Multiple Regression Analysis, mzcars

Thomas Guenther

Thursday, October 23, 2014

```
library(car)
library(gvlma)
library(QuantPsyc)
```

```
## Loading required package: boot
##
## Attaching package: 'boot'
##
## The following object is masked from 'package:car':
##
##     logit
##
## Loading required package: MASS
##
## Attaching package: 'QuantPsyc'
##
## The following object is masked from 'package:base':
##
##     norm
```

```
library(leaps)

data(mtcars)

full.lm <- lm(mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear + carb, mtcars)
summary(gvlma(full.lm))
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp + hp + drat + wt + qsec +
##     vs + gear + carb, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## am           2.52023    2.05665   1.225   0.2340
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
```

```
## vs           0.31776    2.10451   0.151   0.8814
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = full.lm)
##
##                      Value   p-value                   Decision
## Global Stat        13.8933 0.0076436 Assumptions NOT satisfied!
## Skewness            1.4475 0.2289356    Assumptions acceptable.
## Kurtosis            0.2992 0.5843873    Assumptions acceptable.
## Link Function      11.7323 0.0006143 Assumptions NOT satisfied!
## Heteroscedasticity  0.4143 0.5197791    Assumptions acceptable.
```

```
sqrt(car::vif(full.lm))
```

```
##       am      cyl     disp       hp     drat       wt     qsec       vs
## 2.156035 3.920948 4.649757 3.135608 1.837014 3.894212 2.743712 2.228424
##     gear     carb
## 2.314617 2.812249
```

```
# removing disp
removed_disp <- lm(mpg ~ am + cyl + hp + drat + wt + qsec + vs + gear + carb, mtcars)
summary(gvlma(removed_disp))
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + hp + drat + wt + qsec + vs + gear +
##     carb, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7863 -1.4055 -0.2635  1.2029  4.4753
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.55052   18.52585   0.677   0.5052
## am           2.47882    2.03513   1.218   0.2361
## cyl          0.09627    0.99715   0.097   0.9240
## hp          -0.01295    0.01834  -0.706   0.4876
## drat         0.92864    1.60794   0.578   0.5694
## wt          -2.62694    1.19800  -2.193   0.0392 *
```

```
## qsec         0.66523    0.69335   0.959   0.3478
## vs           0.16035    2.07277   0.077   0.9390
## gear         0.74300    1.47360   0.504   0.6191
## carb        -0.61686    0.60566  -1.018   0.3195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.623 on 22 degrees of freedom
## Multiple R-squared:  0.8655, Adjusted R-squared:  0.8105
## F-statistic: 15.73 on 9 and 22 DF,  p-value: 1.183e-07
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = removed_disp)
##
##                      Value   p-value                   Decision
## Global Stat        13.5330 0.0089450 Assumptions NOT satisfied!
## Skewness            0.8754 0.3494780    Assumptions acceptable.
## Kurtosis            0.3521 0.5529355    Assumptions acceptable.
## Link Function      11.6123 0.0006552 Assumptions NOT satisfied!
## Heteroscedasticity  0.6932 0.4050724    Assumptions acceptable.
```

```
sqrt(car::vif(removed_disp))
```

```
##       am      cyl       hp     drat       wt     qsec       vs     gear
## 2.155251 3.779515 2.668962 1.824636 2.487780 2.629529 2.217217 2.307467
##     carb
## 2.076198
```

```
# removing_cyl
removed_cyl <- lm(mpg ~ am + hp + drat + wt + qsec + vs + gear + carb, mtcars)
summary(gvlma(removed_cyl))
```

```
##
## Call:
## lm(formula = mpg ~ am + hp + drat + wt + qsec + vs + gear + carb,
##     data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8187 -1.3903 -0.3045  1.2269  4.5183
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.80810   12.88582   1.072   0.2950
## am           2.42418    1.91227   1.268   0.2176
## hp          -0.01225    0.01649  -0.743   0.4650
## drat         0.88894    1.52061   0.585   0.5645
## wt          -2.60968    1.15878  -2.252   0.0342 *
```

```
## qsec         0.63983    0.62752   1.020   0.3185
## vs           0.08786    1.88992   0.046   0.9633
## gear         0.69390    1.35294   0.513   0.6129
## carb        -0.61286    0.59109  -1.037   0.3106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.566 on 23 degrees of freedom
## Multiple R-squared:  0.8655, Adjusted R-squared:  0.8187
## F-statistic:  18.5 on 8 and 23 DF,  p-value: 2.627e-08
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = removed_cyl)
