```
# R Code for HW7                              Sta108, Fall 2007, Utts


### Problem 9.25
Data=read.table("~/Documents/School/Sta108utts/APPENC01.txt")
names(Data)=c("ID","Stay","Age","Risk","Cultur","Xray","Beds",
              "MedSchool","Region","Census","Nurse","Facility")


#(b)
#Dataset used above (from file "APPENC01.txt") needs to be edited.
#1.Consider observations numbered: 57-113 in your dataset
dim(Data)
Data=Data[57:113,]

#2.Remove two predictors not needed for the analysis:
#these are categorical variables: "MedSchool","Region" (Columns 8,9)
#3.remove ID column not needed for the analysis (Column 1)
Data=Data[,-c(1,8,9)]

#4.transform the response Y to: log10(Y)
Data$Stay=log10(Data$Stay)

#Now, your dataset is clean to be used in the problem
Data

DataY=Data[,1]  #separate response Y
DataX=Data[,-1] #separate predictors, X variables

pairs(DataX)
cor(DataX)

#(c) (NOTE THAT Y IS NOW THE LOG10 TRANSFORMED VERSION OF THE ORIGINAL Y, FROM #4 ABOVE)
library(leaps)
leaps(x=DataX, y=DataY,
      names=c("Age","Risk","Cultur","Xray","Beds","Census","Nurse","Facility"),
      method="Cp")
#Three best models are chosen by 3 lowest Cp criterions

#To automatically print models in the increasing order of Cp criterion:
ModelSel = leaps(x=DataX, y=DataY,
      names=c("Age","Risk","Cultur","Xray","Beds","Census","Nurse","Facility"),
      method="Cp")
ModelSel$which[ order( ModelSel$Cp ), ]

#To print Cp criterion in increasing order
sort( ModelSel$Cp )

#To plot Cp against p, and add reference line: Cp=p
plot( ModelSel$size,  ModelSel$Cp, pch=19)
abline(0,1)

#Fit the best chosen model
Fit = lm(Stay ~ Age + Xray + Census, data=Data)
Fit

#Residual plot
plot(Fit$fitted.values, Fit$residuals, main="Residuals vs. Fitted Values", xlab="Fitted
Values",ylab="Residuals",pch=19)
abline(h=0)
```