Lesson 10

Christopher A. Swenson (chris@cswenson.com)

11/27/2021

Martians (underspecified model)

Load the martians data. Fit a multiple linear regression model of weight vs height + water. Fit a simple linear regression model of weight vs height. Create a scatterplot of weight vs height with points marked by water level and regression lines for each model.

```
martian <- read.table("./Data/martian.txt", header=T)
attach(martian)

model.1 <- lm(weight ~ height + water)
summary(model.1)
```

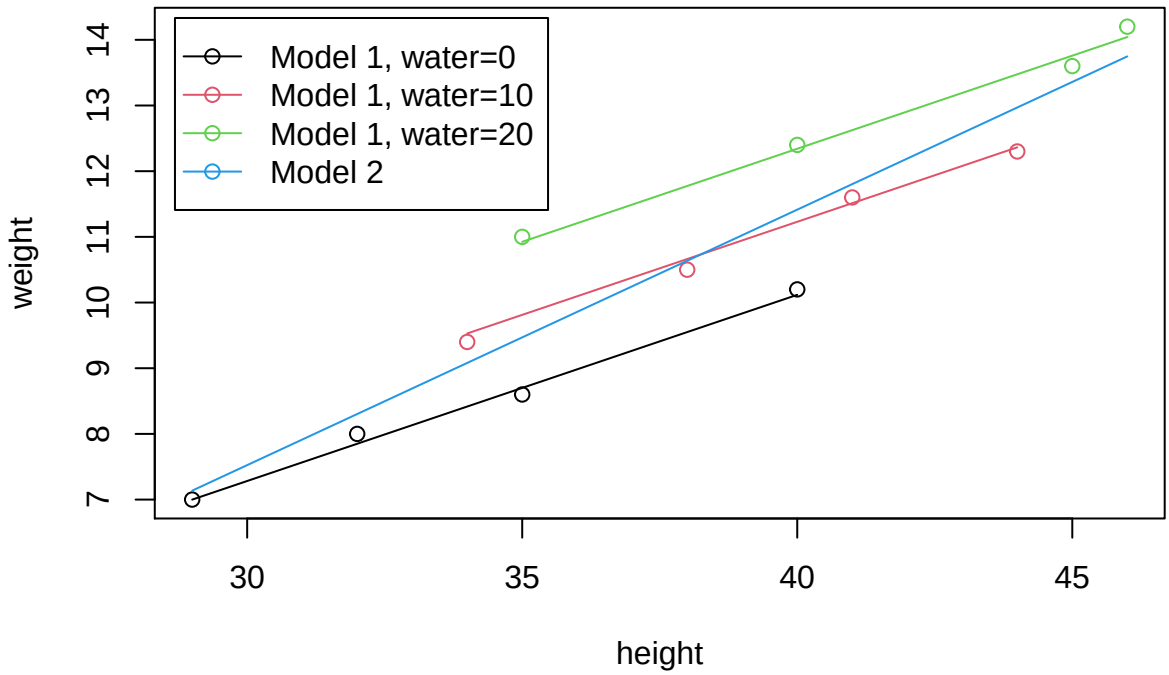
```
##
## Call:
## lm(formula = weight ~ height + water)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16247 -0.10722  0.02955  0.08388  0.15792
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.220194   0.320978  -3.801  0.00421 **
## height       0.283436   0.009142  31.003 1.85e-10 ***
## water        0.111212   0.005748  19.348 1.22e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1305 on 9 degrees of freedom
## Multiple R-squared:  0.9972, Adjusted R-squared:  0.9966
## F-statistic:  1592 on 2 and 9 DF,  p-value: 3.353e-12
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) -1.220194   0.320978  -3.801  0.00421 **
# height       0.283436   0.009142  31.003 1.85e-10 ***
# water        0.111212   0.005748  19.348 1.22e-08 ***
# ---
# Residual standard error: 0.1305 on 9 degrees of freedom

model.2 <- lm(weight ~ height)
summary(model.2)
```

```
##
## Call:
```

```
## lm(formula = weight ~ height)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2140 -0.3943 -0.1359  0.3528  1.5307
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.14335    1.75340  -2.363   0.0397 *
## height       0.38893    0.04543   8.561 6.48e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.808 on 10 degrees of freedom
## Multiple R-squared:  0.8799, Adjusted R-squared:  0.8679
## F-statistic: 73.28 on 1 and 10 DF,  p-value: 6.475e-06
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) -4.14335    1.75340  -2.363   0.0397 *
# height       0.38893    0.04543   8.561 6.48e-06 ***
# ---
# Residual standard error: 0.808 on 10 degrees of freedom

plot(x=height, y=weight, col=water/10+1,
     panel.last = c(lines(sort(height[water==0]),
                          fitted(model.1)[water==0][order(height[water==0])],
                          col=1),
                    lines(sort(height[water==10]),
                          fitted(model.1)[water==10][order(height[water==10])],
                          col=2),
                    lines(sort(height[water==20]),
                          fitted(model.1)[water==20][order(height[water==20])],
                          col=3),
                    lines(sort(height), fitted(model.2)[order(height)], col=4)))
legend("topleft", col=1:4, pch=1, lty=1, inset=0.02,
       legend=c("Model 1, water=0", "Model 1, water=10",
                "Model 1, water=20", "Model 2"))
```

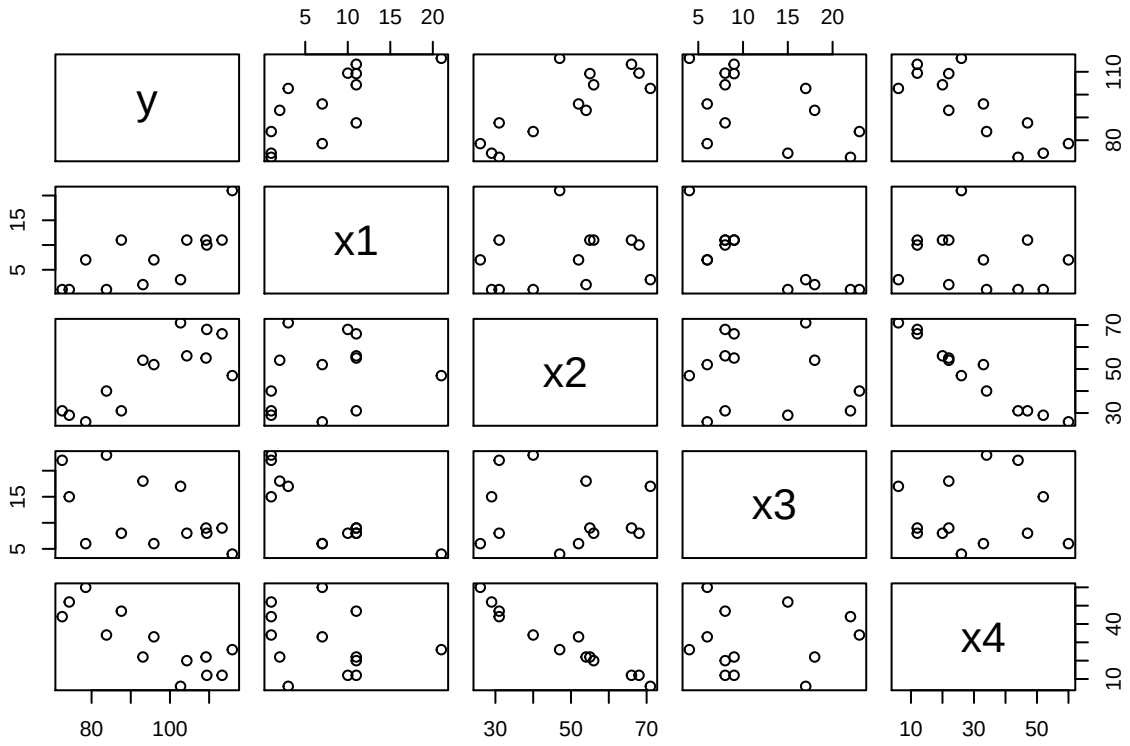



```
detach(martian)
```

Cement hardening (variable selection using stepwise regression)