##
##                      Value  p-value                   Decision
## Global Stat        12.3654 0.014831 Assumptions NOT satisfied!
## Skewness            0.8926 0.344782    Assumptions acceptable.
## Kurtosis            0.3311 0.565007    Assumptions acceptable.
## Link Function      10.4529 0.001225 Assumptions NOT satisfied!
## Heteroscedasticity  0.6888 0.406566    Assumptions acceptable.
```

```
sqrt(car::vif(removed_cyl))
```

```
##       am       hp     drat       wt     qsec       vs     gear     carb
## 2.070221 2.452710 1.763945 2.459904 2.432834 2.066629 2.165684 2.071344
```

```
# removing_wt
removed_wt <- lm(mpg ~ am + hp + drat + qsec + vs + gear + carb, mtcars)
summary(gvlma(removed_wt))
```

```
##
## Call:
## lm(formula = mpg ~ am + hp + drat + qsec + vs + gear + carb,
##     data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7345 -1.3270  0.0299  1.8235  5.2848
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.347877  13.933716   1.030   0.3134
## am           3.429538   2.010998   1.705   0.1010
## hp          -0.026936   0.016379  -1.645   0.1131
## drat         1.473190   1.620444   0.909   0.3723
## qsec         0.002166   0.605655   0.004   0.9972
## vs           1.513483   1.925904   0.786   0.4396
## gear         1.486429   1.412858   1.052   0.3032
```

```
## carb        -1.130568   0.588941  -1.920   0.0669 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.775 on 24 degrees of freedom
## Multiple R-squared:  0.8358, Adjusted R-squared:  0.7879
## F-statistic: 17.45 on 7 and 24 DF,  p-value: 5.353e-08
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = removed_wt)
##
##                       Value p-value                   Decision
## Global Stat        7.2509274  0.1232    Assumptions acceptable.
## Skewness           0.0003991  0.9841    Assumptions acceptable.
## Kurtosis           0.2927931  0.5884    Assumptions acceptable.
## Link Function      5.5293941  0.0187 Assumptions NOT satisfied!
## Heteroscedasticity 1.4283411  0.2320    Assumptions acceptable.
```

```
sqrt(car::vif(removed_wt))
```

```
##       am       hp     drat     qsec       vs     gear     carb
## 2.013020 2.252778 1.738083 2.171101 1.947255 2.091141 1.908280
```

```
# removing_hp
removed_hp <- lm(mpg ~ am + drat + qsec + vs + gear + carb, mtcars)
summary(gvlma(removed_hp))
```

```
##
## Call:
## lm(formula = mpg ~ am + drat + qsec + vs + gear + carb, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0858 -1.2152  0.4257  2.0044  4.7274
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5965    11.5196   0.052  0.95911
## am            4.0671     2.0394   1.994  0.05714 .
## drat          2.1085     1.6265   1.296  0.20669
## qsec          0.4517     0.5586   0.809  0.42640
## vs            1.5001     1.9905   0.754  0.45811
## gear          1.7015     1.4540   1.170  0.25293
## carb         -1.6831     0.4999  -3.367  0.00246 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.869 on 25 degrees of freedom
```

```
## Multiple R-squared:  0.8173, Adjusted R-squared:  0.7735
## F-statistic: 18.64 on 6 and 25 DF,  p-value: 3.991e-08
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = removed_hp)
