Highway Exmaple #2: Best Subsets Selection

We return to the Highway data but this time consider best subsets selection. For this data from Weisberg, the response variable (Y) is accident rate (per million vehicle miles) and there are 13 potential predictors (X's).

We consider AIC selection using the MuMIn package and Cp selection using the leaps package.

```
highway <- read.csv("~/Dropbox/STAT512/Lectures/MultReg3/MR3_Highway.csv")
str(highway)
```

```
## 'data.frame':    39 obs. of  15 variables:
##  $ nr  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ rate: num  4.58 2.86 3.02 2.29 1.61 6.87 3.85 6.12 3.29 5.88 ...
##  $ len : num  4.99 16.11 9.75 10.65 20.01 ...
##  $ adt : int  69 73 49 61 28 30 46 25 43 23 ...
##  $ trks: int  8 8 10 13 12 6 8 9 12 7 ...
##  $ slim: int  55 60 60 65 70 55 55 55 50 50 ...
##  $ lwid: int  12 12 12 12 12 12 12 12 12 12 ...
##  $ shld: int  10 10 10 10 10 10 8 10 4 5 ...
##  $ itg : num  1.2 1.43 1.54 0.94 0.65 0.34 0.47 0.38 0.95 0.12 ...
##  $ sigs: num  0 0 0 0 0 1.84 0.7 0.38 1.39 1.21 ...
##  $ acpt: num  4.6 4.4 4.7 3.8 2.2 24.8 11 18.5 7.5 8.2 ...
##  $ lane: int  8 4 4 6 4 4 4 4 4 4 ...
##  $ fai : int  1 1 1 1 1 0 0 0 0 0 ...
##  $ pa  : int  0 0 0 0 0 1 1 1 1 1 ...
##  $ ma  : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
highway <- highway[,-1]
FullModel <- lm(rate ~ ., data = highway)
```

Best Subset Selection based on AIC using MuMIn package

To use dredge(), you supply the "full" model (the "largest" or most complicated model you want to consider). MuMIn handles factors and produces a nice summary table. For continuous predictors, the partial regression coefficients are shown. For categorical predictors (factors), a + in the summary table would indicate that predictor was included in the model. By default, dredge() will rank models by AICc. That choice is completely reasonable, but I use AIC here to compare to other methods. Note: Watch out for the head() function here. It rescales the model weights to sum to 1 for just the models shown.

```
library(MuMIn)
options(na.action = "na.fail")
AllSubsets <- dredge(FullModel, rank = "AIC", extra = c("R^2"))
```

```
## Fixed term is "(Intercept)"
```

```
head(AllSubsets)
```

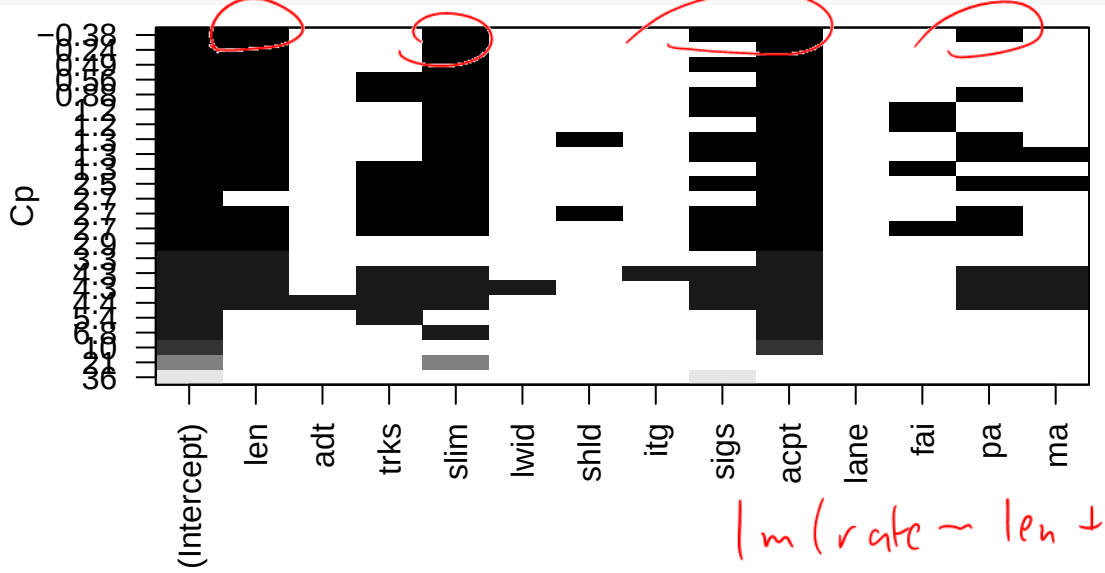
```
## Global model call: lm(formula = rate ~ ., data = highway)
## ---
## Model selection table
##      (Intrc)    acpt   fai   lane      len    ma      pa    shld
## 3362   9.944 0.06428                -0.07405        -0.7744
## 7458  10.570 0.06277                -0.06345        -0.7432
## 3874  10.940 0.06033                -0.07014        -0.9016 0.05678
```

```
## 3490  10.540 0.06993                -0.07105 -0.3785 -0.9725
## 3378   9.933 0.06676        0.05202 -0.07231          -0.7622
## 3366  10.060 0.06721 0.2162         -0.07203          -0.6935
##         sigs    slim     trks     R^2 df  logLik   AIC delta weight
## 3362 0.7974 -0.1051           0.7450  7 -54.942 123.9  0.00  0.301
## 7458 0.7013 -0.1031 -0.08852 0.7521  8 -54.396 124.8  0.91  0.191
## 3874 0.8163 -0.1293          0.7478  8 -54.734 125.5  1.58  0.136
## 3490 0.7622 -0.1136          0.7477  8 -54.734 125.5  1.58  0.136
## 3378 0.7421 -0.1085          0.7459  8 -54.873 125.7  1.86  0.119
## 3366 0.7763 -0.1095          0.7458  8 -54.886 125.8  1.89  0.117
## Models ranked by AIC(x)
```

