Initialization

```
knitr::opts_chunk$set(echo = TRUE)
demand<-read.table("demand.txt")
colnames(demand) = c("X1","X2","X3","Y")
attach(demand)
```

```
## The following objects are masked from cp (pos = 3):
##
##     X1, X2, X3, Y

## The following objects are masked from demand (pos = 4):
##
##     X1, X2, X3, Y

## The following object is masked from drug (pos = 5):
##
##     Y

## The following object is masked from drug (pos = 8):
##
##     Y

## The following objects are masked from demand (pos = 9):
##
##     X1, X2, X3, Y

## The following objects are masked from demand (pos = 14):
##
##     X1, X2, X3, Y

## The following objects are masked from flow (pos = 15):
##
##     X1, X2, X3

## The following objects are masked from cp (pos = 16):
##
##     X1, X2, X3, Y

## The following objects are masked from cp (pos = 17):
##
##     X1, X2, X3, Y

## The following objects are masked from flow (pos = 18):
##
##     X1, X2, X3

## The following objects are masked from flow (pos = 19):
##
##     X1, X2, X3
```

```
## The following objects are masked from flow (pos = 20):
##
##     X1, X2, X3
```

```
# n is the sample size and p is the number of predictors

n = nrow(demand)
p = ncol(demand)-1

X<-matrix(0,n,p+1)
X[,1]<-rep(1,n)
for(j in 1:p)
X[,j+1]<-demand[,j]

fullfit<-lm(Y~X1+X2+X3, data = demand)
summary(fullfit)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3, data = demand)
##
## Residuals:
##       1       2       3       4       5       6       7       8
##  0.8757 -0.9719 -1.4924  0.7621 -4.6125 -5.4619 -2.1292 11.6531
##       9
##  1.3771
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.6159    40.8535   1.410   0.2175
## X1            0.2398     1.0121   0.237   0.8221
## X2           10.7184     4.5296   2.366   0.0642 .
## X3           -0.7510     0.3950  -1.901   0.1157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.29 on 5 degrees of freedom
## Multiple R-squared:  0.8453, Adjusted R-squared:  0.7525
## F-statistic: 9.109 on 3 and 5 DF,  p-value: 0.01807
```

Forward Selection

```
# First consider forward selection with SLE = 0.15
fit1<-lm(Y~1,data=demand)

summary(fit1)
```

```
##
## Call:
## lm(formula = Y ~ 1, data = demand)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.2667  -5.0667   0.0333   7.4333  18.6333
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  167.067      4.215   39.64 1.81e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.65 on 8 degrees of freedom
```

```
add1(fit1,~X1+X2+X3,test = "F", data=demand)
```

```
## Single term additions
##
## Model:
## Y ~ 1
##        Df Sum of Sq     RSS    AIC F value   Pr(>F)
## <none>              1279.20 46.611
## X1      1    822.28  456.92 39.346 12.5972 0.009353 **
## X2      1    754.41  524.79 40.592 10.0627 0.015660 *
## X3      1    505.70  773.50 44.083  4.5764 0.069711 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# X1 has lowest p-value
fit2<-lm(Y~X1,data=demand)

summary(fit2)
```

```
##
## Call:
## lm(formula = Y ~ X1, data = demand)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.039 -8.074  1.114  3.322 11.695
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  41.1881    35.5683   1.158  0.28484
## X1            2.4275     0.6839   3.549  0.00935 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.079 on 7 degrees of freedom
## Multiple R-squared:  0.6428, Adjusted R-squared:  0.5918
## F-statistic:  12.6 on 1 and 7 DF,  p-value: 0.009353
```

```r
add1(fit2,~X1+X2+X3,test = "F", data=demand)
```

```
## Single term additions
##
## Model:
## Y ~ X1
##        Df Sum of Sq    RSS    AIC F value Pr(>F)
## <none>              456.92 39.346
## X2      1   116.015 340.91 38.709  2.0419 0.2030
## X3      1    37.506 419.42 40.575  0.5366 0.4915
```

```r
# All p-values are over 0.15

# model is Y~X1
```

Backward elimination

```r
# Now, consider the backward elimination with SLS = 0.05

# remove the predictor with the highest p-value (as long as that predictor's p-value is >
↪  0.05)

fit.back<-lm(Y~.,data=demand)
summary(fit.back)
```

```
##
## Call:
## lm(formula = Y ~ ., data = demand)
##
## Residuals:
##       1       2       3       4       5       6       7       8
##  0.8757 -0.9719 -1.4924  0.7621 -4.6125 -5.4619 -2.1292 11.6531
##       9
##  1.3771
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.6159    40.8535   1.410   0.2175