##
##                         Value p-value                Decision
## Global Stat        5.357e+00  0.2526 Assumptions acceptable.
## Skewness           1.734e+00  0.1880 Assumptions acceptable.
## Kurtosis           5.186e-05  0.9943 Assumptions acceptable.
## Link Function      1.177e+00  0.2780 Assumptions acceptable.
## Heteroscedasticity 2.447e+00  0.1178 Assumptions acceptable.
```

```
sqrt(car::vif(removed_hp))
```

```
##       am     drat     qsec       vs     gear     carb
## 1.975256 1.687966 1.937454 1.947238 2.082162 1.567286
```

```
lm.beta(full.lm)
```

```
##          am         cyl        disp          hp        drat          wt
##  0.20865790 -0.03302235  0.27422706 -0.24438168  0.06982829 -0.60316876
##        qsec          vs        gear        carb
##  0.24343220  0.02657358  0.08023404 -0.05344363
```

```
lm.beta(removed_disp)
```

```
##          am         cyl          hp        drat          wt        qsec
##  0.20522959  0.02852555 -0.14731759  0.08238388 -0.42647650  0.19723670
##          vs        gear        carb
##  0.01340973  0.09095574 -0.16531518
```

```
lm.beta(removed_cyl)
```

```
##          am          hp        drat          wt        qsec          vs
##  0.20070559 -0.13937460  0.07886160 -0.42367355  0.18970538  0.00734781
##        gear        carb
##  0.08494516 -0.16424436
```

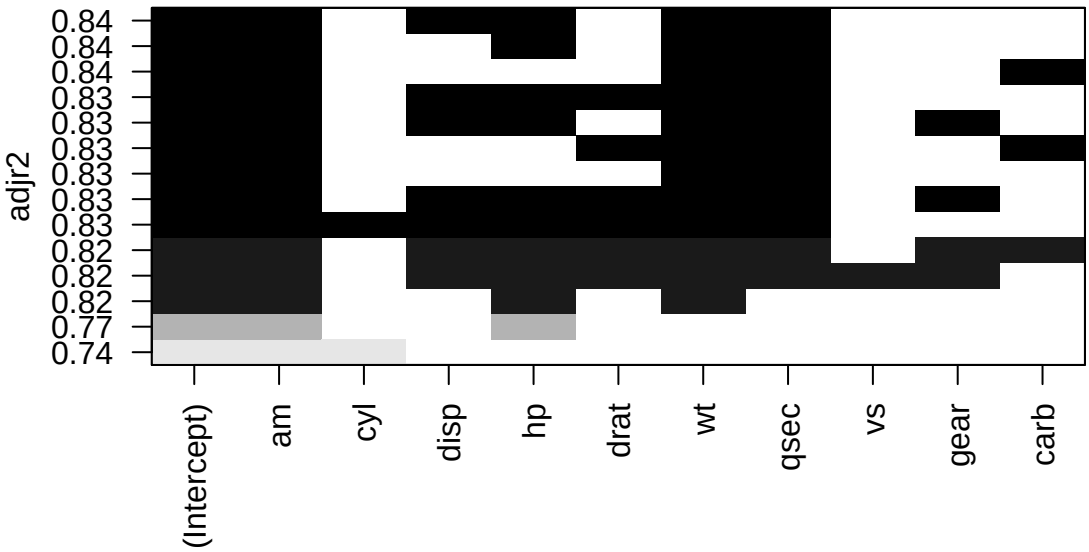
```
lm.beta(removed_wt)
```

```
##            am            hp          drat          qsec            vs
##  0.2839427653 -0.3064284696  0.1306935686  0.0006421891  0.1265682108
##          gear          carb
##  0.1819649849 -0.3029880725
```

```r
lm.beta(removed_hp)
```

```
##         am       drat       qsec         vs       gear       carb
##  0.3367313  0.1870577  0.1339132  0.1254460  0.2082946 -0.4510683
```

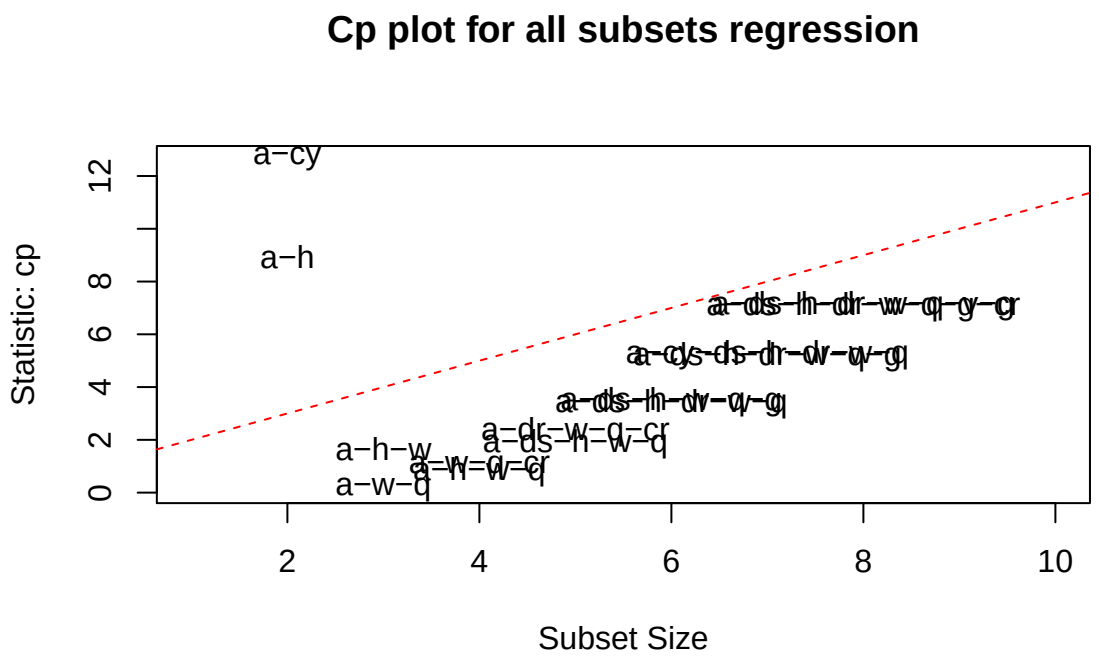
```r
#leaps <- regsubsets(mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear + carb, data=mtcars, nbe
