Assignment 4

Q 4.1

Using the methods described in this chapter and the family lung function data described in Appendix A, and choosing from among the variables OCAGE, OCWEIGHT, MHEIGHT, MWEIGHT, FHEIGHT, and FWEIGHT, select the variables that best predict height in the oldest child. Show your analysis.

```
Start:  AIC=581.65
OCHEIGHT ~ 1

           Df Sum of Sq    RSS    AIC
+ OCAGE     1    5937.9 1212.8 317.51
+ OCWEIGHT  1    5818.3 1332.4 331.62
+ FHEIGHT   1     209.8 6940.9 579.18
+ FWEIGHT   1     109.6 7041.0 581.33
+ MHEIGHT   1     104.9 7045.8 581.43
<none>                  7150.7 581.65
+ MWEIGHT   1      52.0 7098.7 582.55

Step:  AIC=317.51
OCHEIGHT ~ OCAGE

           Df Sum of Sq    RSS    AIC
+ OCWEIGHT  1     439.2  773.6 252.06
+ FHEIGHT   1     219.5  993.3 289.55
+ MHEIGHT   1     145.8 1067.0 300.30
+ FWEIGHT   1      52.2 1160.6 312.91
<none>                  1212.8 317.51
+ MWEIGHT   1       3.6 1209.2 319.06
- OCAGE     1    5937.9 7150.7 581.65

Step:  AIC=252.06
OCHEIGHT ~ OCAGE + OCWEIGHT

           Df Sum of Sq     RSS    AIC
+ FHEIGHT   1    121.28  652.28 228.48
+ MHEIGHT   1    109.68  663.88 231.12
<none>                   773.56 252.06
+ FWEIGHT   1      3.46  770.10 253.38
+ MWEIGHT   1      2.72  770.84 253.53
- OCWEIGHT  1    439.24 1212.80 317.51
- OCAGE     1    558.88 1332.44 331.62

Step:  AIC=228.48
OCHEIGHT ~ OCAGE + OCWEIGHT + FHEIGHT

           Df Sum of Sq     RSS    AIC
+ MHEIGHT   1     60.39  591.89 215.90
+ FWEIGHT   1     17.48  634.80 226.40
<none>                   652.28 228.48
+ MWEIGHT   1      3.15  649.12 229.75
- FHEIGHT   1    121.28  773.56 252.06
- OCWEIGHT  1    340.99  993.26 289.55
- OCAGE     1    631.52 1283.80 328.04

Step:  AIC=215.9
OCHEIGHT ~ OCAGE + OCWEIGHT + FHEIGHT + MHEIGHT
```

```
           Df Sum of Sq     RSS    AIC
+ MWEIGHT   1     20.36  571.52 212.65
+ FWEIGHT   1     10.04  581.85 215.34
<none>                   591.89 215.90
- MHEIGHT   1     60.39  652.28 228.48
- FHEIGHT   1     71.99  663.88 231.12
- OCWEIGHT  1    333.45  925.34 280.93
- OCAGE     1    644.19 1236.08 324.36

Step:  AIC=212.65
OCHEIGHT ~ OCAGE + OCWEIGHT + FHEIGHT + MHEIGHT + MWEIGHT

           Df Sum of Sq     RSS    AIC
<none>                   571.52 212.65
+ FWEIGHT   1      4.67  566.85 213.42
- MWEIGHT   1     20.36  591.89 215.90
- FHEIGHT   1     65.53  637.06 226.93
- MHEIGHT   1     77.60  649.12 229.75
- OCWEIGHT  1    351.66  923.18 282.58
- OCAGE     1    616.35 1187.88 320.39

Call:
lm(formula = OCHEIGHT ~ OCAGE + OCWEIGHT + FHEIGHT + MHEIGHT +
    MWEIGHT)

Coefficients:
(Intercept)        OCAGE     OCWEIGHT      FHEIGHT      MHEIGHT
    0.52776      1.21129      0.07769      0.25345      0.32174
    MWEIGHT
   -0.01282
```

Q 4.2 From among the candidate variables given in Problem 8.11,find the subset of three variables that best predicts height in the oldest child, separately for boys and girls. Are the two sets the same? Find the best subset of three variables for the group as a whole. Does adding OCSEX into the regression equation improve the fit?

