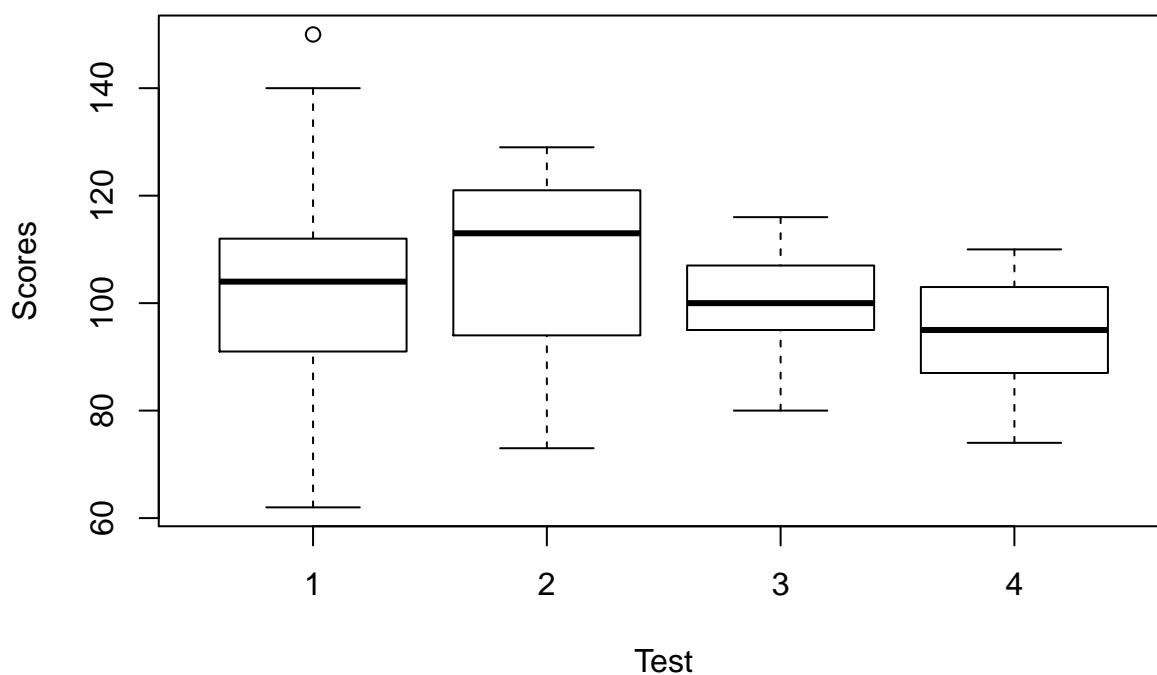


STAT GR5205 – Section 005 HW 7

Bo Rong br2498

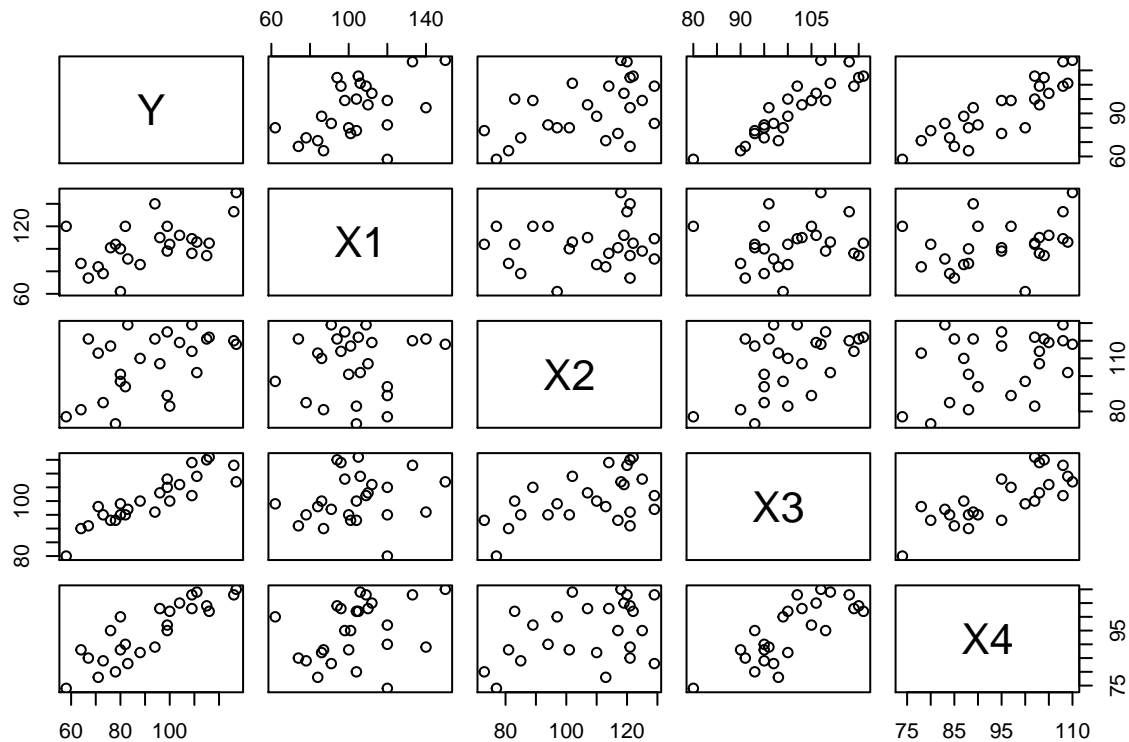
Dec. 2nd, 2016

```
#1.  
#(a)  
filename <- "~/Downloads/job_proficiency.txt"  
JP<-read.table(file=filename, header=T)  
X<-c(JP$X1,JP$X2,JP$X3,JP$X4)  
Test <-rep(1:4,rep(25,4))  
boxplot(X~Test, xlab="Test", ylab="Scores")
```



