module\_9\_assignment\_rmd

library(leaps)  
### Problem 1  
df <- MPV::table.b4  
n <- 14; k <- 5  
set.seed(1); chosen\_rows <- sort(sample(seq(1, nrow(df)), n))  
set.seed(1); chosen\_cols <- sort(sample(seq(1, ncol(df)), k))  
df1 <- df[chosen\_rows, c(1, chosen\_cols + 1)]  
# Reference: http://www.sthda.com/english/articles/37-model-selection-essentials-in-r/155-best-subsets-regression-essentials-in-r/  
best1 <- regsubsets(x = y~., data = df1, nvmax = 5)  
res.sum <- summary(best1)  
p.m <- 2:6  
aic <- n \* log(res.sum$rss / n) + 2 \* p.m  
data.frame(  
 Adj.R2 = which.max(res.sum$adjr2),  
 CP = which.min(res.sum$cp),  
 BIC = which.min(res.sum$bic),  
 AIC = which.min(aic)  
)  
  
### Problem 2  
df <- MPV::table.b11  
n <- 14  
set.seed(2); chosen\_rows <- sort(sample(seq(1,nrow(df)), n))  
df2 <- df[chosen\_rows,]  
  
# Reference: https://stats.idre.ucla.edu/r/modules/coding-for-categorical-variables-in-regression-models/  
best2 <- regsubsets(x = Quality ~ Clarity + Aroma + Body + Flavor + Oakiness +  
 I(Region ==1) + I(Region ==2) + I(Region ==3), data = df2, nvmax = 8)  
res.sum <- summary(best2)  
p.m <- 2:8  
aic <- n \* log(res.sum$rss / n) + 2 \* p.m  
data.frame(  
 Adj.R2 = which.max(res.sum$adjr2),  
 CP = which.min(res.sum$cp),  
 BIC = which.min(res.sum$bic),  
 AIC = which.min(aic)  
)