Best Subset Selection based on Cp using leap package

By Default method="exhaustive" which is best subsets. BIC and Adjusted R2 options are also available, but not aic? Stepwise regression options also available.
Leaps will accept factors, but does not handle them in a convenient way.

```
library(leaps)
Compare <- regsubsets(rate ~ ., data = highway, nbest = 3)
plot(Compare, scale = "Cp")
```




```
summary(Compare)
```

```
## Subset selection object
## Call: regsubsets.formula(rate ~ ., data = highway, nbest = 3)
## 13 Variables  (and intercept)
##      Forced in Forced out
## len      FALSE      FALSE
## adt      FALSE      FALSE
## trks     FALSE      FALSE
## slim     FALSE      FALSE
## lwid     FALSE      FALSE
## shld     FALSE      FALSE
## itg      FALSE      FALSE
## sigs     FALSE      FALSE
```

```
## acpt     FALSE      FALSE
## lane     FALSE      FALSE
## fai      FALSE      FALSE
## pa       FALSE      FALSE
## ma       FALSE      FALSE
## 3 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          len adt trks slim lwid shld itg sigs acpt lane fai pa  ma
## 1  ( 1 ) " " " " " " " "  " "  " "  " "  " " " " " "  "*"  " "  " " " " " " " "
## 1  ( 2 ) " " " " " " " "  "*"  " "  " "  " " " " " "  " "  " "  " " " " " " " "
## 1  ( 3 ) " " " " " " " "  " "  " "  " "  " " " " "*"  " "  " "  " " " " " " " "
## 2  ( 1 ) "*" " " " " " "  " "  " "  " "  " " " " " "  "*"  " "  " " " " " " " "
## 2  ( 2 ) " " " " " " "*"  " "  " "  " "  " " " " " "  "*"  " "  " " " " " " " "
## 2  ( 3 ) " " " " " " " "  "*"  " "  " "  " " " " " "  "*"  " "  " " " " " " " "
## 3  ( 1 ) "*" " " " " " "  "*"  " "  " "  " " " " " "  "*"  " "  " " " " " " " "
## 3  ( 2 ) " " " " " " "*"  "*"  " "  " "  " " " " " "  "*"  " "  " " " " " " " "
## 3  ( 3 ) "*" " " " " " "  " "  " "  " "  " " " " "*"  "*"  " "  " " " " " " " "
## 4  ( 1 ) "*" " " " " " "  "*"  " "  " "  " " " " "*"  "*"  " "  " " " " " " " "
## 4  ( 2 ) "*" " " " " "*"  "*"  " "  " "  " " " " " "  "*"  " "  " " " " " " " "
## 4  ( 3 ) "*" " " " " " "  "*"  " "  " "  " " " " " "  "*"  " "  "*" " " " " "
## 5  ( 1 ) "*" " " " " " "  "*"  " "  " "  " " " " "*"  "*"  " "  " " "*" " " "
## 5  ( 2 ) "*" " " " " " "  "*"  " "  " "  " " " " "*"  "*"  " "  "*" " " " " "
## 5  ( 3 ) "*" " " " " "*"  "*"  " "  " "  " " " " " "  "*"  " "  "*" " " " " "
## 6  ( 1 ) "*" " " " " "*"  "*"  " "  " "  " " " " "*"  "*"  " "  " " "*" " " "
## 6  ( 2 ) "*" " " " " " "  "*"  " "  "*"  " " " " "*"  "*"  " "  " " "*" " " "
## 6  ( 3 ) "*" " " " " " "  "*"  " "  " "  " " " " "*"  "*"  " "  " " "*" "*"
## 7  ( 1 ) "*" " " " " "*"  "*"  " "  " "  " " " " "*"  "*"  " "  " " "*" "*"
## 7  ( 2 ) "*" " " " " "*"  "*"  " "  "*"  " " " " "*"  "*"  " "  " " "*" " " "
## 7  ( 3 ) "*" " " " " "*"  "*"  " "  " "  " " " " "*"  "*"  " "  "*" "*" " " "
## 8  ( 1 ) "*" " " " " "*"  "*"  " "  " "  " " "*" "*"  "*"  " "  " " "*" "*"
## 8  ( 2 ) "*" " " " " "*"  "*"  "*"  " "  " " " " "*"  "*"  " "  " " "*" "*"
## 8  ( 3 ) "*" "*" "*"  "*"  " "  " "  " " " " "*"  "*"  " "  " " "*" "*"
```