Load the cement data. Create a scatterplot matrix of the data. Use the add1 and drop1 functions to conduct stepwise regression.

```
cement <- read.table("./Data/cement.txt", header=T)
attach(cement)

pairs(cement)
```




```
model.0 <- lm(y ~ 1)
add1(model.0, ~ x1 + x2 + x3 + x4, test="F")
```

```
## Single term additions
##
## Model:
## y ~ 1
##        Df Sum of Sq     RSS    AIC F value    Pr(>F)
## <none>              2715.76 71.444
## x1      1   1450.08 1265.69 63.519 12.6025 0.0045520 **
## x2      1   1809.43  906.34 59.178 21.9606 0.0006648 ***
## x3      1    776.36 1939.40 69.067  4.4034 0.0597623 .
## x4      1   1831.90  883.87 58.852 22.7985 0.0005762 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# y ~ 1
#        Df Sum of Sq     RSS    AIC F value    Pr(>F)
#              2715.76 71.444
# x1      1   1450.08 1265.69 63.519 12.6025 0.0045520 **
# x2      1   1809.43  906.34 59.178 21.9606 0.0006648 ***
# x3      1    776.36 1939.40 69.067  4.4034 0.0597623 .
# x4      1   1831.90  883.87 58.852 22.7985 0.0005762 ***

model.4 <- lm(y ~ x4)
add1(model.4, ~ . + x1 + x2 + x3, test="F")
```

```
## Single term additions
##
## Model:
```

```
## y ~ x4
##        Df Sum of Sq    RSS    AIC  F value    Pr(>F)
## <none>              883.87 58.852
## x1      1    809.10  74.76 28.742 108.2239 1.105e-06 ***
## x2      1     14.99 868.88 60.629   0.1725    0.6867
## x3      1    708.13 175.74 39.853  40.2946 8.375e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# y ~ x4
#        Df Sum of Sq    RSS    AIC  F value    Pr(>F)
#              883.87 58.852
# x1      1    809.10  74.76 28.742 108.2239 1.105e-06 ***
# x2      1     14.99 868.88 60.629   0.1725    0.6867
# x3      1    708.13 175.74 39.853  40.2946 8.375e-05 ***

model.14 <- lm(y ~ x1 + x4)
drop1(model.14, ~ ., test="F")
```

```
## Single term deletions
##
## Model:
## y ~ x1 + x4
##        Df Sum of Sq     RSS    AIC F value    Pr(>F)
## <none>               74.76 28.742
## x1      1     809.1  883.87 58.852  108.22 1.105e-06 ***
## x4      1    1190.9 1265.69 63.519  159.30 1.815e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# y ~ x1 + x4
#        Df Sum of Sq     RSS    AIC F value    Pr(>F)
#                74.76 28.742
# x1      1     809.1  883.87 58.852  108.22 1.105e-06 ***
# x4      1    1190.9 1265.69 63.519  159.30 1.815e-07 ***

add1(model.14, ~ . + x2 + x3, test="F")
```

```
## Single term additions
##
## Model:
## y ~ x1 + x4
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              74.762 28.742
## x2      1    26.789 47.973 24.974  5.0259 0.05169 .
## x3      1    23.926 50.836 25.728  4.2358 0.06969 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# y ~ x1 + x4
#        Df Sum of Sq    RSS    AIC F value  Pr(>F)
#              74.762 28.742
# x2      1    26.789 47.973 24.974  5.0259 0.05169 .
```

```
# x3      1    23.926 50.836 25.728  4.2358 0.06969 .

model.124 <- lm(y ~ x1 + x2 + x4)
drop1(model.124, ~ ., test="F")
```

```
## Single term deletions
##
## Model:
## y ~ x1 + x2 + x4
##        Df Sum of Sq    RSS    AIC  F value    Pr(>F)
## <none>              47.97 24.974
## x1      1    820.91 868.88 60.629 154.0076 5.781e-07 ***
## x2      1     26.79  74.76 28.742   5.0259   0.05169 .
## x4      1      9.93  57.90 25.420   1.8633   0.20540
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# y ~ x4 + x1 + x2
#        Df Sum of Sq    RSS    AIC  F value    Pr(>F)
#               47.97 24.974
# x1      1    820.91 868.88 60.629 154.0076 5.781e-07 ***
# x2      1     26.79  74.76 28.742   5.0259   0.05169 .
# x4      1      9.93  57.90 25.420   1.8633   0.20540

model.12 <- lm(y ~ x1 + x2)
add1(model.12, ~ . + x3 + x4, test="F")
```

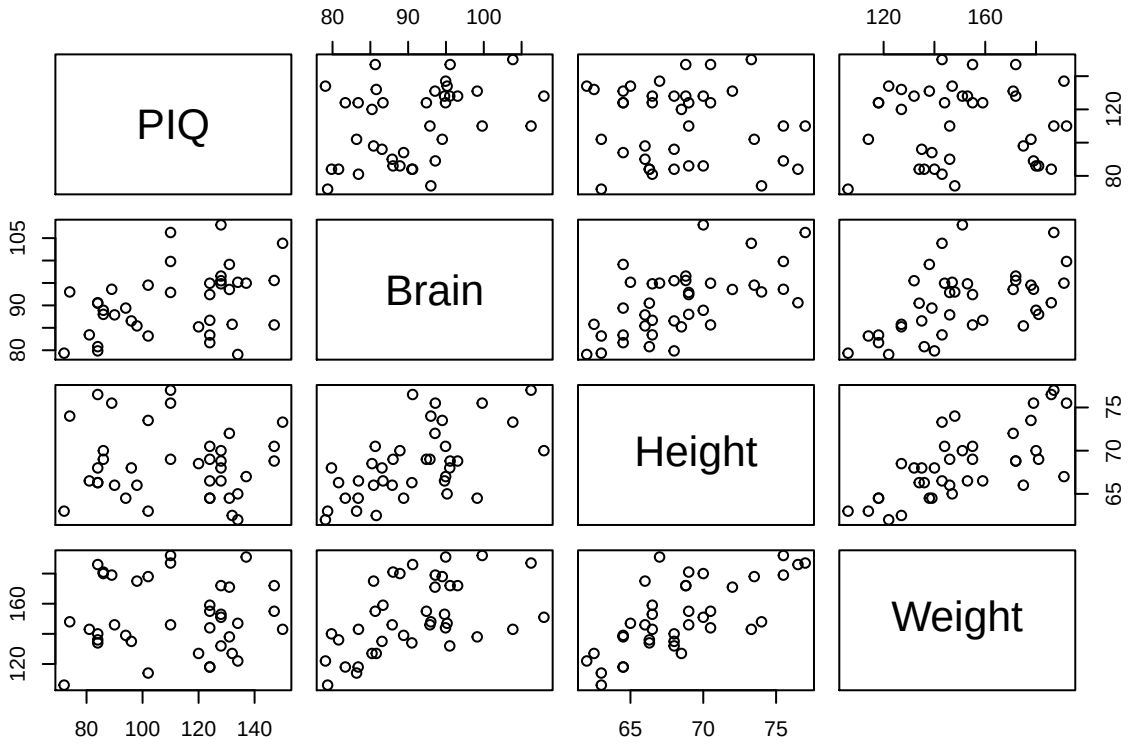
```
## Single term additions
##
## Model:
## y ~ x1 + x2
##        Df Sum of Sq    RSS    AIC F value Pr(>F)
## <none>              57.904 25.420
## x3      1    9.7939 48.111 25.011  1.8321 0.2089
## x4      1    9.9318 47.973 24.974  1.8633 0.2054
```

```
# Model:
# y ~ x1 + x2
#        Df Sum of Sq    RSS    AIC F value Pr(>F)
#               57.904 25.420
# x3      1    9.7939 48.111 25.011  1.8321 0.2089
# x4      1    9.9318 47.973 24.974  1.8633 0.2054

detach(cement)
```

IQ and body size (variable selection using stepwise regression)

Load the iqsize data. Create a scatterplot matrix of the data. Use the add1 and drop1 functions to conduct stepwise regression.

```
iqsize <- read.table("./Data/iqsize.txt", header=T)
attach(iqsize)

pairs(iqsize)
```




```
model.0 <- lm(PIQ ~ 1)
add1(model.0, ~ Brain + Height + Weight, test="F")
```

```
## Single term additions
##
## Model:
## PIQ ~ 1
##        Df Sum of Sq   RSS    AIC F value  Pr(>F)
## <none>              18895 237.94
## Brain   1   2697.09 16198 234.09  5.9945 0.01935 *
## Height  1    163.97 18731 239.61  0.3151 0.57802
## Weight  1      0.12 18894 239.94  0.0002 0.98806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# PIQ ~ 1
#        Df Sum of Sq   RSS    AIC F value  Pr(>F)
#              18895 237.94
# Brain   1   2697.09 16198 234.09  5.9945 0.01935 *
# Height  1    163.97 18731 239.61  0.3151 0.57802
# Weight  1      0.12 18894 239.94  0.0002 0.98806

model.1 <- lm(PIQ ~ Brain)
add1(model.1, ~ . + Height + Weight, test="F")
```

```
## Single term additions
##
## Model:
## PIQ ~ Brain
##        Df Sum of Sq   RSS    AIC F value   Pr(>F)
```

```
## <none>              16198 234.09
## Height  1   2875.65 13322 228.66  7.5551 0.009399 **
## Weight  1    940.94 15256 233.82  2.1586 0.150705
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# PIQ ~ Brain
#        Df Sum of Sq   RSS    AIC F value   Pr(>F)
#               16198 234.09
# Height  1   2875.65 13322 228.66  7.5551 0.009399 **
# Weight  1    940.94 15256 233.82  2.1586 0.150705

model.12 <- lm(PIQ ~ Brain + Height)
drop1(model.12, ~ ., test="F")
```

```
## Single term deletions
##
## Model:
## PIQ ~ Brain + Height
##        Df Sum of Sq   RSS    AIC F value    Pr(>F)
## <none>              13322 228.66
## Brain   1    5408.8 18731 239.61 14.2103 0.0006045 ***
## Height  1    2875.6 16198 234.09  7.5551 0.0093991 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# PIQ ~ Brain + Height
#        Df Sum of Sq   RSS    AIC F value    Pr(>F)
#               13322 228.66
# Brain   1    5408.8 18731 239.61 14.2103 0.0006045 ***
# Height  1    2875.6 16198 234.09  7.5551 0.0093991 **

add1(model.12, ~ . + Weight, test="F")
```