#All four tests have average score around 100, and test 1 has one outlier.

```
##(b)  
pairs(JP)
```



#Y is positively associated with all four test scores(X1,X2,X3,X4). The test scores X1,X2,X3 and X4 are positively associated with each other,there is no serious multicollinearity problems evident.

```
#(c)
mrf <- lm(Y ~ X1 + X2 + X3 + X4, data=JP)
summary(mrf)

##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4, data = JP)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.9779 -3.4506  0.0941  2.4749  5.9959
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -124.38182    9.94106  -12.512 6.48e-11 ***
## X1           0.29573    0.04397   6.725 1.52e-06 ***
## X2           0.04829    0.05662   0.853  0.40383
## X3           1.30601    0.16409   7.959 1.26e-07 ***
## X4           0.51982    0.13194   3.940  0.00081 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.099 on 20 degrees of freedom
## Multiple R-squared:  0.9629, Adjusted R-squared:  0.9555
## F-statistic: 129.7 on 4 and 20 DF,  p-value: 5.262e-14
```

#All four coefficients are positive. P-value of X2 is 0.40383,so X2 can be dropped.

#2.

#(a)

```
mback<-lm(Y ~ ., data=JP)
step(mback, direction="backward")
```

```
## Start:  AIC=74.95
## Y ~ X1 + X2 + X3 + X4
##
##           Df Sum of Sq    RSS    AIC
## - X2      1      12.22  348.20  73.847
## <none>                    335.98  74.954
## - X4      1     260.74  596.72  87.314
## - X1      1     759.83 1095.81 102.509
## - X3      1    1064.15 1400.13 108.636
##
## Step:  AIC=73.85
## Y ~ X1 + X3 + X4
##
##           Df Sum of Sq    RSS    AIC
## <none>                    348.20  73.847
## - X4      1     258.46  606.66  85.727
## - X1      1     763.12 1111.31 100.861
## - X3      1    1324.39 1672.59 111.081

##
## Call:
## lm(formula = Y ~ X1 + X3 + X4, data = JP)
##
## Coefficients:
## (Intercept)          X1          X3          X4
##   -124.2000     0.2963     1.3570     0.5174
```

#Estimated mean function is $E(Y|X = x) = -124.2 + 0.2963x_1 + 1.357x_3 + 0.5174x_4$.

#(b)

```
mfor<-lm(Y ~ 1, data=JP)
step(mfor, scope=~X1+X2+X3+X4)
```

```
## Start:  AIC=149.3
## Y ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + X3      1     7286.0 1768.0 110.47
## + X4      1     6843.3 2210.7 116.06
## + X1      1     2395.9 6658.1 143.62
## + X2      1     2236.5 6817.5 144.21
## <none>                    9054.0 149.30
##
## Step:  AIC=110.47
```

```
## Y ~ X3
##
##      Df Sum of Sq    RSS    AIC
## + X1   1   1161.4   606.7   85.727
## + X4   1    656.7  1111.3  100.861
## <none>          1768.0  110.469
## + X2   1     12.2  1755.8  112.295
## - X3   1   7286.0  9054.0  149.302
##
## Step: AIC=85.73
## Y ~ X3 + X1
##
##      Df Sum of Sq    RSS    AIC
## + X4   1    258.5   348.2   73.847
## <none>          606.7   85.727
## + X2   1      9.9   596.7   87.314
## - X1   1   1161.4  1768.0  110.469
## - X3   1   6051.5  6658.1  143.618
##
## Step: AIC=73.85
## Y ~ X3 + X1 + X4
##
##      Df Sum of Sq    RSS    AIC
## <none>          348.20   73.847
## + X2   1    12.22  335.98   74.954
## - X4   1   258.46   606.66   85.727
## - X1   1   763.12  1111.31  100.861
## - X3   1  1324.39  1672.59  111.081

##
## Call:
## lm(formula = Y ~ X3 + X1 + X4, data = JP)
##
## Coefficients:
## (Intercept)          X3          X1          X4
##   -124.2000     1.3570     0.2963     0.5174
```