## X1            0.2398     1.0121   0.237   0.8221
## X2           10.7184     4.5296   2.366   0.0642 .
## X3           -0.7510     0.3950  -1.901   0.1157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.29 on 5 degrees of freedom
## Multiple R-squared:  0.8453, Adjusted R-squared:  0.7525
## F-statistic: 9.109 on 3 and 5 DF,  p-value: 0.01807
```

```r
# remove X1, highest p-value
fit.back<-update(fit.back, .~.-X1)
summary(fit.back)
```

```
##
## Call:
## lm(formula = Y ~ X2 + X3, data = demand)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8726 -2.7305  0.0653  1.1114 11.4813
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.0211    31.1138   2.026  0.08922 .
## X2           11.5172     2.7773   4.147  0.00603 **
## X3           -0.8158     0.2614  -3.121  0.02057 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.775 on 6 degrees of freedom
## Multiple R-squared:  0.8436, Adjusted R-squared:  0.7915
## F-statistic: 16.18 on 2 and 6 DF,  p-value: 0.003826
```

```r
# remove intercept, highest p-value
fit.back<-lm(Y ~ 0 + X2 + X3, data=demand)
summary(fit.back)
```

```
##
## Call:
## lm(formula = Y ~ 0 + X2 + X3, data = demand)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.867 -3.601 -1.695  1.330 15.867
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## X2  17.0341     0.6527  26.097  3.1e-08 ***