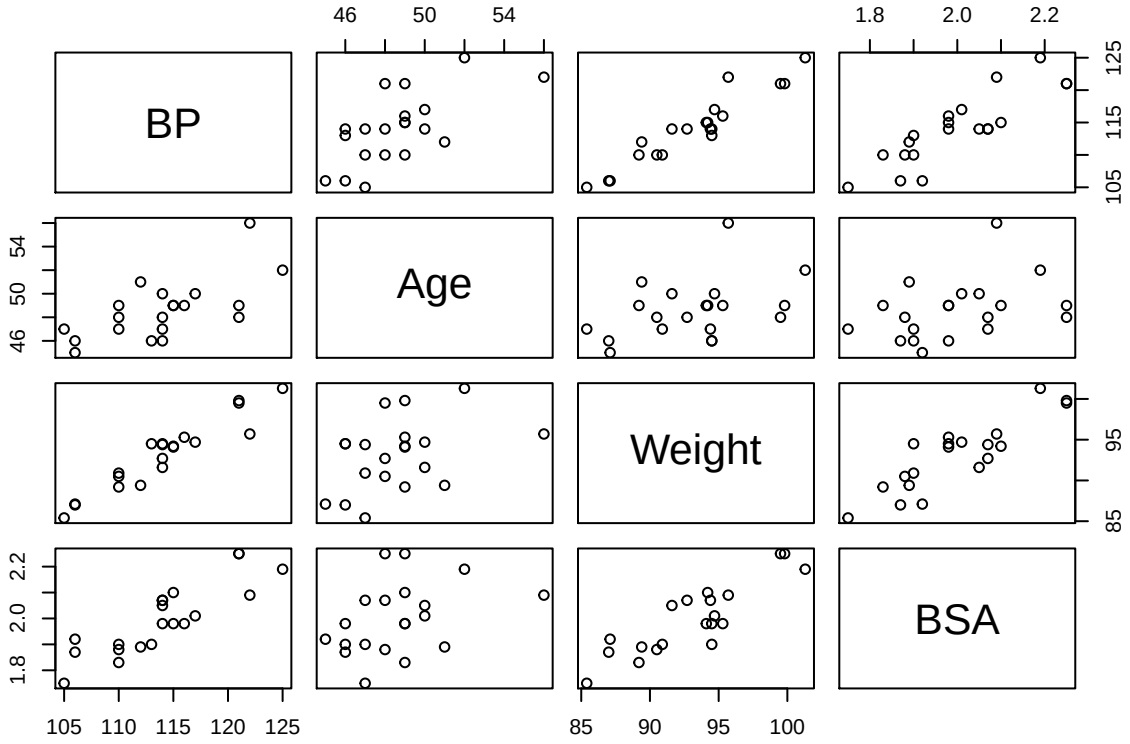
```
## Single term additions
##
## Model:
## PIQ ~ Brain + Height
##        Df Sum of Sq   RSS    AIC F value Pr(>F)
## <none>              13322 228.66
## Weight  1 0.0031633 13322 230.66       0 0.9977
```

```
# Model:
# PIQ ~ Brain + Height
#        Df Sum of Sq   RSS    AIC F value Pr(>F)
#               13322 228.66
# Weight  1 0.0031633 13322 230.66       0 0.9977

detach(iqsize)
```

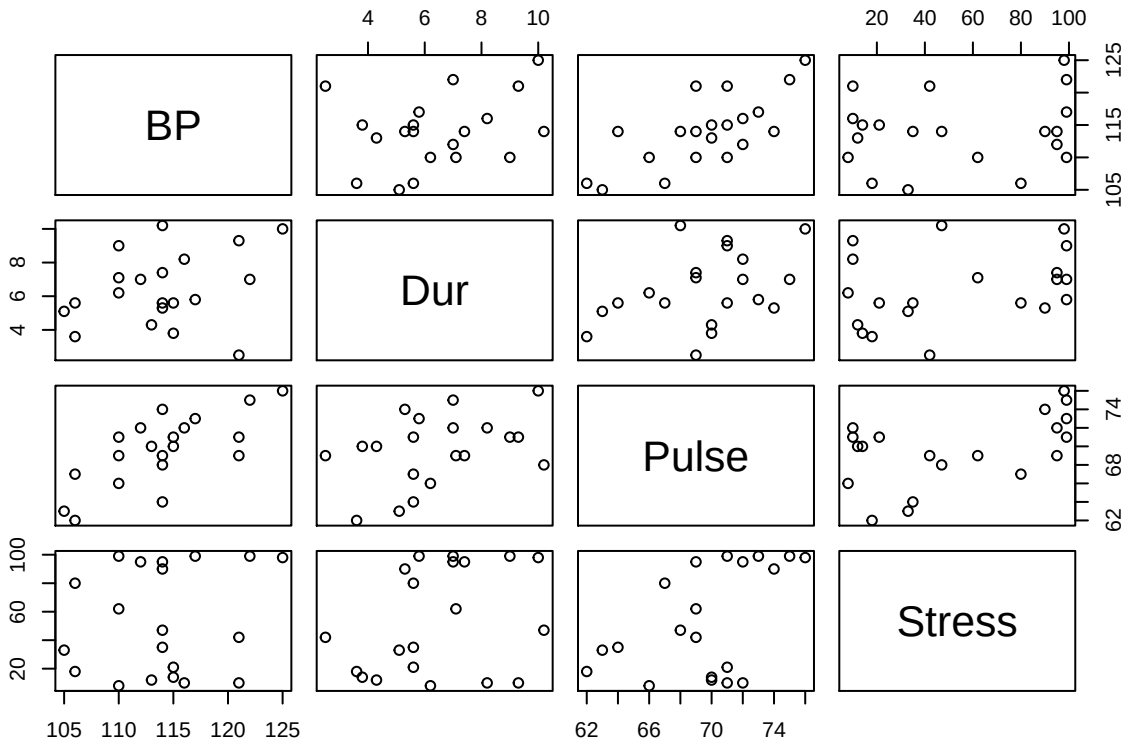
Blood pressure (variable selection using stepwise regression)

Load the bloodpress data. Create scatterplot matrices of the data. Use the add1 and drop1 functions to conduct stepwise regression.

```
bloodpress <- read.table("./Data/bloodpress.txt", header=T)
attach(bloodpress)

pairs(bloodpress[,c(2:5)])
```




```
pairs(bloodpress[,c(2,6:8)])
```



```
model.0 <- lm(BP  ~ 1)
add1(model.0, ~ Age + Weight + BSA + Dur + Pulse + Stress, test="F")
```

```
## Single term additions
##
## Model:
## BP ~ 1
##        Df Sum of Sq    RSS    AIC  F value    Pr(>F)
## <none>              560.00 68.644
## Age     1    243.27 316.73 59.247  13.8248 0.0015737 **
## Weight  1    505.47  54.53 24.060 166.8591 1.528e-10 ***
## BSA     1    419.86 140.14 42.938  53.9270 8.114e-07 ***
## Dur     1     48.02 511.98 68.851   1.6883 0.2102216
## Pulse   1    291.44 268.56 55.946  19.5342 0.0003307 ***
## Stress  1     15.04 544.96 70.099   0.4969 0.4898895
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# BP ~ 1
#        Df Sum of Sq    RSS    AIC  F value    Pr(>F)
#              560.00 68.644
# Age     1    243.27 316.73 59.247  13.8248 0.0015737 **
# Weight  1    505.47  54.53 24.060 166.8591 1.528e-10 ***
# BSA     1    419.86 140.14 42.938  53.9270 8.114e-07 ***
# Dur     1     48.02 511.98 68.851   1.6883 0.2102216
# Pulse   1    291.44 268.56 55.946  19.5342 0.0003307 ***
# Stress  1     15.04 544.96 70.099   0.4969 0.4898895
```

```
model.2 <- lm(BP ~ Weight)
add1(model.2, ~ . + Age + BSA + Dur + Pulse + Stress, test="F")
```

```
## Single term additions
##
## Model:
## BP ~ Weight
##        Df Sum of Sq    RSS     AIC  F value    Pr(>F)
## <none>              54.528  24.060
## Age     1    49.704  4.824 -22.443 175.1622 2.218e-10 ***
## BSA     1     2.814 51.714  25.000   0.9251   0.34962
## Dur     1     6.095 48.433  23.689   2.1393   0.16181
## Pulse   1     8.940 45.588  22.478   3.3338   0.08549 .
## Stress  1     9.660 44.868  22.160   3.6601   0.07273 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# BP ~ Weight
#        Df Sum of Sq    RSS     AIC  F value    Pr(>F)
#              54.528  24.060
# Age     1    49.704  4.824 -22.443 175.1622 2.218e-10 ***
# BSA     1     2.814 51.714  25.000   0.9251   0.34962
# Dur     1     6.095 48.433  23.689   2.1393   0.16181
# Pulse   1     8.940 45.588  22.478   3.3338   0.08549 .
# Stress  1     9.660 44.868  22.160   3.6601   0.07273 .
```

```r
model.12 <- lm(BP ~ Age + Weight)
drop1(model.12, ~ ., test="F")
```

```
## Single term deletions
##
## Model:
## BP ~ Age + Weight
##        Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>               4.82 -22.443
## Age     1    49.704  54.53  24.060  175.16 2.218e-10 ***
## Weight  1   311.910 316.73  59.247 1099.20 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Model:
# BP ~ Age + Weight
#        Df Sum of Sq    RSS     AIC F value    Pr(>F)
#                4.82 -22.443
# Age     1    49.704  54.53  24.060  175.16 2.218e-10 ***
# Weight  1   311.910 316.73  59.247 1099.20 < 2.2e-16 ***

add1(model.12, ~ . + BSA + Dur + Pulse + Stress, test="F")
```

```
## Single term additions
##
## Model:
## BP ~ Age + Weight
##        Df Sum of Sq    RSS     AIC F value   Pr(>F)
## <none>              4.8239 -22.443
## BSA     1   1.76778 3.0561 -29.572  9.2550 0.007764 **
## Dur     1   0.17835 4.6456 -21.196  0.6143 0.444639
## Pulse   1   0.49557 4.3284 -22.611  1.8319 0.194719
## Stress  1   0.16286 4.6611 -21.130  0.5591 0.465486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Model:
# BP ~ Age + Weight
#        Df Sum of Sq    RSS     AIC F value   Pr(>F)
#              4.8239 -22.443
# BSA     1   1.76778 3.0561 -29.572  9.2550 0.007764 **
# Dur     1   0.17835 4.6456 -21.196  0.6143 0.444639
# Pulse   1   0.49557 4.3284 -22.611  1.8319 0.194719
# Stress  1   0.16286 4.6611 -21.130  0.5591 0.465486

model.123 <- lm(BP ~ Age + Weight + BSA)
drop1(model.123, ~ ., test="F")
```

```
## Single term deletions
##
## Model:
## BP ~ Age + Weight + BSA
##        Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>               3.056 -29.572
## Age     1    48.658 51.714  25.000 254.740 3.002e-11 ***
```

```
## Weight  1    65.303 68.359  30.581 341.886 3.198e-12 ***
## BSA     1     1.768  4.824 -22.443   9.255  0.007764 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# BP ~ Age + Weight + BSA
#        Df Sum of Sq    RSS     AIC F value    Pr(>F)
#                3.056 -29.572
# Age     1    48.658 51.714  25.000 254.740 3.002e-11 ***
# Weight  1    65.303 68.359  30.581 341.886 3.198e-12 ***
# BSA     1     1.768  4.824 -22.443   9.255  0.007764 **

add1(model.123, ~ . + Dur + Pulse + Stress, test="F")
```

```
## Single term additions
##
## Model:
## BP ~ Age + Weight + BSA
##        Df Sum of Sq    RSS     AIC F value Pr(>F)
## <none>              3.0561 -29.572
## Dur     1   0.33510 2.7210 -29.894  1.8473 0.1942
## Pulse   1   0.04111 3.0150 -27.842  0.2045 0.6576
## Stress  1   0.21774 2.8384 -29.050  1.1507 0.3004
```

```
# Model:
# BP ~ Age + Weight + BSA
#        Df Sum of Sq    RSS     AIC F value Pr(>F)
#               3.0561 -29.572
# Dur     1   0.33510 2.7210 -29.894  1.8473 0.1942
# Pulse   1   0.04111 3.0150 -27.842  0.2045 0.6576
# Stress  1   0.21774 2.8384 -29.050  1.1507 0.3004

detach(bloodpress)
```

