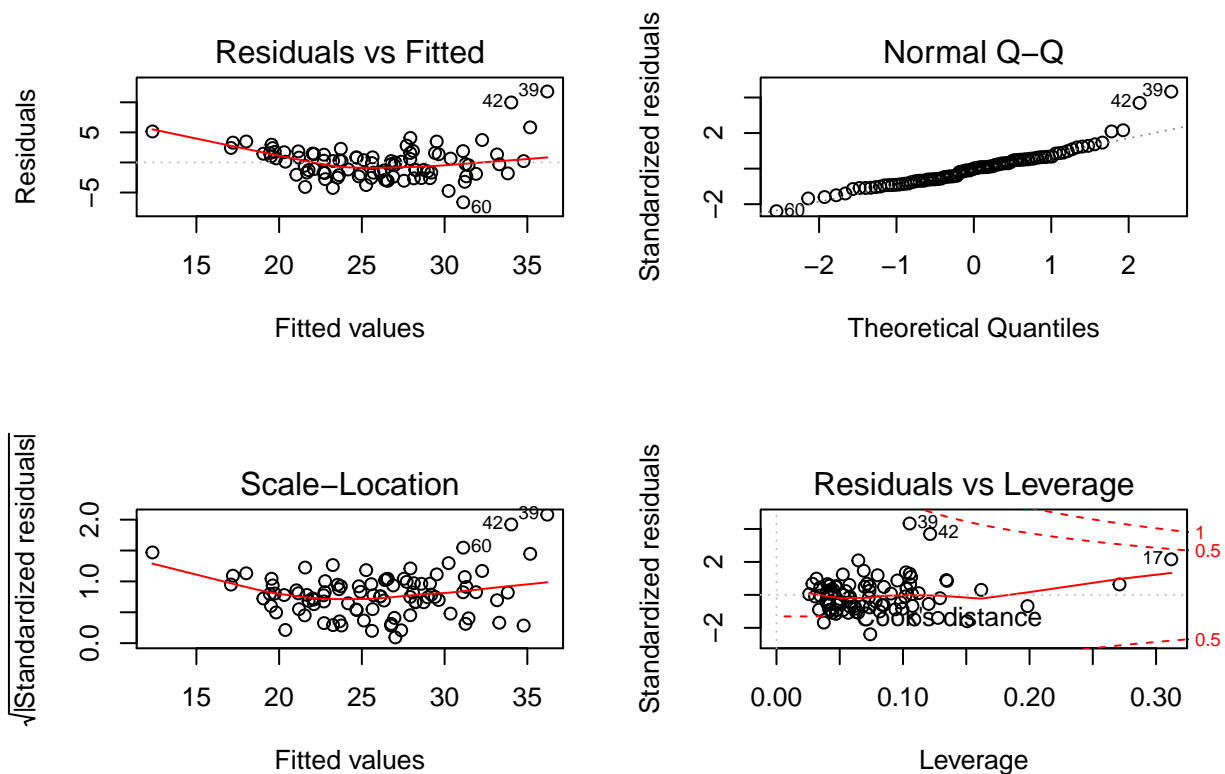
Vi märker att modellen får betydligt högre säkerhet i skattningarna på några av dess parametrar. Detta är typiskt för problem med multikolinearitet. Det finns fortfarande en viss osäkerhet i vissa parametrar, kan detta lösas genom att även ta bort vikt variablen? Detta kan motiveras genom att vikten av en bil bestäms till en stor del av de variabler som vi redan har med, så det är inte så stor mening med att försöka ha kvar denna variabel eller att kombinera den med någon annan.

```
##
## Call:
## lm(formula = mpg ~ wheelbase + fueltankcap + passcap + enginerev +
##     minprice + domestic, data = cars)
##
## Residuals:
##     Min     1Q  Median      3Q     Max
## -6.6268 -1.8208 -0.0243  1.4521 11.7952
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 25.2338745  8.6092177   2.931  0.00433 **
## wheelbase    0.1952111  0.0951716   2.051  0.04330 *
## fueltankcap -0.9542848  0.1613814  -5.913 6.61e-08 ***
## passcap     -1.2217060  0.4657073  -2.623  0.01030 *
```

```
## enginerev     0.0023672  0.0009267     2.554  0.01240 *
## minprice     -0.1597589  0.0527812    -3.027  0.00326 **
## domestic     -0.9068639  0.7235670    -1.253  0.21348
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.879 on 86 degrees of freedom
## Multiple R-squared:  0.7341, Adjusted R-squared:  0.7155
## F-statistic: 39.56 on 6 and 86 DF,  p-value: < 2.2e-16

##   wheelbase fueltankcap      passcap    enginerev     minprice     domestic
##    4.674244    3.107832     2.597820     2.349101     2.364555     1.466595
```

Enligt våra VIF-värden så har vi inte längre några problem med kolinearitet. Modellen har relativt okej R^2-värde, och alla lutningskoefficienter förutom den för domestic har goda t-värden. Detta tyder på att vi inte längre har lika starka multikollinearitet-problem som vi hade tidigare. Vi undersöker residualer och möjliga outliers med nedanstående plottar.