## X3  -0.6295     0.2940  -2.141   0.0695 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.937 on 7 degrees of freedom
## Multiple R-squared:  0.9987, Adjusted R-squared:  0.9983
## F-statistic:  2620 on 2 and 7 DF,  p-value: 8.677e-11
```

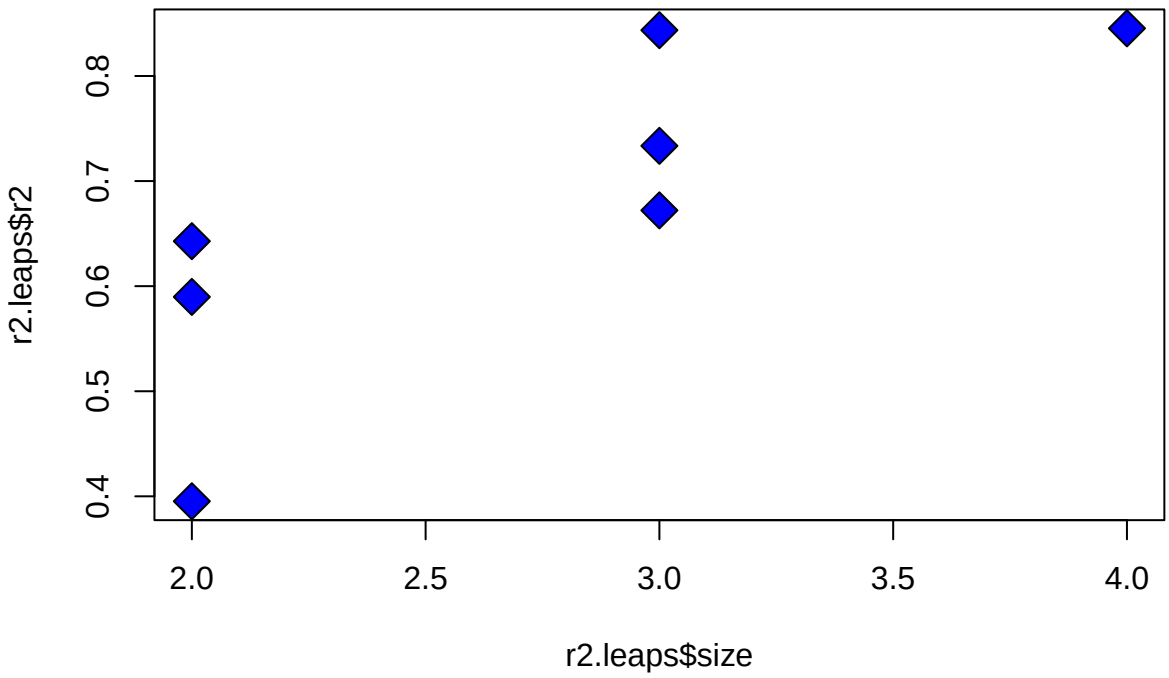
```r
# remove X3, highest p-value
fit.back<-lm(Y ~ 0 + X2, data=demand)
summary(fit.back)
```

```
##
## Call:
## lm(formula = Y ~ 0 + X2, data = demand)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.921 -5.360 -2.894  4.865 17.338
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## X2  15.7162     0.2614   60.12 6.51e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.348 on 8 degrees of freedom
## Multiple R-squared:  0.9978, Adjusted R-squared:  0.9975
## F-statistic:  3615 on 1 and 8 DF,  p-value: 6.507e-12
```

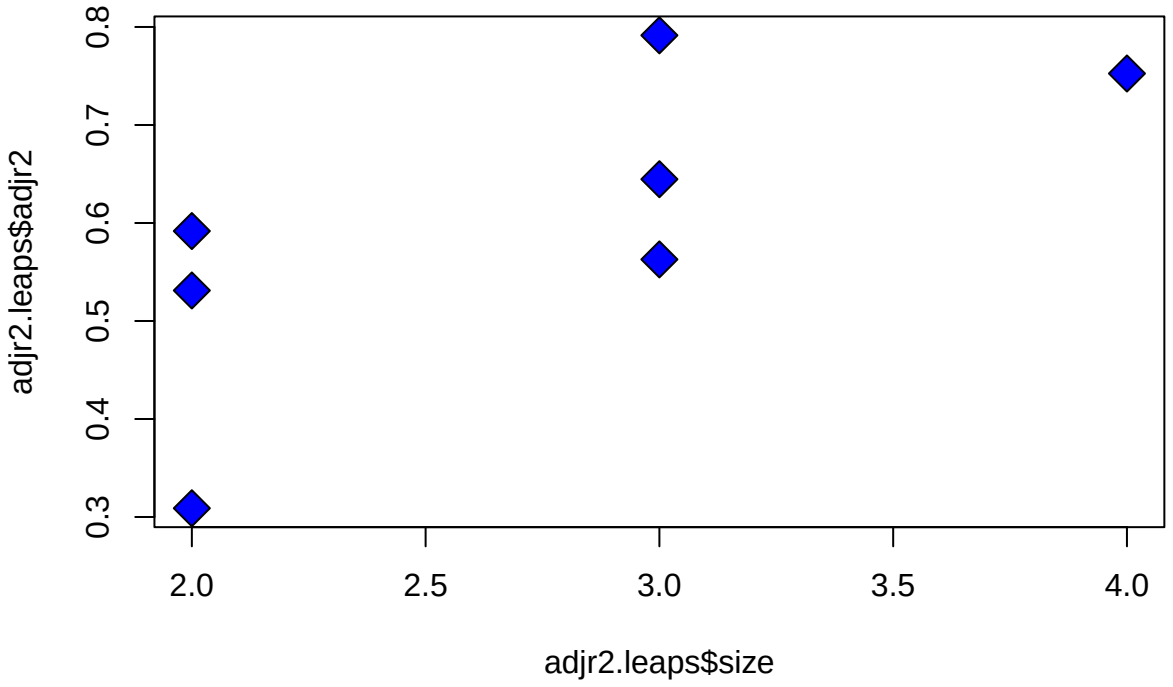
```
# no more p-value > 0.05

# model is Y~0+X2
```

Subset regression

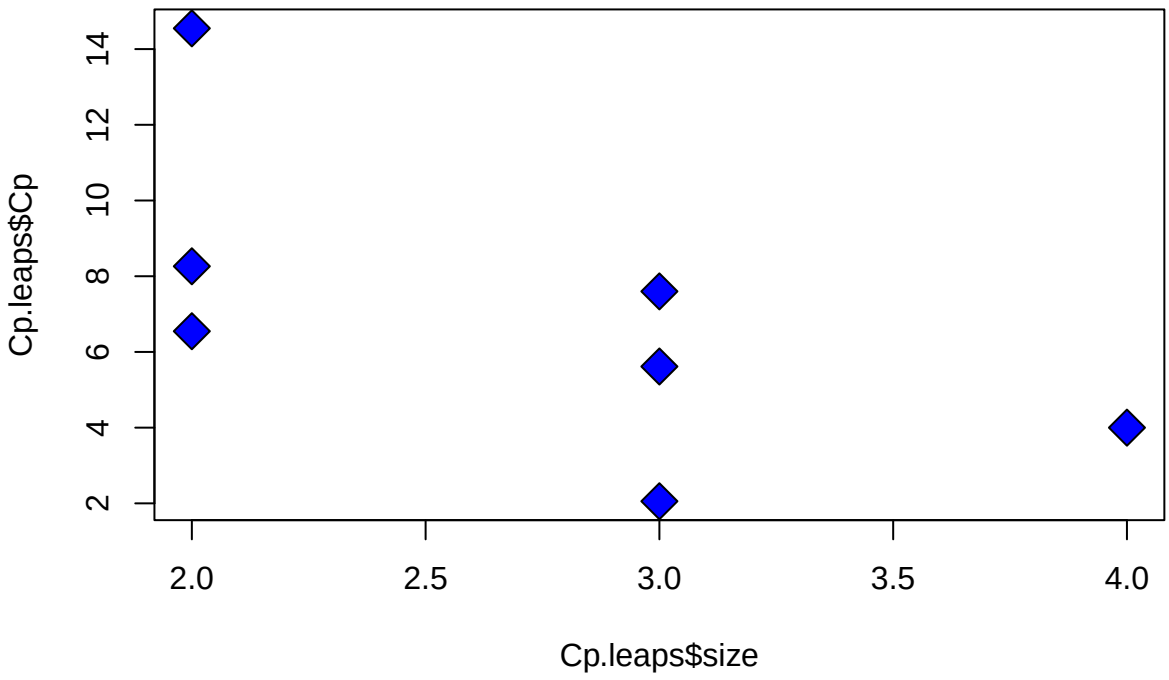
```
# we need to install some packages first before doing subset regression


# R^2 now has all subsets models

library(leaps)
r2.leaps <- leaps(X[,-1], Y, nbest=3, method='r2')
plot(r2.leaps$size, r2.leaps$r2, pch=23, bg='blue', cex=2)
```




```