Cement hardening (variable selection using best subsets regression)

Load the cement data. Use the regsubsets function in the leaps package to conduct variable selection using exhaustive search (i.e., best subsets regression). Note that the nbest=2 argument returns the best two models with 1, 2, . . . , k predictors. Fit models with all four predictors (assumed unbiased) and just two predictors to retrieve the information needed to calculate for the model with just two predictors by hand. Fit model with , , and and note the variance inflation factors for and are very high. Fit model with , , and and note the variance inflation factors are acceptable. Fit model with and and note the variance inflation factors are acceptable, adjusted is high, and a residual analysis and normality test yields no concerns.

```
cement <- read.table("./Data/cement.txt", header=T)
attach(cement)

library(leaps)

subset <- regsubsets(y ~ x1 + x2 + x3 + x4, method="exhaustive", nbest=2, data=cement)
cbind(summary(subset)$outmat, round(summary(subset)$adjr2, 3), round(summary(subset)$cp, 1))
```

```
##          x1  x2  x3  x4
## 1  ( 1 ) " " " " " " "*" "0.645" "138.7"
## 1  ( 2 ) " " "*" " " " " "0.636" "142.5"
```

```
## 2  ( 1 ) "*" "*" " " " " " "0.974" "2.7"
## 2  ( 2 ) "*" " " " " "*" "0.967" "5.5"
## 3  ( 1 ) "*" "*" " " "*" "0.976" "3"
## 3  ( 2 ) "*" "*" "*" " " "0.976" "3"
## 4  ( 1 ) "*" "*" "*" "*" "0.974" "5"
```

```
#          x1  x2  x3  x4
# 1  ( 1 ) " " " " " " "*" "0.645" "138.7"
# 1  ( 2 ) " " "*" " " " " " "0.636" "142.5"
# 2  ( 1 ) "*" "*" " " " " " "0.974" "2.7"
# 2  ( 2 ) "*" " " " " "*" "0.967" "5.5"
# 3  ( 1 ) "*" "*" " " "*" "0.976" "3"
# 3  ( 2 ) "*" "*" "*" " " "0.976" "3"
# 4  ( 1 ) "*" "*" "*" "*" "0.974" "5"

model.1234 <- lm(y ~ x1 + x2 + x3 + x4)
model.12 <- lm(y ~ x1 + x2)

SSE.k <- sum(residuals(model.12)^2) # SSE_k = 57.90448
MSE.all <- summary(model.1234)$sigma^2 # MSE_all = 5.982955
params <- summary(model.12)$df[1] # k+1 = 3
n <- sum(summary(model.1234)$df[1:2]) # n = 13
SSE.k/MSE.all + 2*params - n # Cp = 2.678242
```

```
## [1] 2.678242
```

```
model.14 <- lm(y ~ x1 + x4)

SSE.k <- sum(residuals(model.14)^2) # SSE_k = 74.76211
params <- summary(model.14)$df[1] # k+1 = 3
SSE.k/MSE.all + 2*params - n # Cp = 5.495851
```

```
## [1] 5.495851
```

```
model.124 <- lm(y ~ x1 + x2 + x4)
library(car)
```

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

```
vif(model.124)
```

```
##       x1       x2       x4
##  1.06633 18.78031 18.94008
```

```
#      x1       x2       x4
# 1.06633 18.78031 18.94008

model.123 <- lm(y ~ x1 + x2 + x3)
vif(model.123)
```

```
##       x1       x2       x3
## 3.251068 1.063575 3.142125
```

```
#       x1       x2       x3
# 3.251068 1.063575 3.142125

summary(model.12)
```

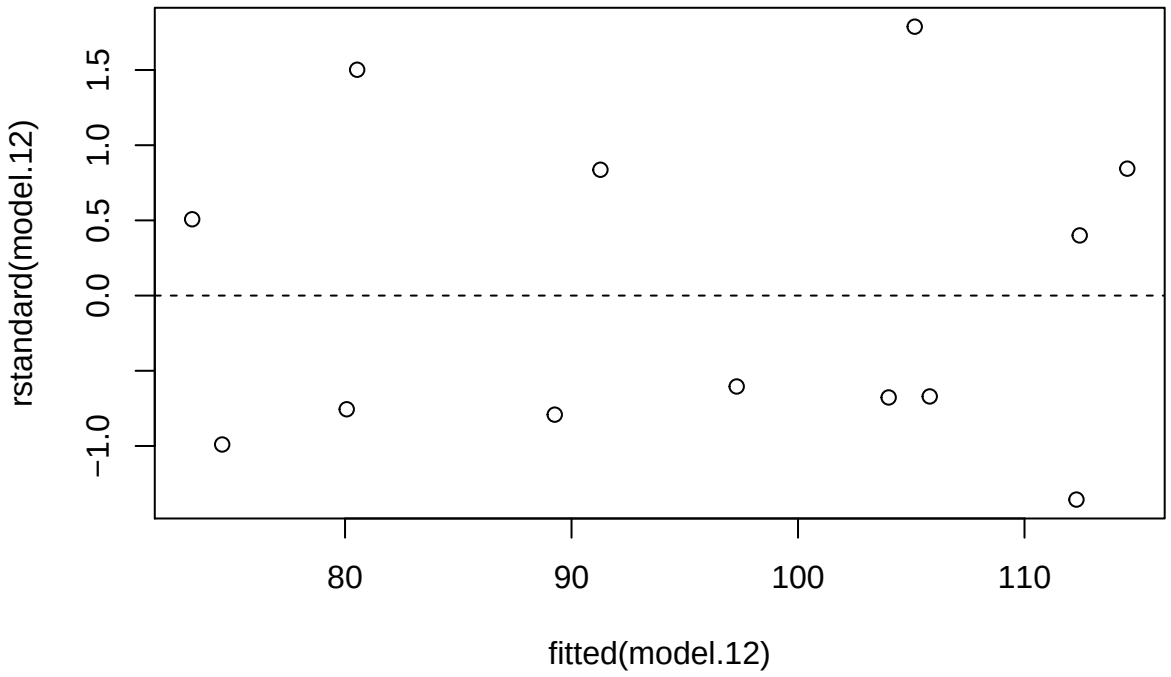
```
##
```

```
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.893 -1.574 -1.302  1.363  4.048
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 52.57735    2.28617   23.00 5.46e-10 ***
## x1           1.46831    0.12130   12.11 2.69e-07 ***
## x2           0.66225    0.04585   14.44 5.03e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.406 on 10 degrees of freedom
## Multiple R-squared:  0.9787, Adjusted R-squared:  0.9744
## F-statistic: 229.5 on 2 and 10 DF,  p-value: 4.407e-09
```

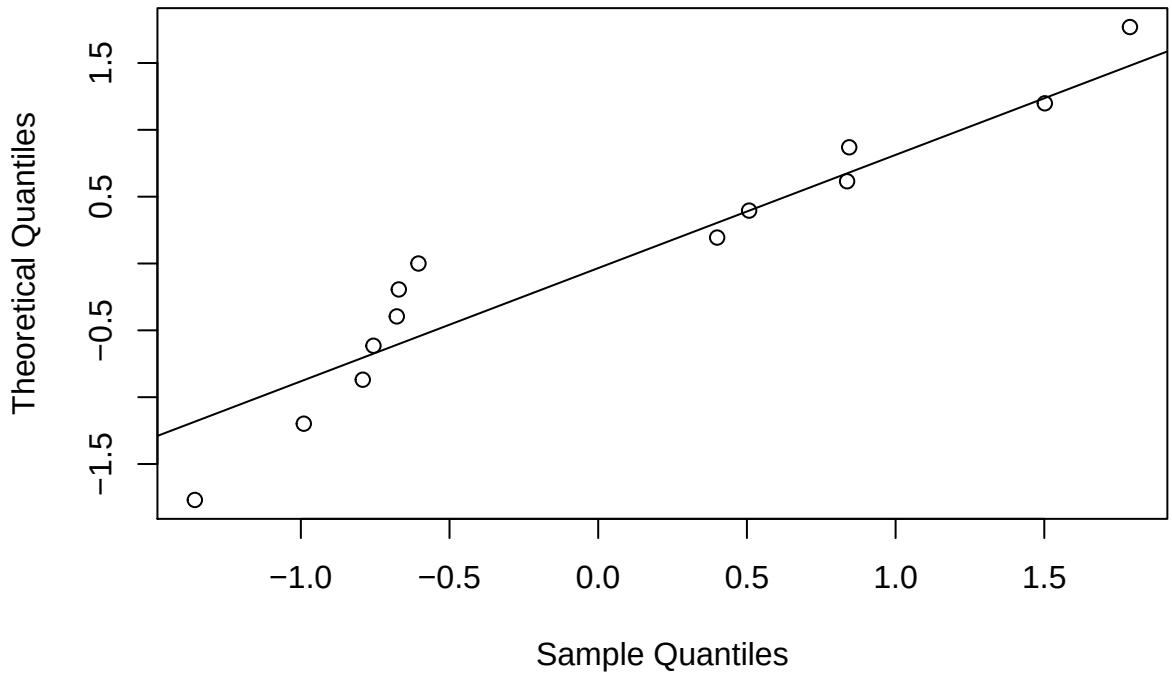
```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 52.57735    2.28617   23.00 5.46e-10 ***
# x1           1.46831    0.12130   12.11 2.69e-07 ***
# x2           0.66225    0.04585   14.44 5.03e-08 ***
# ---
# Residual standard error: 2.406 on 10 degrees of freedom
# Multiple R-squared:  0.9787,  Adjusted R-squared:  0.9744
# F-statistic: 229.5 on 2 and 10 DF,  p-value: 4.407e-09

vif(model.12)
```

```
##       x1       x2
## 1.055129 1.055129
```

```
#       x1       x2
# 1.055129 1.055129

plot(x=fitted(model.12), y=rstandard(model.12),
     panel.last = abline(h=0, lty=2))
```




```
qqnorm(rstandard(model.12), main="", datax=TRUE)
qqline(rstandard(model.12), datax=TRUE)
```




```
library(nortest)
ad.test(rstandard(model.12)) # A = 0.6136, p-value = 0.08628
```

```
##
##  Anderson-Darling normality test
##
## data:  rstandard(model.12)
## A = 0.61361, p-value = 0.08628
```

```
detach(cement)
```