leaps <- regsubsets(mpg ~ ., data=mtcars, nbest=2, force.in = "am")
plot(leaps, scale="adjr2")
```



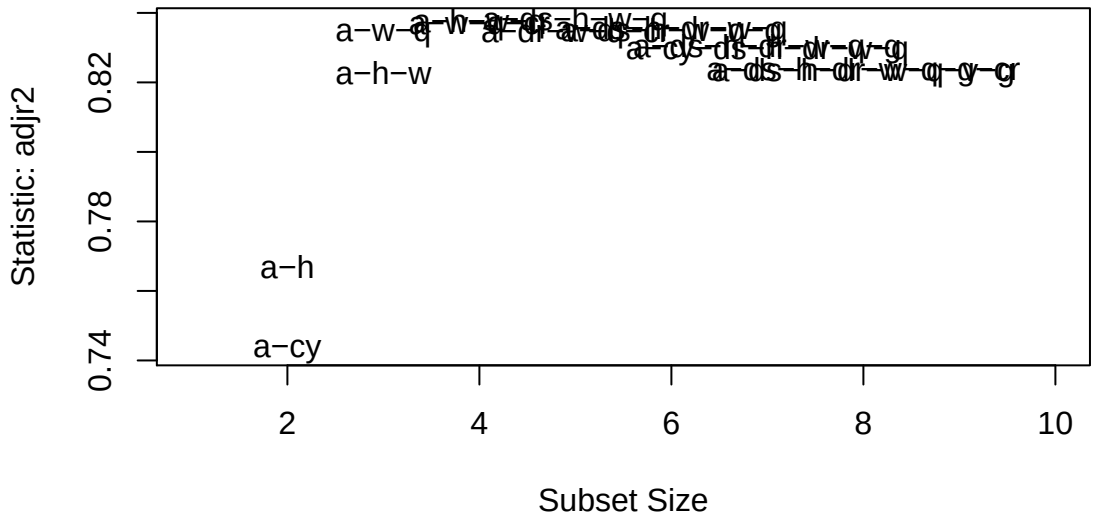

```r
subsets(leaps, statistic="cp", legend = FALSE, main="Cp plot for all subsets regression")
```

```
##      Abbreviation
## am              a
## cyl            cy
## disp           ds
## hp              h
## drat           dr
## wt              w
## qsec            q
## vs              v
## gear            g
## carb           cr
```

```r
abline(a = 1, b = 1,lty=2,col="red")
```




```r
subsets(leaps, statistic="adjr2", legend = FALSE, main = "Adjusted R^2")
```




```
##      Abbreviation
## am              a
## cyl            cy
## disp           ds
## hp              h
## drat           dr
## wt              w
## qsec            q
## vs              v
## gear            g
## carb           cr
```

```
final <- lm(mpg ~ am + disp + hp + wt + qsec, mtcars)

lm.beta(final)
```

```
##         am       disp         hp         wt       qsec
##  0.2873303  0.2310920 -0.2408372 -0.6630794  0.2985371
```

```
summary.out <- summary(leaps)
max.adjr2 <- which.max(summary.out$adjr2)
summary.out$which[max.adjr2,]
```

```
## (Intercept)          am         cyl        disp          hp        drat
```

```
##        TRUE        TRUE       FALSE        TRUE        TRUE       FALSE
##          wt        qsec          vs        gear        carb
##        TRUE        TRUE       FALSE       FALSE       FALSE
```

```
summary(gvlma(final))
```

```
##
## Call:
## lm(formula = mpg ~ am + disp + hp + wt + qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5399 -1.7398 -0.3196  1.1676  4.5534
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.36190    9.74079   1.474  0.15238
