

# MATH-650 Assignment 9

Saket Choudhary (USCID: 2170058637) ([skchoudh@usc.edu](mailto:skchoudh@usc.edu))

11/04/2015

## Chapter 12: 14

```
require(leaps)
```

```
## Loading required package: leaps
```

```
data <- read.csv('case1102.csv')
data$logY = log(data$Brain/data$Liver)
Y <- data$logY
X <- data[,c('Days', 'Sex', 'Weight', 'Loss', 'Tumor')]
```

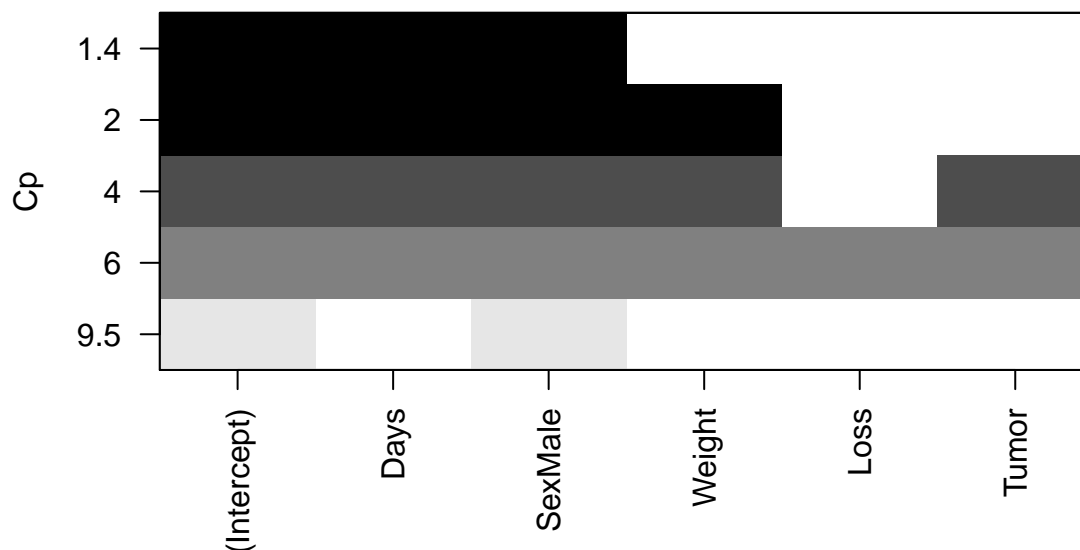
We use the leaps package to perform subset selection.

```
rsubsets <- regsubsets(logY ~ Days+Sex+Weight+Loss+Tumor, data=data)
s <- summary(rsubsets, matrix.logical=TRUE)
s$cp
```

```
## [1] 9.457598 1.430200 2.006538 4.000835 6.000000
```

Part (a):  $C_p$

```
plot(rsubsets, scale='Cp')
```



The way to interpret this plot is to look at first the smallest  $C_p$  values, which happens to be around 1.4 and see the black dots which in this case are given by Days, SexMale. So if we were to choose the covariates based only on  $C_p$  values, we select: Days and Sex. Here  $p = 5$  and in principle any model with  $C_p < p$  is better than the full model, so we can also select these:

- Days, Sex:  $C_p = 1.43$
- Days, Sex, Weight:  $C_p = 2.006$
- Days, Sex, Weight, Tumor:  $C_p = 4.00008$

### Part (b): Forward Selection

```
rsubsets <- regsubsets(logY ~ Days+Sex+Weight+Loss+Tumor,
                      data=data,
                      method='forward')
sforward <- summary(rsubsets, matrix.logical=TRUE)
sforward
```

```
## Subset selection object
## Call: regsubsets.formula(logY ~ Days + Sex + Weight + Loss + Tumor,
##      data = data, method = "forward")
## 5 Variables (and intercept)
##      Forced in Forced out
## Days      FALSE      FALSE
## SexMale    FALSE      FALSE
## Weight     FALSE      FALSE
## Loss       FALSE      FALSE
## Tumor      FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: forward
##      Days SexMale Weight Loss Tumor
## 1 ( 1 ) FALSE      TRUE  FALSE FALSE FALSE
## 2 ( 1 ) TRUE      TRUE  FALSE FALSE FALSE
## 3 ( 1 ) TRUE      TRUE   TRUE  FALSE FALSE
## 4 ( 1 ) TRUE      TRUE   TRUE  FALSE  TRUE
## 5 ( 1 ) TRUE      TRUE   TRUE   TRUE  TRUE
```

### Part (c): Backward Selection

```
rsubsets <- regsubsets(logY ~ Days+Sex+Weight+Loss+Tumor,
                      data=data,
                      method='backward')
sbackward <- summary(rsubsets, matrix.logical=TRUE)
sbackward
```

```
## Subset selection object
## Call: regsubsets.formula(logY ~ Days + Sex + Weight + Loss + Tumor,
##      data = data, method = "backward")
## 5 Variables (and intercept)
##      Forced in Forced out
```

```
## Days          FALSE      FALSE
## SexMale       FALSE      FALSE
## Weight        FALSE      FALSE
## Loss          FALSE      FALSE
## Tumor         FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: backward
##           Days SexMale Weight  Loss Tumor
## 1  ( 1 ) FALSE      TRUE  FALSE FALSE FALSE
## 2  ( 1 )  TRUE      TRUE  FALSE FALSE FALSE
## 3  ( 1 )  TRUE      TRUE   TRUE  FALSE FALSE
## 4  ( 1 )  TRUE      TRUE   TRUE  FALSE  TRUE
## 5  ( 1 )  TRUE      TRUE   TRUE   TRUE  TRUE
```

#### Part(d): Stepwise Regression

```
rsubsets <- regsubsets(logY ~ Days+Sex+Weight+Loss+Tumor,
                      data=data,
                      method="seqrep")
sboth <- summary(rsubsets, matrix.logical=TRUE)
sboth
```

```
## Subset selection object
## Call: regsubsets.formula(logY ~ Days + Sex + Weight + Loss + Tumor,
##      data = data, method = "seqrep")
## 5 Variables (and intercept)
##           Forced in Forced out
## Days          FALSE      FALSE
## SexMale       FALSE      FALSE
## Weight        FALSE      FALSE
## Loss          FALSE      FALSE
## Tumor         FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: 'sequential replacement'
##           Days SexMale Weight  Loss Tumor
## 1  ( 1 ) FALSE      TRUE  FALSE FALSE FALSE
## 2  ( 1 )  TRUE      TRUE  FALSE FALSE FALSE
## 3  ( 1 )  TRUE      TRUE   TRUE  FALSE FALSE
## 4  ( 1 )  TRUE      TRUE   TRUE  FALSE  TRUE
## 5  ( 1 )  TRUE      TRUE   TRUE   TRUE  TRUE
```

#### Conclusion

From the above, we conclude that the variable selection in this case gives us the same set for all four methods.