IQ and body size (variable selection using best subsets regression)

Load the iqsize data. Use the regsubsets function in the leaps package to conduct variable selection using exhaustive search (i.e., best subsets regression). Fit model with Brain and Height and note the variance inflation factors are acceptable, adjusted is as good as it gets with this dataset, and a residual analysis and normality test yields no concerns.

```
iqsize <- read.table("./Data/iqsize.txt", header=T)
attach(iqsize)

subset <- regsubsets(PIQ ~ Brain + Height + Weight, method="exhaustive", nbest=2, data=iqsize)
cbind(summary(subset)$outmat, round(summary(subset)$adjr2, 3), round(summary(subset)$cp, 1))
```

```
##          Brain Height Weight
## 1  ( 1 ) "*"   " "    " "    "0.119"  "7.3"
## 1  ( 2 ) " "   "*"    " "    "-0.019" "13.8"
## 2  ( 1 ) "*"   "*"    " "    "0.255"  "2"
## 2  ( 2 ) "*"   " "    "*"    "0.146"  "6.9"
## 3  ( 1 ) "*"   "*"    "*"    "0.233"  "4"
```

```
#          Brain Height Weight
# 1  ( 1 ) "*"   " "    " "    "0.119"  "7.3"
# 1  ( 2 ) " "   "*"    " "    "-0.019" "13.8"
# 2  ( 1 ) "*"   "*"    " "    "0.255"  "2"
# 2  ( 2 ) "*"   " "    "*"    "0.146"  "6.9"
# 3  ( 1 ) "*"   "*"    "*"    "0.233"  "4"

model.12 <- lm(PIQ ~ Brain + Height)
summary(model.12)
```

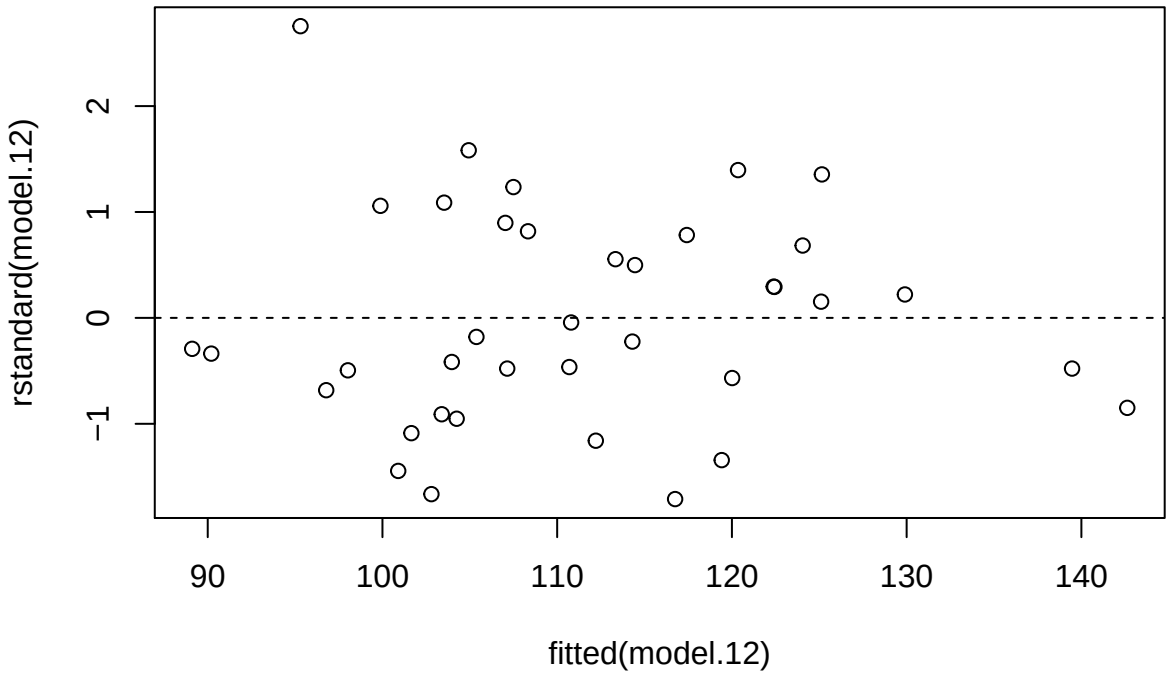
```
##
## Call:
## lm(formula = PIQ ~ Brain + Height)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.750 -12.090  -3.841  14.174  51.690
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 111.2757    55.8673   1.992 0.054243 .
## Brain         2.0606     0.5466   3.770 0.000604 ***
## Height       -2.7299     0.9932  -2.749 0.009399 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.51 on 35 degrees of freedom
## Multiple R-squared:  0.2949, Adjusted R-squared:  0.2546
## F-statistic: 7.321 on 2 and 35 DF,  p-value: 0.002208
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 111.2757    55.8673   1.992 0.054243 .
# Brain         2.0606     0.5466   3.770 0.000604 ***
```

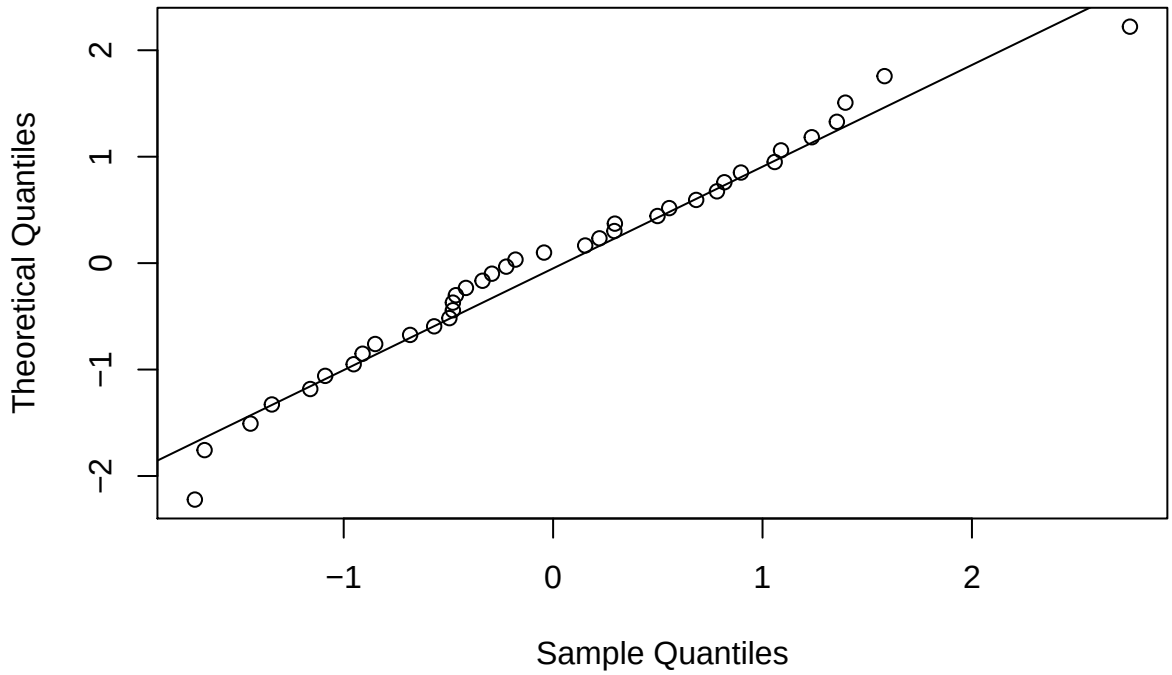
```
# Height       -2.7299     0.9932  -2.749 0.009399 **
# ---
# Residual standard error: 19.51 on 35 degrees of freedom
# Multiple R-squared:  0.2949,  Adjusted R-squared:  0.2546
# F-statistic: 7.321 on 2 and 35 DF,  p-value: 0.002208

vif(model.12)
```

```
##    Brain   Height
## 1.529463 1.529463
```

```
#    Brain   Height
# 1.529463 1.529463

plot(x=fitted(model.12), y=rstandard(model.12),
     panel.last = abline(h=0, lty=2))
```




```
qqnorm(rstandard(model.12), main="", datax=TRUE)
qqline(rstandard(model.12), datax=TRUE)
```




```
ad.test(rstandard(model.12)) # A = 0.2629, p-value = 0.6829
```

```
##
##  Anderson-Darling normality test
##
## data:  rstandard(model.12)
## A = 0.26289, p-value = 0.6829
```

```
detach(iqsize)
```

Blood pressure (variable selection using best subsets regression)

