Samantha L. Misurda (smisurda)

HW3

Problem 1

3)

Approximately 80% of the variance is explained by 2 components.

```r
{r}
######
# Problem 1.3
#####

# Iterate over components and stop when variance is 80%
for(i in 1:length(returns.pca$sdev)) {
  total.variance <- sum(returns.pca$sdev[1:i]^2/sum(returns.pca$sdev^2))
  if(total.variance > .80) {
    cat(sprintf("A total variance of %f > .80 uses %d components", total.variance, i))
    break
  }
}
```

```
A total variance of 0.804254 > .80 uses 2 components
```

4)

The magnitude of the estimated reconstruction error if we only retain top two of the PCA components is 0.1957456

Problem 2

A.

The three features with the highest correlation with body fat percentage are:

```r
# Print out top 3
feature.correlations[1:3,]
```

```
##      Feature Correlation
## 6 Abdomen    0.8134323
## 5   Chest    0.7026203
## 7     Hip    0.6252009
```

B.

The three best subsets of sizes 1, 2, and 3 are shown below:

```
## Subset selection object
## Call: regsubsets.formula(fatpctg ~ ., data = bmi, nbest = 1, nvmax = 3,
##      method = "exhaustive")
## 13 Variables  (and intercept)
##          Forced in Forced out
## Age           FALSE       FALSE
## Weight        FALSE       FALSE
## Height        FALSE       FALSE
## Neck          FALSE       FALSE
## Chest         FALSE       FALSE
## Abdomen       FALSE       FALSE
## Hip           FALSE       FALSE
## Thigh         FALSE       FALSE
## Knee          FALSE       FALSE
## Ankle         FALSE       FALSE
## Biceps        FALSE       FALSE
## Forearm       FALSE       FALSE
## Wrist         FALSE       FALSE
## 1 subsets of each size up to 3
## Selection Algorithm: exhaustive
##          Age Weight Height Neck Chest Abdomen Hip Thigh Knee Ankle Biceps
## 1  ( 1 ) " " " "    " "    " "  " "   "*"     " " " "   " "  " "   " "
## 2  ( 1 ) " " "*"    " "    " "  " "   "*"     " " " "   " "  " "   " "
## 3  ( 1 ) " " "*"    " "    " "  " "   "*"     " " " "   " "  " "   " "
##          Forearm Wrist
## 1  ( 1 ) " "     " "
## 2  ( 1 ) " "     " "
## 3  ( 1 ) " "     "*"
```

N = 1: Abdomen

N = 2: Weight, Abdomen

N = 3: Weight, Abdomen, Wrist

C.

Stepwise regression produced its best model for fat percentage by using the age, weight, neck, abdomen, hip, thigh, forearm, and wrist variables.

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## fatpctg ~ Age + Weight + Height + Neck + Chest + Abdomen + Hip +
##     Thigh + Knee + Ankle + Biceps + Forearm + Wrist
##
## Final Model:
## fatpctg ~ Age + Weight + Neck + Abdomen + Hip + Thigh + Forearm +
##     Wrist
##
##
##       Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                            238    4452.816 751.7095
## 2  - Knee   1  3.460083       239    4456.276 749.9052
## 3  - Chest  1  4.607796       240    4460.884 748.1656
## 4  - Ankle  1  8.075864       241    4468.960 746.6214
## 5 - Biceps  1 11.665862       242    4480.626 745.2784
## 6 - Height  1 15.176342       243    4495.802 744.1305
```