R Material for Chapter 09

```
> attach(bill.data)

> bill.data
   bill income persons sqft
1   228   3220       2 1160           ## our data
2   156   2750       1 1080
3   648   3620       2 1720
4   528   3940       1 1840
5   552   4510       3 2240
6   636   3990       4 2190
7   444   2430       1  830
8   144   3070       1 1150
9   744   3750       2 1570
10 1104   4790       5 2660
11  204   2490       1  900
12  420   3600       3 1680
13  876   5370       1 2550
14  840   3180       7 1770
15  876   5910       2 2960
16  276    320       2 1190
17 1236   5920       3 3130
18  372   3520       2 1560
19  276   3720       1 1510
20  450   4840       1 2190
```

All-Subsets Selection of a Model – Requires a Special Library

```
> library(leaps)    ##  loads the special library of functions
                    ##  SEE THE END OF THIS DOCUMENT

> subs <- regsubsets(bill~.,data=bill.data,nbest=3,method=c("exhaustive"))
     ## (response variable~., name of data set, number of best subsets per model size,
     ## "exhaustive" means all-subsets
> summary(subs)

Subset selection object
Call: regsubsets.formula(bill ~ ., data = bill.data, nbest = 3, method = c("exhaustive"))
3 Variables  (and intercept)
        Forced in Forced out
income      FALSE      FALSE
persons     FALSE      FALSE
sqft        FALSE      FALSE
3 subsets of each size up to 3
Selection Algorithm: exhaustive
         income persons sqft
1  ( 1 ) " "    " "     "*"    ## orders the one variable models by R^2
1  ( 2 ) "*"    " "     " "
1  ( 3 ) " "    "*"     " "
2  ( 1 ) " "    "*"     "*"    ## orders the two variable models by R^2
2  ( 2 ) "*"    " "     "*"
2  ( 3 ) "*"    "*"     " "
3  ( 1 ) "*"    "*"     "*"    ## the only three variable model

> result1 <- with(bill.data,leaps(cbind(income, persons,sqft),bill,method="r2", nbest=3))
     ## with(name of data set, leaps( list of predictors, response variable, selection
     ## criterion,  number of best subsets per model size))
> plot(result1$size,result1$r2,xlab="# predictors +1",ylab="R-Square") ## plot of r^2 values
> result1$r2   ## printing the r^2 variable in the result1 object

[1] 0.7171831 0.4879795 0.2947749 0.7834572 0.7231502 0.6983454 0.7842642
```

```
      ## these are the r^2 values for each of the above models
> result2 <- with(bill.data,leaps(cbind(income, persons,sqft),bill,method="adjr2", nbest=3))
> plot(result2$size,result2$adjr2,xlab="# predictors +1",ylab="Adjusted R-Square")
> result2$adjr2
[1] 0.7014711 0.4595339 0.2555957 0.7579816 0.6905797 0.6628566 0.7438137
      ## these are the adjusted r^2 values for each of the above models

> result3 <- with(bill.data,leaps(cbind(income, persons,sqft),bill,method="Cp", nbest=3))
> plot(result3$size,result3$Cp,xlab="# predictors +1",ylab="Cp")
> result3$Cp
[1]  4.975051 21.973893 36.302866  2.059850  6.532504  8.372148  4.000000
      ## these are the Cp values for each of the above models

> n <- nrow(bill.data)     ## number of data points
> n
[1] 20

> SSE <- (1-result1$r2)*var(bill)*(n-1)             ## error sum of squares
> AIC <- n*log(SSE) -n*log(n) +2*result1$size       ## computing AIC values
> plot(result1$size,AIC,xlab="# predictors +1",ylab="AIC")
> AIC
[1] 207.8968 219.7681 226.1712 204.5566 209.4703 211.1865 206.4819
      ## these are the AIC values for each of the above models

> SBC <- n*log(SSE) -n*log(n) + log(n)*result1$size     ## computing SBC values
> plot(result1$size,SBC,xlab="# predictors +1",ylab="SBC")
> SBC
[1] 209.8883 221.7596 228.1626 207.5438 212.4575 214.1737 210.4649
      ## these are the SBC values for each of the above models
```

A Convenient Stepwise Search Process Based on AIC – Does not require a Special Library

```
> reg <- lm(bill~income+persons+sqft)    ## model with all potential predictors

> step(reg)         ## begins stepwise elimination of variables based on the AIC criterion
Start:  AIC=206.48   ## starting AIC with all variables in the model
bill ~ income + persons + sqft

          Df Sum of Sq    RSS    AIC
- income   1      1527 409856 204.56     ## It considers eliminating one of the variables
<none>                 408329 206.48     ## basing a decision on trying to reduce the AIC.
- persons  1    115672 524001 209.47     ## It chooses income because its elimination
- sqft     1    162621 570950 211.19     ## produces the smallest AIC and it is smaller
                                         ## than eliminating none of the predictors.
Step:  AIC=204.56
bill ~ persons + sqft

          Df Sum of Sq     RSS    AIC
<none>                  409856 204.56   ## Eliminating none of the predictors has the
- persons  1    125439  535295 207.90   ## smallest AIC, so this model is chosen and is
- sqft     1    924942 1334799 226.17   ## displayed below.

Call:
lm(formula = bill ~ persons + sqft)

Coefficients:
(Intercept)      persons         sqft
   -202.670       54.874        0.351
```

Validation of a Model using PRESS – Requires a Special Library

```
> reg <- lm(bill~persons+sqft)
> anova(reg)
Analysis of Variance Table

Response: bill
          Df Sum Sq Mean Sq F value    Pr(>F)
persons    1 557928  557928  23.142 0.0001632 ***
sqft       1 924942  924942  38.365  9.81e-06 ***
Residuals 17 409856   24109
---
Signif. Codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> library(boot)   ## Special Library   -  SEE THE END OF THIS DOCUMENT

> model2 <- glm(bill~sqft+persons)   ##using the best two predictor model (best overall)
> MSPRESS <- cv.glm(bill.data, model2)$delta[1]
> MSPRESS
      1
27437.5
> n <- nrow(bill.data)     ## number of data points
> n
[1] 20
> PRESS <- n*MSPRESS
> PRESS
     1
548750
```

Loading External Libraries:

From time to time we need to use libraries of programs stored on the R system but not included in the installation of R.

Suppose you need to download the set of programs in the library called “leaps”. This can be accomplished in the following manner. First you must be in a setting in which you are connected to the internet. When in the R workspace window type:

```
> install.packages(“leaps”)
                        ## You will have to answer a few questions in the process,
                        ## but it will download onto your computer.
                        ## Then type:

> library(leaps)    ## This brings the library into your active workspace.  You
                    ## should have access to the programs in “leaps” now.
```