Load the bloodpress data. Use the regsubsets function in the leaps package to conduct variable selection using exhaustive search (i.e., best subsets regression). Fit model with Age and Weight and note the variance inflation factors are acceptable, adjusted can't get much better, and a residual analysis and normality test yields no concerns.

```
bloodpress <- read.table("./Data/bloodpress.txt", header=T)
attach(bloodpress)

subset <- regsubsets(BP ~ Age + Weight + BSA + Dur + Pulse + Stress,
                     method="exhaustive", nbest=2, data=bloodpress)
cbind(summary(subset)$outmat, round(summary(subset)$adjr2, 3),
      round(summary(subset)$cp, 1))
```

```
##          Age Weight BSA Dur Pulse Stress
## 1  ( 1 ) " " "*"    " " " " " " "    " "    "0.897" "312.8"
## 1  ( 2 ) " " " "    "*" " " " " "    " "    "0.736" "829.1"
## 2  ( 1 ) "*" "*"    " " " " " " "    " "    "0.99"  "15.1"
## 2  ( 2 ) " " "*"    " " " " " " "    "*"    "0.91"  "256.6"
## 3  ( 1 ) "*" "*"    "*" " " " " "    " "    "0.994" "6.4"
## 3  ( 2 ) "*" "*"    " " " " " " "*"  " "    "0.991" "14.1"
## 4  ( 1 ) "*" "*"    "*" "*" " " "    " "    "0.994" "6.4"
## 4  ( 2 ) "*" "*"    "*" " " " " "    "*"    "0.994" "7.1"
```

```
## 5  ( 1 ) "*" "*"    "*" " " "*"   "*"    "0.994" "7"
## 5  ( 2 ) "*" "*"    "*" "*" "*"   " "    "0.994" "7.7"
## 6  ( 1 ) "*" "*"    "*" "*" "*"   "*"    "0.994" "7"
```

```
#          Age Weight BSA Dur Pulse Stress
# 1  ( 1 ) " " "*"    " " " " " "   " "    "0.897" "312.8"
# 1  ( 2 ) " " " "    "*" " " " "   " "    "0.736" "829.1"
# 2  ( 1 ) "*" "*"    " " " " " "   " "    "0.99"  "15.1"
# 2  ( 2 ) " " "*"    " " " " " "   "*"    "0.91"  "256.6"
# 3  ( 1 ) "*" "*"    "*" " " " "   " "    "0.994" "6.4"
# 3  ( 2 ) "*" "*"    " " " " "*"   " "    "0.991" "14.1"
# 4  ( 1 ) "*" "*"    "*" "*" " "   " "    "0.994" "6.4"
# 4  ( 2 ) "*" "*"    "*" " " " "   "*"    "0.994" "7.1"
# 5  ( 1 ) "*" "*"    "*" " " "*"   "*"    "0.994" "7"
# 5  ( 2 ) "*" "*"    "*" "*" "*"   " "    "0.994" "7.7"
# 6  ( 1 ) "*" "*"    "*" "*" "*"   "*"    "0.994" "7"
```

```
model.12 <- lm(BP ~ Age + Weight)
summary(model.12)
```

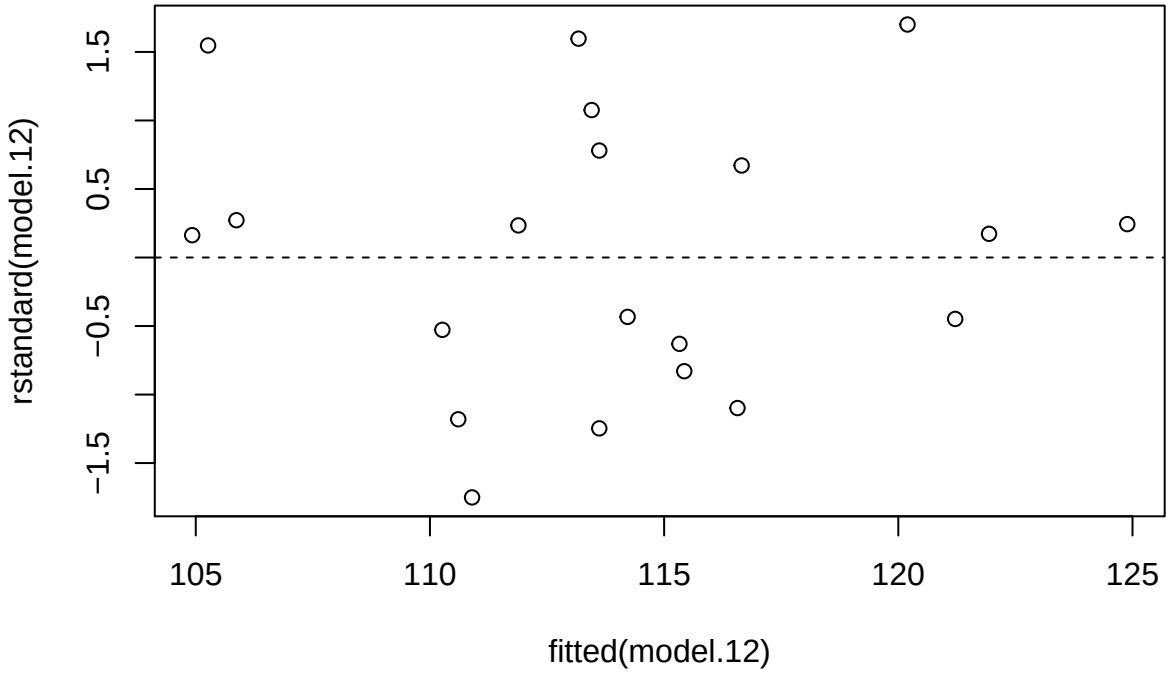
```
##
## Call:
## lm(formula = BP ~ Age + Weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89968 -0.35242  0.06979  0.35528  0.82781
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16.57937    3.00746  -5.513 3.80e-05 ***
## Age           0.70825    0.05351  13.235 2.22e-10 ***
## Weight        1.03296    0.03116  33.154  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5327 on 17 degrees of freedom
## Multiple R-squared:  0.9914, Adjusted R-squared:  0.9904
## F-statistic: 978.2 on 2 and 17 DF,  p-value: < 2.2e-16
```

```
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept) -16.57937    3.00746  -5.513 3.80e-05 ***
# Age           0.70825    0.05351  13.235 2.22e-10 ***
# Weight        1.03296    0.03116  33.154  < 2e-16 ***
# ---
# Residual standard error: 0.5327 on 17 degrees of freedom
# Multiple R-squared:  0.9914,  Adjusted R-squared:  0.9904
# F-statistic: 978.2 on 2 and 17 DF,  p-value: < 2.2e-16

vif(model.12)
```

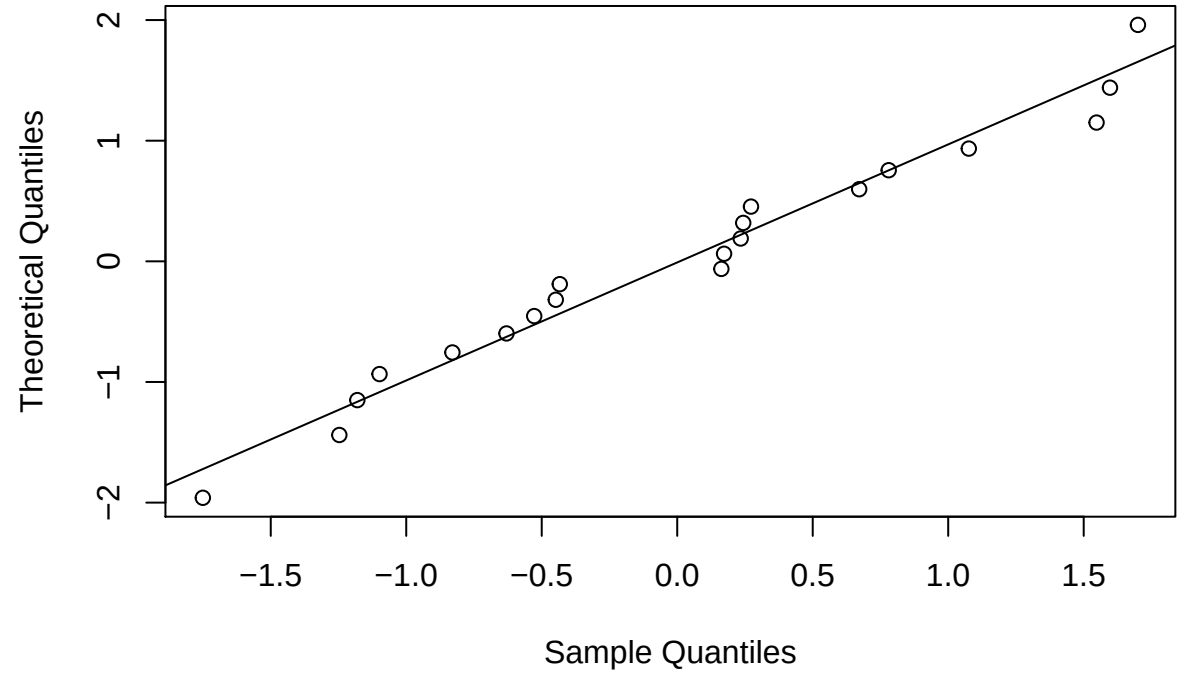
```
##      Age   Weight
## 1.198945 1.198945
```

```
#      Age   Weight
# 1.198945 1.198945
```

```
plot(x=fitted(model.12), y=rstandard(model.12),
     panel.last = abline(h=0, lty=2))
```




```
qqnorm(rstandard(model.12), main="", datax=TRUE)
qqline(rstandard(model.12), datax=TRUE)
```




```
ad.test(rstandard(model.12)) # A = 0.275, p-value = 0.6225
```

```
##
##  Anderson-Darling normality test
##
```

```
## data:  rstandard(model.12)
## A = 0.27496, p-value = 0.6225
```

```
detach(bloodpress)
```

Peruvian blood pressure (variable selection using best subsets regression)