#E(Y|X = x) = -124.2 + 0.2963x1 + 1.357x3 + 0.5174x4. The same as part(a).

```
##(c)
step(mback, direction="backward", k=log(25))
```

```
## Start: AIC=81.05
## Y ~ X1 + X2 + X3 + X4
##
##      Df Sum of Sq    RSS    AIC
## - X2   1    12.22   348.20   78.723
## <none>          335.98   81.049
## - X4   1   260.74   596.72   92.190
## - X1   1   759.83  1095.81  107.385
## - X3   1  1064.15  1400.13  113.512
##
## Step: AIC=78.72
```

```
## Y ~ X1 + X3 + X4
##
##           Df Sum of Sq    RSS    AIC
## <none>                348.20  78.723
## - X4      1      258.46  606.66  89.384
## - X1      1      763.12 1111.31 104.517
## - X3      1     1324.39 1672.59 114.738

##
## Call:
## lm(formula = Y ~ X1 + X3 + X4, data = JP)
##
## Coefficients:
## (Intercept)          X1          X3          X4
##   -124.2000      0.2963      1.3570      0.5174
```

```
step(mfor, scope=~X1+X2+X3+X4, k=log(25))
```

```
## Start:  AIC=150.52
## Y ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + X3      1      7286.0 1768.0 112.91
## + X4      1      6843.3 2210.7 118.49
## + X1      1      2395.9 6658.1 146.06
## + X2      1      2236.5 6817.5 146.65
## <none>                9054.0 150.52
##
## Step:  AIC=112.91
## Y ~ X3
##
##           Df Sum of Sq    RSS    AIC
## + X1      1      1161.4  606.7  89.384
## + X4      1        656.7 1111.3 104.517
## <none>                1768.0 112.906
## + X2      1         12.2 1755.8 115.952
## - X3      1      7286.0 9054.0 150.521
##
## Step:  AIC=89.38
## Y ~ X3 + X1
##
##           Df Sum of Sq    RSS    AIC
## + X4      1       258.5  348.2  78.723
## <none>                606.7  89.384
## + X2      1         9.9  596.7  92.190
## - X1      1      1161.4 1768.0 112.906
## - X3      1     6051.5 6658.1 146.056
##
## Step:  AIC=78.72
## Y ~ X3 + X1 + X4
##
##           Df Sum of Sq    RSS    AIC
## <none>                348.20  78.723
```

```
## + X2      1      12.22  335.98  81.049
## - X4      1      258.46  606.66  89.384
## - X1      1      763.12 1111.31 104.517
## - X3      1     1324.39 1672.59 114.738
```

```
##
## Call:
## lm(formula = Y ~ X3 + X1 + X4, data = JP)
##
## Coefficients:
## (Intercept)          X3          X1          X4
##   -124.2000      1.3570      0.2963      0.5174
```