(i)**Both girls and Boys of Oldest Child:**

```
Subset selection object
Call: regsubsets.formula(OCHEIGHT ~ OCAGE + OCWEIGHT + MHEIGHT + MWEIGHT +
    FHEIGHT + FWEIGHT, data = lung_data, nvmax = 3, method = "backward")
6 Variables  (and intercept)
         Forced in Forced out
OCAGE        FALSE      FALSE
OCWEIGHT     FALSE      FALSE
MHEIGHT      FALSE      FALSE
MWEIGHT      FALSE      FALSE
FHEIGHT      FALSE      FALSE
FWEIGHT      FALSE      FALSE
1 subsets of each size up to 3
Selection Algorithm: backward
         OCAGE OCWEIGHT MHEIGHT MWEIGHT FHEIGHT FWEIGHT
1  ( 1 ) "*"   " "      " "     " "     " "     " "
2  ( 1 ) "*"   "*"      " "     " "     " "     " "
3  ( 1 ) "*"   "*"      " "     " "     "*"     " "
```

Best variables are "OCAGE","OCWEIGHT","FHEIGHT"

(ii) **Oldest Child Boy**

```
Subset selection object
Call: regsubsets.formula(OCHEIGHT ~ OCAGE + OCWEIGHT + MHEIGHT + MWEIGHT +
    FHEIGHT + FWEIGHT, data = lung_data_boys, nvmax = 3, method = "forward")
6 Variables  (and intercept)
         Forced in Forced out
OCAGE        FALSE      FALSE
OCWEIGHT     FALSE      FALSE
MHEIGHT      FALSE      FALSE
MWEIGHT      FALSE      FALSE
FHEIGHT      FALSE      FALSE
FWEIGHT      FALSE      FALSE
1 subsets of each size up to 3
Selection Algorithm: forward
         OCAGE OCWEIGHT MHEIGHT MWEIGHT FHEIGHT FWEIGHT
1  ( 1 ) "*"   " "      " "     " "     " "     " "
2  ( 1 ) "*"   "*"      " "     " "     " "     " "
3  ( 1 ) "*"   "*"      " "     " "     "*"     " "
```

Best variables are "OCAGE","OCWEIGHT","FHEIGHT"

(iii)**Oldest Child Girl**

```
Subset selection object
Call: regsubsets.formula(OCHEIGHT ~ OCAGE + OCWEIGHT + MHEIGHT + MWEIGHT +
    FHEIGHT + FWEIGHT, data = lung_data_girls, nvmax = 3, method = "forward")
6 Variables  (and intercept)
         Forced in Forced out
OCAGE        FALSE      FALSE
OCWEIGHT     FALSE      FALSE
MHEIGHT      FALSE      FALSE
MWEIGHT      FALSE      FALSE
FHEIGHT      FALSE      FALSE
FWEIGHT      FALSE      FALSE
1 subsets of each size up to 3
Selection Algorithm: forward
         OCAGE OCWEIGHT MHEIGHT MWEIGHT FHEIGHT FWEIGHT
1  ( 1 ) "*"   " "      " "     " "     " "     " "
2  ( 1 ) "*"   " "      " "     " "     "*"     " "
3  ( 1 ) "*"   "*"      " "     " "     "*"     " "
```

Best variables are "OCAGE","OCWEIGHT","FHEIGHT"

4.3

Using the Parental HIV data find the best model that predicts the age at which adolescents started drinking alcohol. Since the data were collected retrospectively, only consider variables which might be considered representativeofthetime beforethe adolescentstarteddrinkingalcohol.

The best model is:

```
AGEALC ~ AGEMAR + AGESMOKE + SMOKEP3M + NGHB5 +
    NGHB2 + GENDER + SIBLINGS + NGHB1 + NGHB3
```

```
Call:
lm(formula = AGEALC ~ AGEMAR + AGESMOKE + SMOKEP3M + NGHB5 +
    NGHB2 + GENDER + SIBLINGS + NGHB1 + NGHB3)

Coefficients:
(Intercept)      AGEMAR1      AGEMAR2      AGEMAR3      AGEMAR4    AGESMOKE1
AGESMOKE2    AGESMOKE3
    0.26873      1.34823      1.61988      1.76995      1.32280      1.17813
1.99694      2.74984
  AGESMOKE4    AGESMOKE8    SMOKEP3M0   SMOKEP3M10   SMOKEP3M11   SMOKEP3M12
SMOKEP3M13   SMOKEP3M14
    2.67029      0.77423     -0.08244      0.99010      0.68595      0.56477
1.27721      2.63449
 SMOKEP3M15   SMOKEP3M16   SMOKEP3M17   SMOKEP3M18    SMOKEP3M2    SMOKEP3M5
SMOKEP3M6    SMOKEP3M7
    2.32120      2.35280      1.84522      0.23712      2.71643     -0.30685
-0.70201      0.78853
  SMOKEP3M8    SMOKEP3M9        NGHB5        NGHB2       GENDER    SIBLINGS1
SIBLINGS2    SIBLINGS3
   -2.14369      3.53918     -0.07046      0.11987     -0.30297     -0.06449
0.13255      0.32963
      NGHB1        NGHB3
```

0.11221 0.21885