Load the peru data. Use the regsubsets function in the leaps package to conduct variable selection using exhaustive search (i.e., best subsets regression). Fit the best 5-predictor and 4-predictor models. Calculate AIC and BIC by hand. Use the stepAIC function in the MASS package to conduct variable selection using a stepwise algorithm based on AIC or BIC.

```
peru <- read.table("./Data/peru.txt", header=T)
attach(peru)

fraclife <- Years/Age

n <- length(Systol) # 39

subset <- regsubsets(Systol ~ Age + Years + fraclife + Weight + Height + Chin +
                       Forearm + Pulse,
                     method="exhaustive", nbest=2, data=peru)
cbind(summary(subset)$outmat, round(summary(subset)$rsq, 3),
      round(summary(subset)$adjr2, 3), round(summary(subset)$cp, 1),
      round(sqrt(summary(subset)$rss/(n-c(rep(1:7,rep(2,7)),8)-1)), 4))
```

```
##          Age Years fraclife Weight Height Chin Forearm Pulse
## 1  ( 1 ) " " " "   " "      "*"    " "    " "  " "     " "   "0.272" "0.252"
## 1  ( 2 ) " " " "   "*"      " "    " "    " "  " "     " "   "0.076" "0.051"
## 2  ( 1 ) " " " "   "*"      "*"    " "    " "  " "     " "   "0.473" "0.444"
## 2  ( 2 ) " " "*"   " "      "*"    " "    " "  " "     " "   "0.421" "0.389"
## 3  ( 1 ) " " " "   "*"      "*"    " "    "*"  " "     " "   "0.503" "0.461"
## 3  ( 2 ) " " "*"   "*"      "*"    " "    " "  " "     " "   "0.49"  "0.447"
## 4  ( 1 ) "*" "*"   "*"      "*"    " "    " "  " "     " "   "0.597" "0.55"
## 4  ( 2 ) "*" "*"   "*"      " "    "*"    " "  " "     " "   "0.525" "0.469"
## 5  ( 1 ) "*" "*"   "*"      "*"    " "    "*"  " "     " "   "0.639" "0.584"
## 5  ( 2 ) "*" "*"   "*"      "*"    " "    " "  "*"     " "   "0.631" "0.576"
## 6  ( 1 ) "*" "*"   "*"      "*"    " "    "*"  "*"     " "   "0.649" "0.583"
## 6  ( 2 ) "*" "*"   "*"      "*"    "*"    "*"  " "     " "   "0.643" "0.576"
## 7  ( 1 ) "*" "*"   "*"      "*"    "*"    "*"  "*"     " "   "0.661" "0.584"
## 7  ( 2 ) "*" "*"   "*"      "*"    " "    "*"  "*"     "*"   "0.655" "0.577"
## 8  ( 1 ) "*" "*"   "*"      "*"    "*"    "*"  "*"     "*"   "0.666" "0.577"
##
## 1  ( 1 ) "30.5" "11.3376"
## 1  ( 2 ) "48.1" "12.7697"
## 2  ( 1 ) "14.4" "9.7772"
## 2  ( 2 ) "19.1" "10.2512"
## 3  ( 1 ) "13.7" "9.6273"
## 3  ( 2 ) "14.8" "9.7509"
## 4  ( 1 ) "7.2"  "8.7946"
## 4  ( 2 ) "13.7" "9.5502"
## 5  ( 1 ) "5.5"  "8.4571"
## 5  ( 2 ) "6.1"  "8.5417"
## 6  ( 1 ) "6.6"  "8.4663"
## 6  ( 2 ) "7.1"  "8.5337"
```

```
## 7  ( 1 ) "7.5"  "8.4556"
## 7  ( 2 ) "8"    "8.522"
## 8  ( 1 ) "9"    "8.5228"
```

```
#          Age Years fraclife Weight Height Chin Forearm Pulse
# 1  ( 1 ) " " " "   " "      "*"    " "    " "  " "     " "   "0.272" "0.252" "30.5" "11.3376"
# 1  ( 2 ) " " " "   "*"      " "    " "    " "  " "     " "   "0.076" "0.051" "48.1" "12.7697"
# 2  ( 1 ) " " " "   "*"      "*"    " "    " "  " "     " "   "0.473" "0.444" "14.4" "9.7772"
# 2  ( 2 ) " " "*"   " "      "*"    " "    " "  " "     " "   "0.421" "0.389" "19.1" "10.2512"
# 3  ( 1 ) " " " "   "*"      "*"    " "    "*"  " "     " "   "0.503" "0.461" "13.7" "9.6273"
# 3  ( 2 ) " " "*"   "*"      "*"    " "    " "  " "     " "   "0.49"  "0.447" "14.8" "9.7509"
# 4  ( 1 ) "*" "*"   "*"      "*"    " "    " "  " "     " "   "0.597" "0.55"  "7.2"  "8.7946"
# 4  ( 2 ) "*" "*"   "*"      " "    "*"    " "  " "     " "   "0.525" "0.469" "13.7" "9.5502"
# 5  ( 1 ) "*" "*"   "*"      "*"    " "    "*"  " "     " "   "0.639" "0.584" "5.5"  "8.4571"
# 5  ( 2 ) "*" "*"   "*"      "*"    " "    " "  "*"     " "   "0.631" "0.576" "6.1"  "8.5417"
# 6  ( 1 ) "*" "*"   "*"      "*"    " "    "*"  "*"     " "   "0.649" "0.583" "6.6"  "8.4663"
# 6  ( 2 ) "*" "*"   "*"      "*"    "*"    "*"  " "     " "   "0.643" "0.576" "7.1"  "8.5337"
# 7  ( 1 ) "*" "*"   "*"      "*"    "*"    "*"  "*"     " "   "0.661" "0.584" "7.5"  "8.4556"
# 7  ( 2 ) "*" "*"   "*"      "*"    " "    "*"  "*"     "*"   "0.655" "0.577" "8"    "8.522"
# 8  ( 1 ) "*" "*"   "*"      "*"    "*"    "*"  "*"     "*"   "0.666" "0.577" "9"    "8.5228"
```

```
model.5 <- lm(Systol ~ Age + Years + fraclife + Weight + Chin)
summary(model.5)
```

```
##
## Call:
## lm(formula = Systol ~ Age + Years + fraclife + Weight + Chin)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.520  -6.640  -1.093   4.893  16.366
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  109.3590    21.4843   5.090 1.41e-05 ***
## Age           -1.0120     0.3059  -3.308 0.002277 **
## Years          2.4067     0.7426   3.241 0.002723 **
## fraclife    -110.8112    27.2795  -4.062 0.000282 ***
## Weight         1.0976     0.2980   3.683 0.000819 ***
## Chin          -1.1918     0.6140  -1.941 0.060830 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.457 on 33 degrees of freedom
## Multiple R-squared:  0.6386, Adjusted R-squared:  0.5839
## F-statistic: 11.66 on 5 and 33 DF,  p-value: 1.531e-06
```

```
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  109.3590    21.4843   5.090 1.41e-05 ***
# Age           -1.0120     0.3059  -3.308 0.002277 **
# Years          2.4067     0.7426   3.241 0.002723 **
# fraclife    -110.8112    27.2795  -4.062 0.000282 ***
# Weight         1.0976     0.2980   3.683 0.000819 ***
# Chin          -1.1918     0.6140  -1.941 0.060830 .
# ---
```

```
# Residual standard error: 8.457 on 33 degrees of freedom
# Multiple R-squared:  0.6386,  Adjusted R-squared:  0.5839
# F-statistic: 11.66 on 5 and 33 DF,  p-value: 1.531e-06

k <- 5
n*log(sum(residuals(model.5)^2))-n*log(n)+2*(k+1) # AIC = 172.0151
```

```
## [1] 172.0151
```

```
n*log(sum(residuals(model.5)^2))-n*log(n)+log(n)*(k+1) # BIC = 181.9965
```

```
## [1] 181.9965
```

```
model.4 <- lm(Systol ~ Age + Years + fraclife + Weight)
summary(model.4)
```

```
##
## Call:
## lm(formula = Systol ~ Age + Years + fraclife + Weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.890  -5.976   0.058   5.407  16.835
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  116.8354    21.9797   5.316 6.69e-06 ***
## Age           -0.9507     0.3164  -3.004 0.004971 **
## Years          2.3393     0.7714   3.032 0.004621 **
## fraclife    -108.0728    28.3302  -3.815 0.000549 ***
## Weight         0.8324     0.2754   3.022 0.004742 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.795 on 34 degrees of freedom
## Multiple R-squared:  0.5974, Adjusted R-squared:   0.55
## F-statistic: 12.61 on 4 and 34 DF,  p-value: 2.142e-06
```

```
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  116.8354    21.9797   5.316 6.69e-06 ***
# Age           -0.9507     0.3164  -3.004 0.004971 **
# Years          2.3393     0.7714   3.032 0.004621 **
# fraclife    -108.0728    28.3302  -3.815 0.000549 ***
# Weight         0.8324     0.2754   3.022 0.004742 **
# ---
# Residual standard error: 8.795 on 34 degrees of freedom
# Multiple R-squared:  0.5974,  Adjusted R-squared:   0.55
# F-statistic: 12.61 on 4 and 34 DF,  p-value: 2.142e-06

k <- 4
n*log(sum(residuals(model.4)^2))-n*log(n)+2*(k+1) # AIC = 174.2316
```

```
## [1] 174.2316
```

```
n*log(sum(residuals(model.4)^2))-n*log(n)+log(n)*(k+1) # BIC = 182.5494
```

```
## [1] 182.5494
```

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
##     cement
```

```r
subset.aic <- stepAIC(lm(Systol ~ Age + Years + fraclife + Weight + Height +
                           Chin + Forearm + Pulse), direction="both", k=2)
```