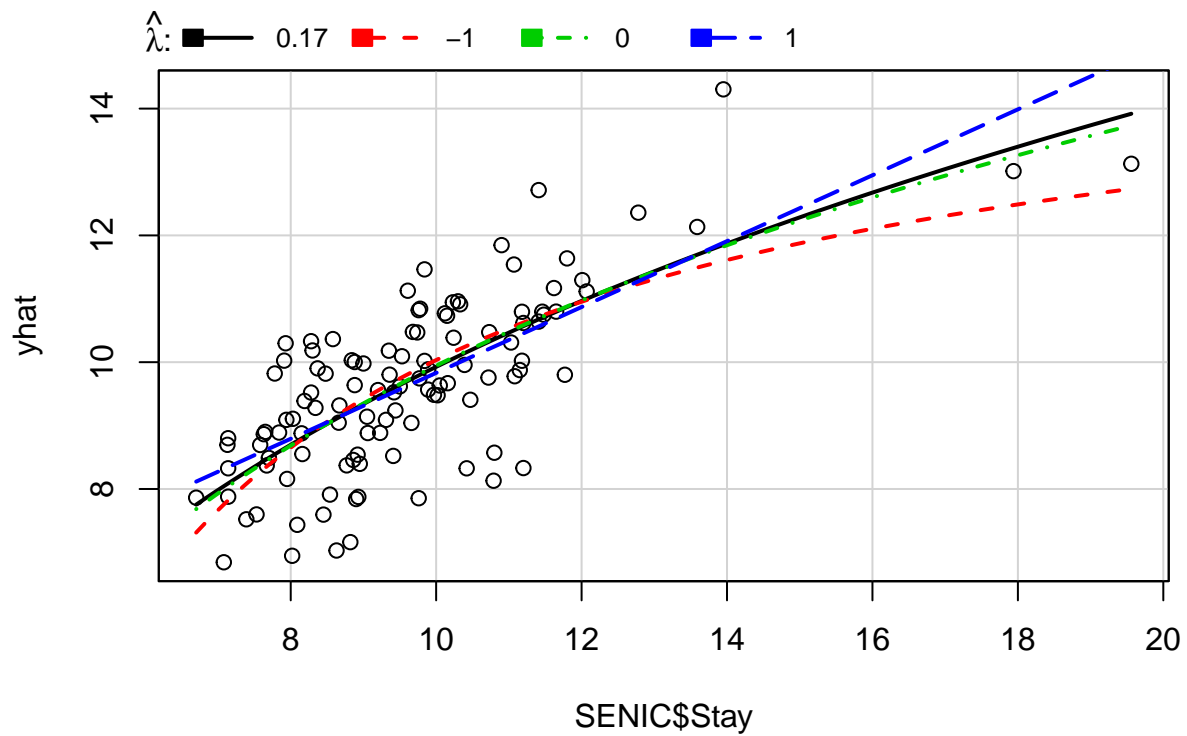
#Still the same mean function.

```
#3.
#(a)
filename <- "~/Downloads/SENIC.txt"
SENIC<-read.table(file=filename, header=T)
library(car)
transformation<-powerTransform(cbind(Age,Risk,Cult,Xray,Beds,Cen,Nurses,AFS)~1,data=SENIC)
summary(transformation)
```

```
## bcPower Transformations to Multinormality
##
##      Est.Power Std.Err. Wald Lower Bound Wald Upper Bound
## Age      1.6803   0.7892      0.1333      3.2272
## Risk      0.8007   0.2009      0.4069      1.1945
## Cult      0.1428   0.0971     -0.0476      0.3332
## Xray      0.8184   0.3179      0.1954      1.4414
## Beds      0.1096   0.0597     -0.0073      0.2265
## Cen       0.1968   0.0631      0.0731      0.3205
## Nurses    0.1450   0.0636      0.0203      0.2698
## AFS       0.9753   0.1584      0.6649      1.2857
##
## Likelihood ratio tests about transformation parameters
##
##      LRT df      pval
## LR test, lambda = (0 0 0 0 0 0 0 0)      84.39533  8 6.328271e-15
## LR test, lambda = (1 1 1 1 1 1 1 1)      339.79456  8 0.000000e+00
## LR test, lambda = (1 1 0 1 0 0.2 0.15 1)  12.79795  8 1.189931e-01
```

#Box-Cox method chooses log-transformations (lambda = 0) for Cult and Beds, but lambda = .20 for Cen and lambda = .15 for Nurses.

```
X1 <- SENIC$Age
X2 <- SENIC$Risk
X3 <- log(SENIC$Cult)
X4 <- SENIC$Xray
X5 <- log(SENIC$Beds)
X6 <- log(SENIC$Cen)
X7 <- log(SENIC$Nurses)
X8 <- SENIC$AFS
m1<-lm(SENIC$Stay ~ X1+X2+X3+X4+X5+X6+X7+X8)
inverseResponsePlot(m1)
```