best.model.r2 <- r2.leaps$which[which((r2.leaps$r2 == max(r2.leaps$r2))),]
#print(best.model.r2)




# adjusted R^2 now has all subset models
adjr2.leaps <- leaps(X[,-1], Y, nbest=3, method='adjr2')
plot(adjr2.leaps$size, adjr2.leaps$adjr2, pch=23, bg='blue', cex=2)
```




```
best.model.adjr2 <- adjr2.leaps$which[which((adjr2.leaps$adjr2 ==
↪  max(adjr2.leaps$adjr2))),]
#print(best.model.adjr2)



# C_p now has all subset models
Cp.leaps <- leaps(X[,-1],Y, nbest=3, method='Cp')
plot(Cp.leaps$size, Cp.leaps$Cp, pch=23, bg='blue', cex=2)
```




```
Cp.leaps2=regsubsets(Y~., data=demand,nvmax=5)
summary.Cp = summary(Cp.leaps2)
summary.Cp$cp
```

```
## [1] 6.547063 2.056159 4.000000
```

```
#plot(Cp.leaps2,scale="Cp")



# Cp plot is needed from faraway R package.


library(faraway)
#Cpplot(Cp.leaps)


# Stepwise regression performed using the AIC criteria.

null = lm(Y~1,data = demand)
full = lm(Y~., data = demand)
summary(fit <- lm(Y~., data = demand))
```

```
##
```

```
## Call:
## lm(formula = Y ~ ., data = demand)
##
## Residuals:
##       1       2       3       4       5       6       7       8
##  0.8757 -0.9719 -1.4924  0.7621 -4.6125 -5.4619 -2.1292 11.6531
##       9
##  1.3771
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.6159    40.8535   1.410   0.2175