```
## Start:  AIC=174.9
## Systol ~ Age + Years + fraclife + Weight + Height + Chin + Forearm +
##     Pulse
##
##            Df Sum of Sq    RSS    AIC
## - Pulse     1     37.31 2216.4 173.56
## - Height    1     72.23 2251.3 174.17
## <none>                  2179.1 174.90
## - Chin      1    130.95 2310.1 175.18
## - Forearm   1    146.06 2325.2 175.43
## - Years     1    677.20 2856.3 183.46
## - Weight    1    829.55 3008.7 185.48
## - Age       1    873.30 3052.4 186.04
## - fraclife  1   1103.76 3282.9 188.88
##
## Step:  AIC=173.56
## Systol ~ Age + Years + fraclife + Weight + Height + Chin + Forearm
##
##            Df Sum of Sq    RSS    AIC
## - Height    1     77.26 2293.7 172.90
## - Forearm   1    113.91 2330.3 173.52
## <none>                  2216.4 173.56
## - Chin      1    152.11 2368.5 174.15
## + Pulse     1     37.31 2179.1 174.90
## - Years     1    651.50 2867.9 181.61
## - Age       1    856.29 3072.7 184.30
## - Weight    1    880.23 3096.7 184.61
## - fraclife  1   1067.91 3284.3 186.90
##
## Step:  AIC=172.9
## Systol ~ Age + Years + fraclife + Weight + Chin + Forearm
##
##            Df Sum of Sq    RSS    AIC
## - Forearm   1     66.53 2360.2 172.01
## - Chin      1    114.02 2407.7 172.79
## <none>                  2293.7 172.90
## + Height    1     77.26 2216.4 173.56
## + Pulse     1     42.35 2251.3 174.17
## - Years     1    811.30 3105.0 182.71
## - Age       1    848.93 3142.6 183.18
## - Weight    1   1036.53 3330.2 185.44
## - fraclife  1   1246.44 3540.1 187.83
##
```

```
## Step:  AIC=172.02
## Systol ~ Age + Years + fraclife + Weight + Chin
##
##            Df Sum of Sq    RSS    AIC
## <none>                  2360.2 172.01
## + Forearm   1     66.53 2293.7 172.90
## + Height    1     29.88 2330.3 173.52
## + Pulse     1      9.84 2350.4 173.85
## - Chin      1    269.48 2629.7 174.23
## - Years     1    751.19 3111.4 180.79
## - Age       1    782.65 3142.9 181.18
## - Weight    1    970.26 3330.5 183.44
## - fraclife  1   1180.14 3540.4 185.83
```

```
# Step:  AIC=172.02
# Systol ~ Age + Years + fraclife + Weight + Chin

subset.bic <- stepAIC(lm(Systol ~ Age + Years + fraclife + Weight + Height +
                           Chin + Forearm + Pulse), direction="both", k=log(n))
```

```
## Start:  AIC=189.87
## Systol ~ Age + Years + fraclife + Weight + Height + Chin + Forearm +
##     Pulse
##
##            Df Sum of Sq    RSS    AIC
## - Pulse     1     37.31 2216.4 186.87
## - Height    1     72.23 2251.3 187.48
## - Chin      1    130.95 2310.1 188.49
## - Forearm   1    146.06 2325.2 188.74
## <none>                  2179.1 189.87
## - Years     1    677.20 2856.3 196.76
## - Weight    1    829.55 3008.7 198.79
## - Age       1    873.30 3052.4 199.35
## - fraclife  1   1103.76 3282.9 202.19
##
## Step:  AIC=186.87
## Systol ~ Age + Years + fraclife + Weight + Height + Chin + Forearm
##
##            Df Sum of Sq    RSS    AIC
## - Height    1     77.26 2293.7 184.54
## - Forearm   1    113.91 2330.3 185.16
## - Chin      1    152.11 2368.5 185.80
## <none>                  2216.4 186.87
## + Pulse     1     37.31 2179.1 189.87
## - Years     1    651.50 2867.9 193.26
## - Age       1    856.29 3072.7 195.95
## - Weight    1    880.23 3096.7 196.25
## - fraclife  1   1067.91 3284.3 198.55
##
## Step:  AIC=184.54
## Systol ~ Age + Years + fraclife + Weight + Chin + Forearm
##
##            Df Sum of Sq    RSS    AIC
## - Forearm   1     66.53 2360.2 182.00
## - Chin      1    114.02 2407.7 182.77
```

```
## <none>                  2293.7 184.54
## + Height    1     77.26 2216.4 186.87
## + Pulse     1     42.35 2251.3 187.48
## - Years     1    811.30 3105.0 192.69
## - Age       1    848.93 3142.6 193.16
## - Weight    1   1036.53 3330.2 195.42
## - fraclife  1   1246.44 3540.1 197.81
##
## Step:  AIC=182
## Systol ~ Age + Years + fraclife + Weight + Chin
##
##            Df Sum of Sq    RSS    AIC
## <none>                  2360.2 182.00
## - Chin      1    269.48 2629.7 182.55
## + Forearm   1     66.53 2293.7 184.54
## + Height    1     29.88 2330.3 185.16
## + Pulse     1      9.84 2350.4 185.50
## - Years     1    751.19 3111.4 189.11
## - Age       1    782.65 3142.9 189.50
## - Weight    1    970.26 3330.5 191.76
## - fraclife  1   1180.14 3540.4 194.15
```

```
# Step:  AIC=182
# Systol ~ Age + Years + fraclife + Weight + Chin

detach(peru)
```

Measurements of college students (variable selection using stepwise regression)

Load the Physical data. Use the add1 and drop1 functions to conduct stepwise regression. Use the regsubsets function to conduct variable selection using backward elimination. Use the regsubsets function to conduct variable selection using forward selection.

```
physical <- read.table("./Data/Physical.txt", header=T)
attach(physical)

gender <- ifelse(Sex=="Female",1,0)

model.0 <- lm(Height ~ 1)
add1(model.0, ~ LeftArm + LeftFoot + LeftHand + HeadCirc + nose + gender, test="F")
```

```
## Single term additions
##
## Model:
## Height ~ 1
##          Df Sum of Sq     RSS    AIC  F value    Pr(>F)
## <none>                1054.75 164.46
## LeftArm   1    590.21  464.53 121.35  67.3396 5.252e-11 ***
## LeftFoot  1    707.42  347.33 105.36 107.9484 2.172e-14 ***
## LeftHand  1    143.59  911.15 158.41   8.3525  0.005570 **
## HeadCirc  1    189.24  865.51 155.58  11.5880  0.001272 **
## nose      1     85.25  969.49 161.82   4.6605  0.035412 *
## gender    1    533.24  521.51 127.72  54.1923 1.181e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# Height ~ 1
#          Df Sum of Sq     RSS    AIC  F value    Pr(>F)
#                1054.75 164.46
# LeftArm   1    590.21  464.53 121.35  67.3396 5.252e-11 ***
# LeftFoot  1    707.42  347.33 105.36 107.9484 2.172e-14 ***
# LeftHand  1    143.59  911.15 158.41   8.3525  0.005570 **
# HeadCirc  1    189.24  865.51 155.58  11.5880  0.001272 **
# nose      1     85.25  969.49 161.82   4.6605  0.035412 *
# gender    1    533.24  521.51 127.72  54.1923 1.181e-09 ***

model.2 <- lm(Height ~ LeftFoot)
add1(model.2, ~ . + LeftArm + LeftHand + HeadCirc + nose + gender, test="F")
```

```
## Single term additions
##
## Model:
## Height ~ LeftFoot
##          Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                347.33 105.361
## LeftArm   1   107.143 240.18  87.074 23.1967 1.305e-05 ***
## LeftHand  1    15.359 331.97 104.874  2.4059    0.1269
## HeadCirc  1     2.313 345.01 106.994  0.3486    0.5575
## nose      1     1.449 345.88 107.131  0.2178    0.6427
## gender    1    15.973 331.35 104.772  2.5066    0.1194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# Height ~ LeftFoot
#         Df Sum of Sq    RSS     AIC F value    Pr(>F)
#                347.33 105.361
# LeftArm   1   107.143 240.18  87.074 23.1967 1.305e-05 ***
# LeftHand  1    15.359 331.97 104.874  2.4059    0.1269
# HeadCirc  1     2.313 345.01 106.994  0.3486    0.5575
# nose      1     1.449 345.88 107.131  0.2178    0.6427
# gender    1    15.973 331.35 104.772  2.5066    0.1194

model.12 <- lm(Height ~ LeftArm + LeftFoot)
drop1(model.12, ~ ., test="F")
```

```
## Single term deletions
##
## Model:
## Height ~ LeftArm + LeftFoot
##          Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                240.18  87.074
## LeftArm   1    107.14 347.33 105.361  23.197 1.305e-05 ***
## LeftFoot  1    224.35 464.53 121.353  48.572 5.538e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Model:
# Height ~ LeftArm + LeftFoot
#          Df Sum of Sq    RSS     AIC F value    Pr(>F)
```

```
#                 240.18  87.074
# LeftArm   1    107.14 347.33 105.361  23.197 1.305e-05 ***
# LeftFoot  1    224.35 464.53 121.353  48.572 5.538e-09 ***

add1(model.12, ~ . + LeftHand + HeadCirc + nose + gender, test="F")
```

```
## Single term additions
##
## Model:
## Height ~ LeftArm + LeftFoot
##          Df Sum of Sq    RSS    AIC F value Pr(>F)
## <none>                240.18 87.074
## LeftHand  1    3.7854 236.40 88.200  0.8167 0.3704
## HeadCirc  1    1.4016 238.78 88.752  0.2994 0.5867
## nose      1    0.4463 239.74 88.971  0.0950 0.7592
## gender    1    3.7530 236.43 88.207  0.8096 0.3725
```

```
# Model:
# Height ~ LeftArm + LeftFoot
#          Df Sum of Sq    RSS    AIC F value Pr(>F)
#                 240.18 87.074
# LeftHand  1    3.7854 236.40 88.200  0.8167 0.3704
# HeadCirc  1    1.4016 238.78 88.752  0.2994 0.5867
# nose      1    0.4463 239.74 88.971  0.0950 0.7592
# gender    1    3.7530 236.43 88.207  0.8096 0.3725

subset <- regsubsets(Height ~ LeftArm + LeftFoot + LeftHand + HeadCirc + nose + gender,
                     method="backward", data=physical)

subset <- regsubsets(Height ~ LeftArm + LeftFoot + LeftHand + HeadCirc + nose + gender,
                     method="forward", data=physical)

detach(physical)
```