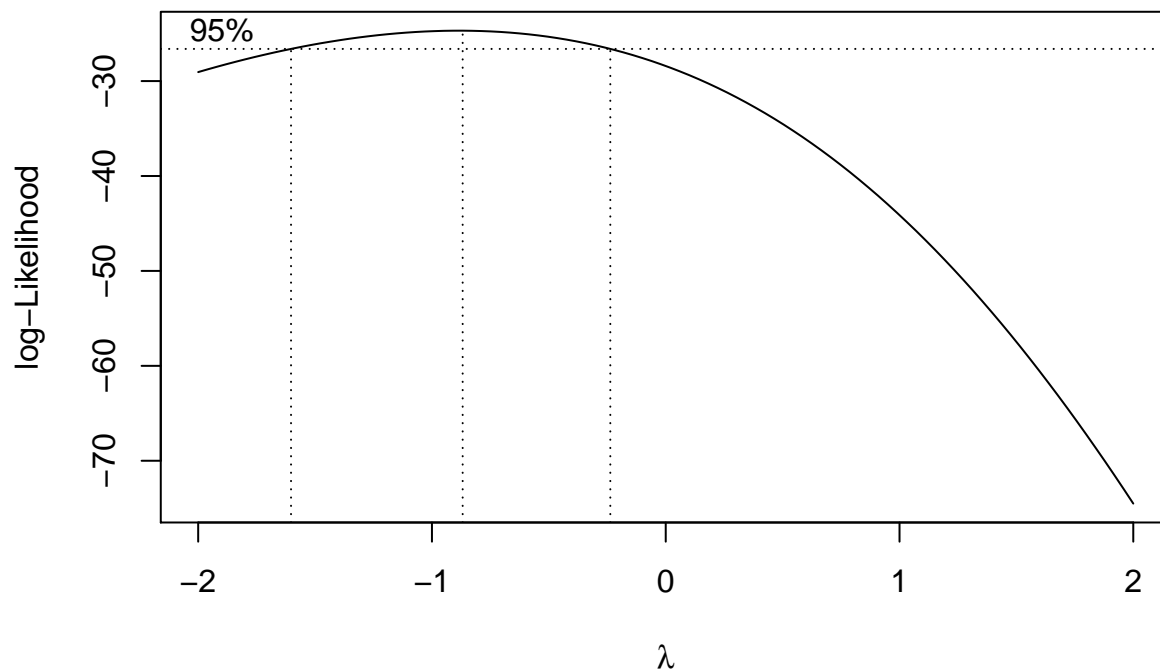


```
##      lambda      RSS
## 1  0.1711874  99.34549
## 2 -1.0000000 103.45688
## 3  0.0000000  99.45095
## 4  1.0000000 102.14551
```

```
unlist(invTranEstimate(SENIC$Stay, fitted(m1)))
```

