Stat 482, Homework #13 Solutions

X_1 = triceps, X_2 = thigh, X_3 = midarm, Y = body fat

1. $b_1^* = 4.264$, $b_2^* = -2.929$, $b_3^* = -1.561$

Standardized regression coefficients are based on input variables having the same scale of measurement, so effect sizes are computed in common units.

It is estimated that triceps measurement has the strongest (partial) effect on body fat measurement.

2. see attached for plot

Shrinkage estimation features a weighted average of the candidate estimators (LSE if an input is included, zero if an input is excluded), using information from all models under consideration.

3. The goal of principal components analysis is to explain the relationships among a set of variables using a reduced set formed through linear combinations.

$\hat{\beta}_1^{(PC)} = 0.4156$, $\hat{\beta}_2^{(PC)} = 0.5042$, $\hat{\beta}_3^{(PC)} = -0.0894$

Body Fat Example: Standardized Regression, Model Selection, Ridge Regression, and Principal Components Regression

Data from Table 7.1

A body fat measurement is made for each of 20 healthy females through a cumbersome and expensive procedure requiring the immersion of the person in water. The predictor variables triceps skinfold thickness (x1), thigh circumference (x2) and midarm circumference (x3) are easier to measure. The study aims to determine the relationship between body fat (y) and the varibles x1,x2,x3.