## am           3.47045    1.48578   2.336  0.02749 *
## disp         0.01124    0.01060   1.060  0.29897
## hp          -0.02117    0.01450  -1.460  0.15639
## wt          -4.08433    1.19410  -3.420  0.00208 **
## qsec         1.00690    0.47543   2.118  0.04391 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.429 on 26 degrees of freedom
## Multiple R-squared:  0.8637, Adjusted R-squared:  0.8375
## F-statistic: 32.96 on 5 and 26 DF,  p-value: 1.844e-10
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = final)
##
##                       Value  p-value                   Decision
## Global Stat        10.93127 0.027347 Assumptions NOT satisfied!
## Skewness            1.27563 0.258714    Assumptions acceptable.
## Kurtosis            0.52035 0.470693    Assumptions acceptable.
## Link Function       9.08797 0.002573 Assumptions NOT satisfied!
## Heteroscedasticity  0.04732 0.827787    Assumptions acceptable.
```

```
anova(final, removed_hp, removed_wt, removed_cyl, removed_disp, full.lm)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + disp + hp + wt + qsec
## Model 2: mpg ~ am + drat + qsec + vs + gear + carb
## Model 3: mpg ~ am + hp + drat + qsec + vs + gear + carb
## Model 4: mpg ~ am + hp + drat + wt + qsec + vs + gear + carb
## Model 5: mpg ~ am + cyl + hp + drat + wt + qsec + vs + gear + carb
```

```
## Model 6: mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear + carb
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     26 153.44
## 2     25 205.71  1   -52.274
## 3     24 184.88  1    20.834 2.9663 0.09971 .
## 4     23 151.47  1    33.403 4.7558 0.04072 *
## 5     22 151.41  1     0.064 0.0091 0.92477
## 6     21 147.49  1     3.917 0.5576 0.46349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(final, removed_hp, removed_wt, removed_cyl, removed_disp, full.lm)
```

```
##              df      AIC
## final         7 154.9740
## removed_hp    8 166.3558
## removed_wt    9 164.9388
## removed_cyl  10 160.5620
## removed_disp 11 162.5485
## full.lm      12 163.7098
```