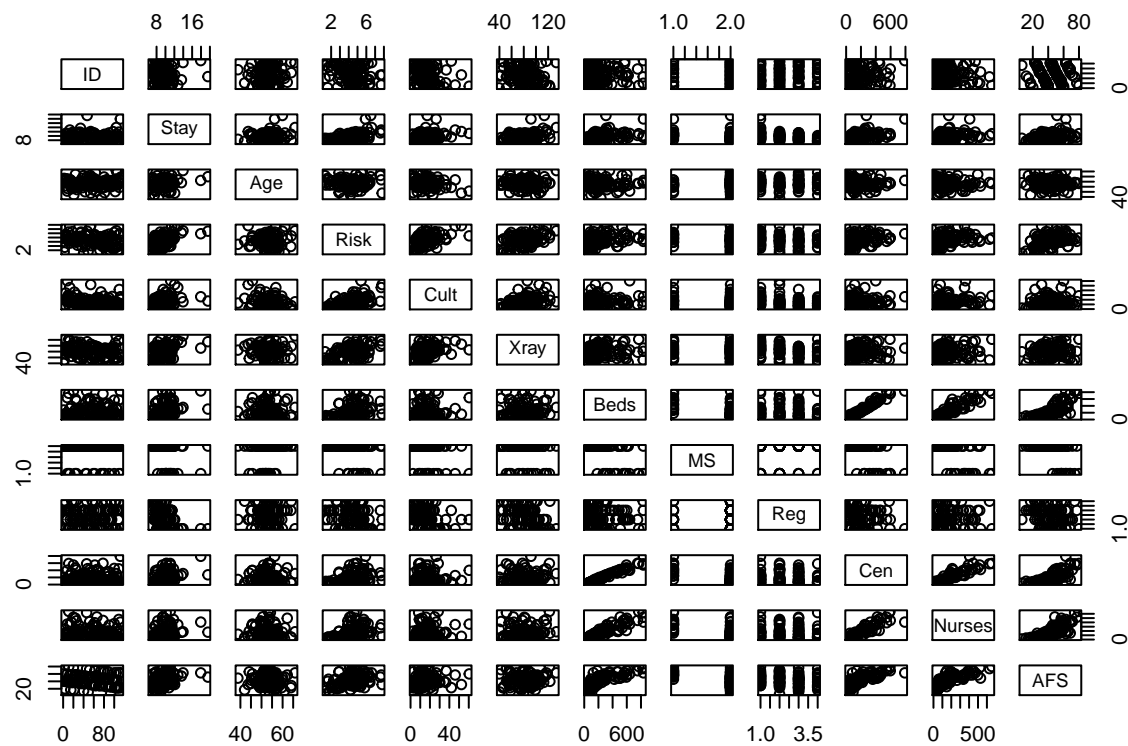
```
##      lambda  lowerCI  upperCI
## 0.1711874 -0.550554  0.8228023
```

```
library(MASS)
boxcox(m1)
```



```
Y <- 1 / SENIC$Stay
```

```
#(b)
pairs(SENIC)
```



```
#The strongest correlations are X5 and X6.
```



```
#(c)
mforward <-lm(Y~1)
step(mforward, scope=~X1+X2+X3+X4+X5+X6+X7+X8)
```

```
## Start:  AIC=-908.53
## Y ~ 1
##
##      Df Sum of Sq      RSS      AIC
## + X2   1 0.0106374 0.025141 -946.40
## + X6   1 0.0098106 0.025967 -942.75
## + X5   1 0.0081831 0.027595 -935.88
## + X7   1 0.0074095 0.028369 -932.75
## + X8   1 0.0053558 0.030422 -924.86
## + X4   1 0.0050737 0.030704 -923.81
## + X3   1 0.0040332 0.031745 -920.05
## + X1   1 0.0008214 0.034957 -909.16
## <none>          0.035778 -908.53
##
## Step:  AIC=-946.4
## Y ~ X2
##
##      Df Sum of Sq      RSS      AIC
## + X6   1 0.0033629 0.021778 -960.63
## + X5   1 0.0024493 0.022691 -955.99
## + X7   1 0.0012631 0.023877 -950.23
## + X8   1 0.0011306 0.024010 -949.60
## + X1   1 0.0008149 0.024326 -948.13
## + X4   1 0.0007536 0.024387 -947.84
## <none>          0.025141 -946.40
## + X3   1 0.0000054 0.025135 -944.43
## - X2   1 0.0106374 0.035778 -908.53
##
## Step:  AIC=-960.63
## Y ~ X2 + X6
##
##      Df Sum of Sq      RSS      AIC
## + X1   1 0.0014006 0.020377 -966.14
## + X4   1 0.0013231 0.020455 -965.71
## + X7   1 0.0011907 0.020587 -964.98
## + X5   1 0.0010367 0.020741 -964.14
## <none>          0.021778 -960.63
## + X8   1 0.0003322 0.021445 -960.37
## + X3   1 0.0000005 0.021777 -958.63
## - X6   1 0.0033629 0.025141 -946.40
## - X2   1 0.0041897 0.025967 -942.75
##
## Step:  AIC=-966.14
## Y ~ X2 + X6 + X1
##
##      Df Sum of Sq      RSS      AIC
## + X4   1 0.0014489 0.018928 -972.48
## + X5   1 0.0011745 0.019203 -970.85
## + X7   1 0.0010162 0.019361 -969.92
```

```

## + X8      1 0.0005062 0.019871 -966.99
## <none>          0.020377 -966.14
## + X3      1 0.0001465 0.020231 -964.96
## - X1      1 0.0014006 0.021778 -960.63
## - X2      1 0.0038429 0.024220 -948.62
## - X6      1 0.0039486 0.024326 -948.13
##
## Step: AIC=-972.48
## Y ~ X2 + X6 + X1 + X4
##
##           Df Sum of Sq      RSS      AIC
## + X7      1 0.0011222 0.017806 -977.38
## + X5      1 0.0010809 0.017847 -977.12
## + X8      1 0.0005783 0.018350 -973.98
## <none>          0.018928 -972.48
## + X3      1 0.0000008 0.018927 -970.48
## - X2      1 0.0013755 0.020304 -966.55
## - X4      1 0.0014489 0.020377 -966.14
## - X1      1 0.0015264 0.020455 -965.71
## - X6      1 0.0046091 0.023537 -949.85
##
## Step: AIC=-977.38
## Y ~ X2 + X6 + X1 + X4 + X7
##
##           Df Sum of Sq      RSS      AIC
## + X5      1 0.0005239 0.017282 -978.76
## <none>          0.017806 -977.38
## + X8      1 0.0000987 0.017707 -976.01
## + X3      1 0.0000734 0.017733 -975.85
## - X7      1 0.0011222 0.018928 -972.48
## - X1      1 0.0013399 0.019146 -971.19
## - X4      1 0.0015549 0.019361 -969.92
## - X2      1 0.0020531 0.019859 -967.05
## - X6      1 0.0037727 0.021579 -957.67
##
## Step: AIC=-978.76
## Y ~ X2 + X6 + X1 + X4 + X7 + X5
##
##           Df Sum of Sq      RSS      AIC
## <none>          0.017282 -978.76
## - X5      1 0.00052393 0.017806 -977.38
## - X7      1 0.00056525 0.017847 -977.12
## + X8      1 0.00003308 0.017249 -976.98
## + X3      1 0.00003060 0.017251 -976.96
## - X4      1 0.00145353 0.018736 -971.63
## - X1      1 0.00146953 0.018752 -971.54
## - X2      1 0.00179886 0.019081 -969.57
## - X6      1 0.00243527 0.019717 -965.86

##
## Call:
## lm(formula = Y ~ X2 + X6 + X1 + X4 + X7 + X5)
##
## Coefficients:

```

```
## (Intercept)          X2          X6          X1          X4
##  0.2127046   -0.0040211   -0.0297851   -0.0008262   -0.0002117
##          X7          X5
##  0.0074390    0.0160117
```

*#Forward selection chooses the mean function :0.2127046 -0.0008262X1 - 0.0040211X2
#- 0.0002117X4 + 0.0160117X5 - 0.0297851X6 + 0.0074390X7.*

```
##(d)
mbackward<-lm(Y~X1+X2+X3+X4+X5+X6+X7+X8)
step(mbackward, direction="backward")
```

```
## Start:  AIC=-975.23
## Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8
##
##      Df Sum of Sq    RSS    AIC
## - X3   1 0.00003881 0.017249 -976.98
## - X8   1 0.00004128 0.017251 -976.96
## <none>          0.017210 -975.23
## - X5   1 0.00040764 0.017618 -974.58
## - X7   1 0.00042006 0.017630 -974.50
## - X2   1 0.00116721 0.018377 -969.81
## - X4   1 0.00121736 0.018427 -969.51
## - X1   1 0.00151398 0.018724 -967.70
## - X6   1 0.00237245 0.019583 -962.64
##
## Step:  AIC=-976.98
## Y ~ X1 + X2 + X4 + X5 + X6 + X7 + X8
##
##      Df Sum of Sq    RSS    AIC
## - X8   1 0.00003308 0.017282 -978.76
## <none>          0.017249 -976.98
## - X7   1 0.00038641 0.017635 -976.47
## - X5   1 0.00045827 0.017707 -976.01
## - X4   1 0.00146698 0.018716 -969.75
## - X1   1 0.00150261 0.018752 -969.54
## - X2   1 0.00173970 0.018989 -968.12
## - X6   1 0.00238438 0.019633 -964.34
##
## Step:  AIC=-978.76
## Y ~ X1 + X2 + X4 + X5 + X6 + X7
##
##      Df Sum of Sq    RSS    AIC
## <none>          0.017282 -978.76
## - X5   1 0.00052393 0.017806 -977.38
## - X7   1 0.00056525 0.017847 -977.12
## - X4   1 0.00145353 0.018736 -971.63
## - X1   1 0.00146953 0.018752 -971.54
## - X2   1 0.00179886 0.019081 -969.57
## - X6   1 0.00243527 0.019717 -965.86
##
##
## Call:
```

```
## lm(formula = Y ~ X1 + X2 + X4 + X5 + X6 + X7)
##
## Coefficients:
## (Intercept)          X1          X2          X4          X5
##  0.2127046   -0.0008262   -0.0040211   -0.0002117    0.0160117
##          X6          X7
##  -0.0297851    0.0074390

#Backward selection also has the same mean function:0.2127046 -0.0008262X1 - 0.0040211X2
#- 0.0002117X4 + 0.0160117X5 - 0.0297851X6 + 0.0074390X7.
```

```
##(e)
step(mbackward, direction="backward", k=log(length(Y)))
```

```
## Start:  AIC=-950.68
## Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8
##
##      Df Sum of Sq  RSS    AIC
## - X3   1 0.00003881 0.017249 -955.16
## - X8   1 0.00004128 0.017251 -955.14
## - X5   1 0.00040764 0.017618 -952.77
## - X7   1 0.00042006 0.017630 -952.69
## <none>          0.017210 -950.68
## - X2   1 0.00116721 0.018377 -948.00
## - X4   1 0.00121736 0.018427 -947.69
## - X1   1 0.00151398 0.018724 -945.88
## - X6   1 0.00237245 0.019583 -940.82
##
## Step:  AIC=-955.16
## Y ~ X1 + X2 + X4 + X5 + X6 + X7 + X8
##
##      Df Sum of Sq  RSS    AIC
## - X8   1 0.00003308 0.017282 -959.67
## - X7   1 0.00038641 0.017635 -957.38
## - X5   1 0.00045827 0.017707 -956.92
## <none>          0.017249 -955.16
## - X4   1 0.00146698 0.018716 -950.66
## - X1   1 0.00150261 0.018752 -950.45
## - X2   1 0.00173970 0.018989 -949.03
## - X6   1 0.00238438 0.019633 -945.25
##
## Step:  AIC=-959.67
## Y ~ X1 + X2 + X4 + X5 + X6 + X7
##
##      Df Sum of Sq  RSS    AIC
## - X5   1 0.00052393 0.017806 -961.02
## - X7   1 0.00056525 0.017847 -960.76
## <none>          0.017282 -959.67
## - X4   1 0.00145353 0.018736 -955.27
## - X1   1 0.00146953 0.018752 -955.17
## - X2   1 0.00179886 0.019081 -953.21
## - X6   1 0.00243527 0.019717 -949.50
##
```

```