```
##
## Call:
## lm(formula = mpg ~ wheelbase + fueltankcap + passcap + enginerev +
##     minprice + domestic, data = cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6268 -1.8208 -0.0243  1.4521 11.7952
##
## Coefficients:
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) 25.2338745  8.6092177   2.931  0.00433 **
## wheelbase    0.1952111  0.0951716   2.051  0.04330 *
## fueltankcap -0.9542848  0.1613814  -5.913 6.61e-08 ***
## passcap     -1.2217060  0.4657073  -2.623  0.01030 *
## enginerev    0.0023672  0.0009267   2.554  0.01240 *
## minprice    -0.1597589  0.0527812  -3.027  0.00326 **
## domestic    -0.9068639  0.7235670  -1.253  0.21348
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.879 on 86 degrees of freedom
## Multiple R-squared:  0.7341, Adjusted R-squared:  0.7155
## F-statistic: 39.56 on 6 and 86 DF,  p-value: < 2.2e-16

##    wheelbase fueltankcap     passcap    enginerev    minprice    domestic
##     4.674244    3.107832    2.597820     2.349101    2.364555    1.466595

##
## Call:
## lm(formula = mpg ~ fueltankcap + passcap + enginerev + minprice +
##     domestic, data = cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2536 -1.7167  0.0233  1.2022  7.4715
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.2930033  3.1103282  12.955  < 2e-16 ***
## fueltankcap -0.7751716  0.1206219  -6.426 7.24e-09 ***
## passcap     -0.5457420  0.2792933  -1.954  0.05399 .
## enginerev    0.0013391  0.0007269   1.842  0.06895 .
## minprice    -0.1183507  0.0390999  -3.027  0.00327 **
## domestic    -0.4921866  0.5703853  -0.863  0.39062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.263 on 85 degrees of freedom
## Multiple R-squared:  0.7625, Adjusted R-squared:  0.7485
## F-statistic: 54.58 on 5 and 85 DF,  p-value: < 2.2e-16

## fueltankcap     passcap   enginerev    minprice    domestic
##    2.638116    1.475880    2.117328    2.044312    1.440809
```
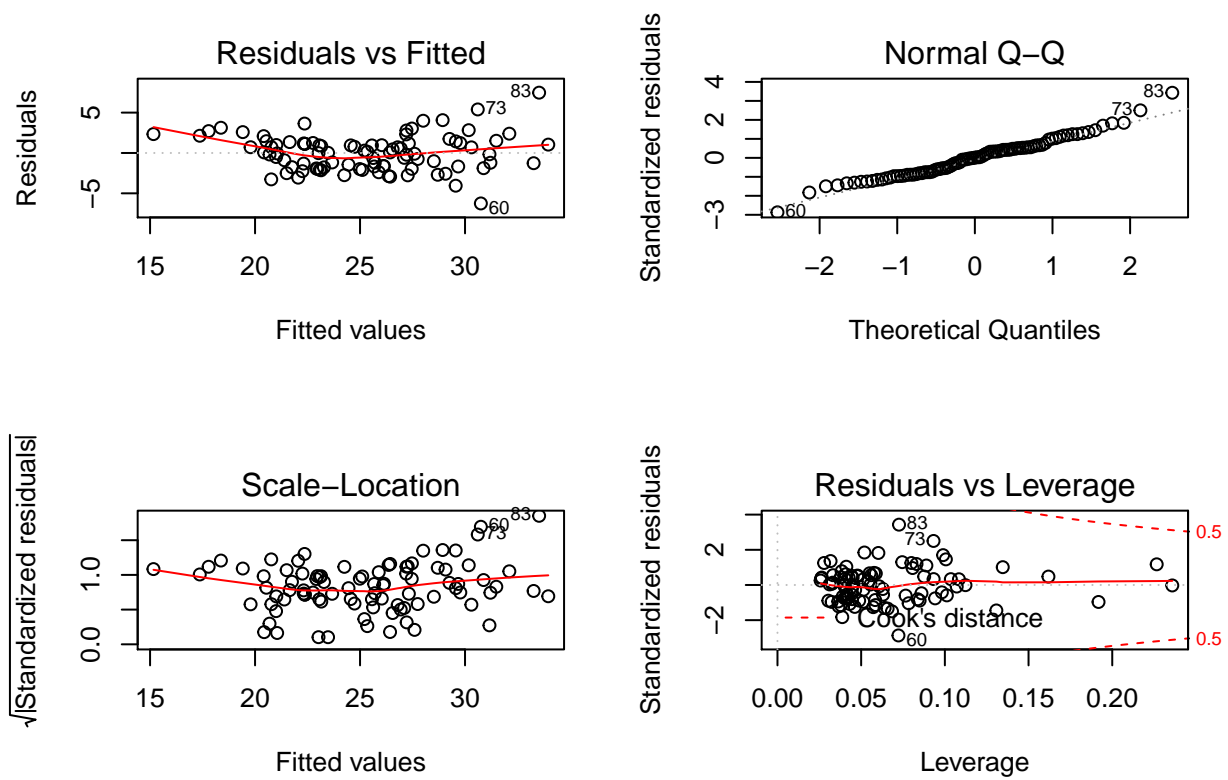
Konstruktion av modell

Jämförelse av amerikanska - och icke-amerikanska bilar