## X1            0.2398     1.0121   0.237   0.8221
## X2           10.7184     4.5296   2.366   0.0642 .
## X3           -0.7510     0.3950  -1.901   0.1157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.29 on 5 degrees of freedom
## Multiple R-squared:  0.8453, Adjusted R-squared:  0.7525
## F-statistic: 9.109 on 3 and 5 DF,  p-value: 0.01807
```

```
sfit <- step(null, scope = list(lower=null, upper=full),direction='both')
```

```
## Start:  AIC=46.61
## Y ~ 1
##
##        Df Sum of Sq     RSS    AIC
## + X1    1    822.28  456.92 39.346
## + X2    1    754.41  524.79 40.592
## + X3    1    505.70  773.50 44.083
## <none>              1279.20 46.611
##
## Step:  AIC=39.35
## Y ~ X1
##
##        Df Sum of Sq     RSS    AIC
## + X2    1    116.01  340.91 38.709
## <none>               456.92 39.346
## + X3    1     37.51  419.42 40.575
## - X1    1    822.28 1279.20 46.611
##
## Step:  AIC=38.71
## Y ~ X1 + X2
##
##        Df Sum of Sq    RSS    AIC
## + X3    1    143.06 197.85 35.813
## <none>              340.91 38.709
## - X2    1    116.02 456.92 39.346
## - X1    1    183.89 524.79 40.592
##
## Step:  AIC=35.81
## Y ~ X1 + X2 + X3
```

```
##
##        Df Sum of Sq    RSS    AIC
## - X1    1     2.222 200.07 33.913
## <none>              197.85 35.813
## - X3    1   143.055 340.91 38.709
## - X2    1   221.564 419.42 40.575
##
## Step:  AIC=33.91
## Y ~ X2 + X3
##
##        Df Sum of Sq    RSS    AIC
## <none>              200.07 33.913
## + X1    1      2.22 197.85 35.813
## - X3    1    324.72 524.79 40.592
## - X2    1    573.43 773.50 44.083
```

```
summary(sfit)
```

```
##
## Call:
## lm(formula = Y ~ X2 + X3, data = demand)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8726 -2.7305  0.0653  1.1114 11.4813
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.0211    31.1138   2.026  0.08922 .
## X2           11.5172     2.7773   4.147  0.00603 **
## X3           -0.8158     0.2614  -3.121  0.02057 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.775 on 6 degrees of freedom
## Multiple R-squared:  0.8436, Adjusted R-squared:  0.7915
## F-statistic: 16.18 on 2 and 6 DF,  p-value: 0.003826
```

```
#sfit$anova



# Step function used to perform the forward selection, backward elimination.

null = lm(Y~1,data = demand)
full = lm(Y~., data = demand)

## forward selection
sfit_f <- step(null, scope = list(lower=null, upper=full),direction='forward')
```

```
## Start:  AIC=46.61
## Y ~ 1
```

```
##
##        Df Sum of Sq     RSS    AIC
## + X1    1    822.28  456.92 39.346
## + X2    1    754.41  524.79 40.592
## + X3    1    505.70  773.50 44.083
## <none>              1279.20 46.611
##
## Step:  AIC=39.35
## Y ~ X1
##
##        Df Sum of Sq    RSS    AIC
## + X2    1   116.015 340.91 38.709
## <none>              456.92 39.346
## + X3    1    37.506 419.42 40.575
##
## Step:  AIC=38.71
## Y ~ X1 + X2
##
##        Df Sum of Sq    RSS    AIC
## + X3    1    143.06 197.85 35.813
## <none>              340.91 38.709
##
## Step:  AIC=35.81
## Y ~ X1 + X2 + X3
```

```
## backward selection

sfit_b <- step(full, direction='backward')
```

```
## Start:  AIC=35.81
## Y ~ X1 + X2 + X3
##
##        Df Sum of Sq    RSS    AIC
## - X1    1     2.222 200.07 33.913
## <none>              197.85 35.813
## - X3    1   143.055 340.91 38.709
## - X2    1   221.564 419.42 40.575
##
## Step:  AIC=33.91
## Y ~ X2 + X3
##
##        Df Sum of Sq    RSS    AIC
## <none>              200.07 33.913
## - X3    1    324.72 524.79 40.592
## - X2    1    573.43 773.50 44.083
```

```
# model is Y ~ X2 + X3
```

Best model for forward selection is Y~X1, best model for backwards elimination is Y~0+X2, best model for subset regression is Y~X2+X3.