## Step: AIC=-961.02
## Y ~ X1 + X2 + X4 + X6 + X7
##
##      Df Sum of Sq      RSS      AIC
## <none>            0.017806 -961.02
## - X7      1 0.0011222 0.018928 -958.84
## - X1      1 0.0013399 0.019146 -957.55
## - X4      1 0.0015549 0.019361 -956.29
## - X2      1 0.0020531 0.019859 -953.42
## - X6      1 0.0037727 0.021579 -944.03

##
## Call:
## lm(formula = Y ~ X1 + X2 + X4 + X6 + X7)
##
## Coefficients:
## (Intercept)          X1          X2          X4          X6
##  0.2247364   -0.0007864   -0.0042680   -0.0002187   -0.0176324
##          X7
##  0.0098629

#By the Bayesian criterion, the mean function :0.2247364 - 0.0007864X1 - 0.0042680X2
#- 0.0002187X4 - 0.0176324X6 + 0.0098629X7, which means also drop X5.
m.final <- step(mbackward, direction="backward", k=log(length(Y)))
```

```
## Start: AIC=-950.68
## Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8
##
##      Df Sum of Sq      RSS      AIC
## - X3      1 0.00003881 0.017249 -955.16
## - X8      1 0.00004128 0.017251 -955.14
## - X5      1 0.00040764 0.017618 -952.77
## - X7      1 0.00042006 0.017630 -952.69
## <none>            0.017210 -950.68
## - X2      1 0.00116721 0.018377 -948.00
## - X4      1 0.00121736 0.018427 -947.69
## - X1      1 0.00151398 0.018724 -945.88
## - X6      1 0.00237245 0.019583 -940.82
##
## Step: AIC=-955.16
## Y ~ X1 + X2 + X4 + X5 + X6 + X7 + X8
##
##      Df Sum of Sq      RSS      AIC
## - X8      1 0.00003308 0.017282 -959.67
## - X7      1 0.00038641 0.017635 -957.38
## - X5      1 0.00045827 0.017707 -956.92
## <none>            0.017249 -955.16
## - X4      1 0.00146698 0.018716 -950.66
## - X1      1 0.00150261 0.018752 -950.45
## - X2      1 0.00173970 0.018989 -949.03
## - X6      1 0.00238438 0.019633 -945.25
##
## Step: AIC=-959.67
```

```
## Y ~ X1 + X2 + X4 + X5 + X6 + X7
##
##          Df Sum of Sq      RSS      AIC
## - X5      1 0.00052393 0.017806 -961.02
## - X7      1 0.00056525 0.017847 -960.76
## <none>                                0.017282 -959.67
## - X4      1 0.00145353 0.018736 -955.27
## - X1      1 0.00146953 0.018752 -955.17
## - X2      1 0.00179886 0.019081 -953.21
## - X6      1 0.00243527 0.019717 -949.50
##
## Step: AIC=-961.02
## Y ~ X1 + X2 + X4 + X6 + X7
##
##          Df Sum of Sq      RSS      AIC
## <none>                                0.017806 -961.02
## - X7      1 0.0011222 0.018928 -958.84
## - X1      1 0.0013399 0.019146 -957.55
## - X4      1 0.0015549 0.019361 -956.29
## - X2      1 0.0020531 0.019859 -953.42
## - X6      1 0.0037727 0.021579 -944.03
```

```
summary(m.final)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X4 + X6 + X7)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0294128 -0.0094486 -0.0001517  0.0096839  0.0258416
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.247e-01  1.825e-02  12.312 < 2e-16 ***
## X1           -7.864e-04  2.771e-04  -2.838 0.005439 **
## X2           -4.268e-03  1.215e-03  -3.513 0.000651 ***
## X4           -2.187e-04  7.153e-05  -3.057 0.002826 **
## X6           -1.763e-02  3.703e-03  -4.761 6.06e-06 ***
## X7            9.863e-03  3.798e-03   2.597 0.010732 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0129 on 107 degrees of freedom
## Multiple R-squared:  0.5023, Adjusted R-squared:  0.4791
## F-statistic: 21.6 on 5 and 107 DF, p-value: 6.823e-15
```

```
#The average length of stay is an increasing function of X1(Age), X2(Risk), X4(Xray),
#and X6(Cen), but a decreasing function of X7(Nurses).
```