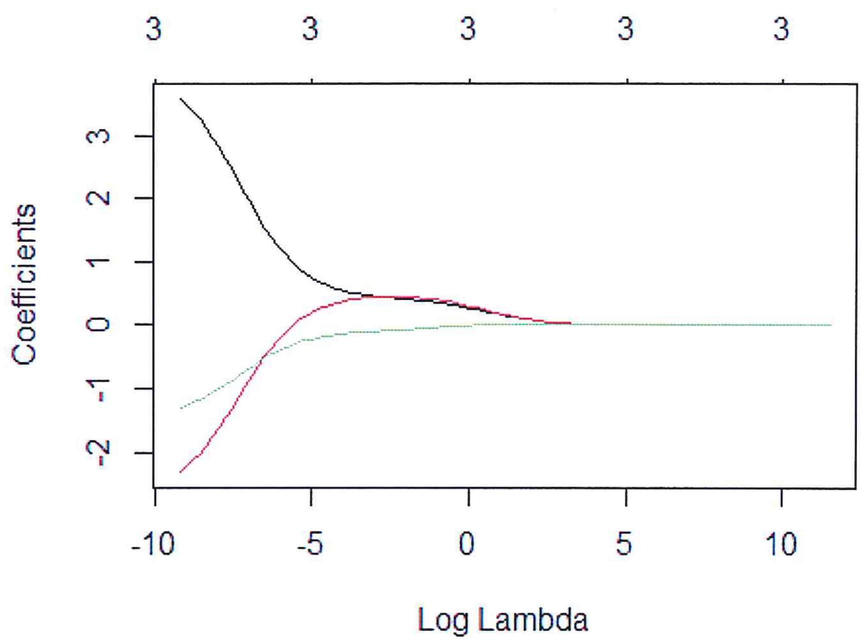
:

```
setwd("F:/Lexar/stat 482 data sets")
multcoll.dat = read.csv('TABLE0701.csv')

library(leaps)

library(glmnet)

multcoll.dat$y = scale(multcoll.dat$body_fat)
multcoll.dat$z1 = scale(multcoll.dat$triceps)
multcoll.dat$z2 = scale(multcoll.dat$thigh)
multcoll.dat$z3 = scale(multcoll.dat$midarm)

mod = lm(y~0+z1+z2+z3,data = multcoll.dat)
summary(mod)
```

```
##
## Call:
## lm(formula = y ~ 0 + z1 + z2 + z3, data = multcoll.dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.72976 -0.31552  0.07683  0.28702  0.80837
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## z1    4.264      2.878   1.481    0.157
## z2   -2.929      2.568  -1.140    0.270
## z3   -1.561      1.106  -1.412    0.176
##
## Residual standard error: 0.4712 on 17 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7663
## F-statistic: 22.86 on 3 and 17 DF,  p-value: 3.363e-06
```

```
z = model.matrix(mod)


full.mod = regsubsets( y ~ z1+z2+z3, multcoll.dat, intercept = FALSE)
selection = summary(full.mod)
selection$cp

## [1] 1.594581 2.300725 3.000000

selection

## Subset selection object
## Call: regsubsets.formula(y ~ z1 + z2 + z3, multcoll.dat, intercept =
FALSE)
## 3 Variables
##    Forced in Forced out
## z2     FALSE      FALSE
## z3     FALSE      FALSE
## 1 subsets of each size up to 3
## Selection Algorithm: exhaustive
##          z1  z2  z3
## 1  ( 1 ) " " "*" " "
## 2  ( 1 ) "*" " " "*"
## 3  ( 1 ) "*" "*" "*"

coef(full.mod,1:3)

## [[1]]
##        z2
## 0.8780896
##
## [[2]]
##         z1         z3
##  0.9843350 -0.3081621
##
## [[3]]
##        z1        z2        z3
##  4.263705 -2.928701 -1.561417
```

```
grid=10^seq(-4,5,length.out = 100) ##get lambda sequence
mod.ridge=glmnet(z,multcoll.dat$y,alpha=0,lambda=grid,standardize = FALSE)
plot(mod.ridge,xvar = "lambda")
```




```
#cv.ridge = cv.glmnet(z,y,alpha=0,standardize = FALSE)
#best.lam.ridge = cv.ridge$lambda.min

#print(log10(best.lam.ridge))

#best.mod.ridge = glmnet(z,y,alpha = 0,lambda = best.lam.ridge,standardize =
FALSE)
#coef(best.mod.ridge)

#results = predict(best.mod.ridge, z, type="response")
```

```
pca = prcomp(z)
summary(pca)
```

```
## Importance of components:
##                           PC1    PC2     PC3
## Standard deviation     1.4375 0.9658 0.02696
## Proportion of Variance 0.6888 0.3109 0.00024
## Cumulative Proportion  0.6888 0.9998 1.00000
```

```
pca$rotation
```

```
##          PC1         PC2        PC3
## z1 0.6946957 -0.05010563  0.7175565
## z2 0.6294279 -0.44050902 -0.6401347
## z3 0.3481645  0.89634883 -0.2744818
```

```
P = pca$rotation[,1:2]
```

```
head(pca$x)
```

```
##            PC1        PC2          PC3
## 1 -1.631888014  1.1007545  0.046261568
## 2 -0.193034481  0.2638694  0.037463010
## 3  1.729318426  2.1900499 -0.024544755
## 4  1.330209942  0.5470338 -0.002570155
## 5 -1.623593621  1.6228630 -0.036283958
## 6 -0.005149624 -1.1960877  0.003307760
```

```
head(z %*% pca$rotation)
```

```
##            PC1        PC2          PC3
## 1 -1.631888014  1.1007545  0.046261568
## 2 -0.193034481  0.2638694  0.037463010
## 3  1.729318426  2.1900499 -0.024544755
## 4  1.330209942  0.5470338 -0.002570155
## 5 -1.623593621  1.6228630 -0.036283958
## 6 -0.005149624 -1.1960877  0.003307760
```

```
multcoll.dat$w1 = pca$x[,1]
multcoll.dat$w2 = pca$x[,2]
```

```
cor(multcoll.dat[,5:10])
```

```
##             y          z1         z2        z3           w1            w2
## y   1.0000000  0.84326545  0.8780896 0.1424440 8.264917e-01 -3.120462e-01
## z1  0.8432654  1.00000000  0.9238425 0.4577772 9.986411e-01 -4.839283e-02
## z2  0.8780896  0.92384251  1.0000000 0.0846675 9.048171e-01 -4.254507e-01
## z3  0.1424440  0.45777716  0.0846675 1.0000000 5.004945e-01  8.657081e-01
## w1  0.8264917  0.99864108  0.9048171 0.5004945 1.000000e+00  8.716134e-17
## w2 -0.3120462 -0.04839283 -0.4254507 0.8657081 8.716134e-17  1.000000e+00
```

```r
mod.pcr = lm(y ~ w1 + w2 -1, data = multcoll.dat)
summary(mod.pcr)
```

```
##
## Call:
## lm(formula = y ~ w1 + w2 - 1, data = multcoll.dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.78066 -0.36921  0.05219  0.25810  0.78815
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## w1  0.57494    0.07683   7.484 6.25e-07 ***
## w2 -0.32309    0.11435  -2.826   0.0112 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4814 on 18 degrees of freedom
## Multiple R-squared:  0.7805, Adjusted R-squared:  0.7561
## F-statistic:    32 on 2 and 18 DF,  p-value: 1.185e-06
```

```r
b.hat.pcr = P %*% mod.pcr$coefficients
b.hat.pcr
```

```
##           [,1]
## z1  0.41559803
## z2  0.50420855
## z3 -0.08942776
```