Stat 415/615, Lab 7. Weighted Least Squares (and Ridge Regression)

Jun Lu

Stat 415/615 Regression, 2023

Contents

Comments and explanations are not included here. We'll discuss them in class.

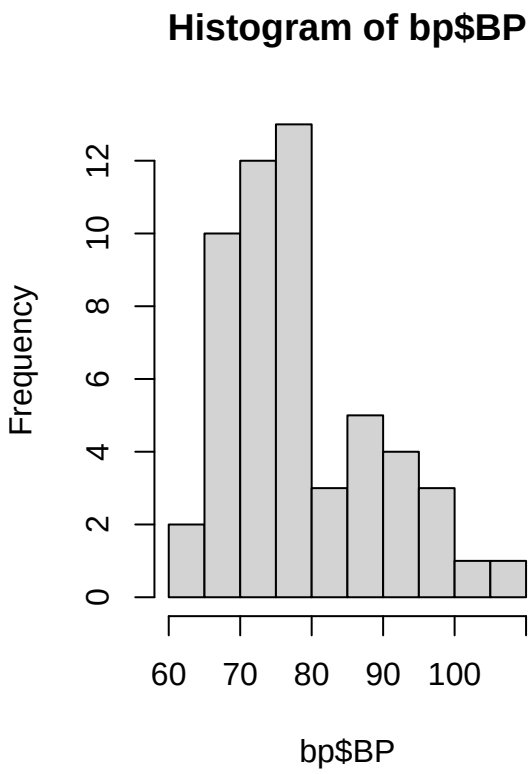
1 Blood Pressure Example, text p.427

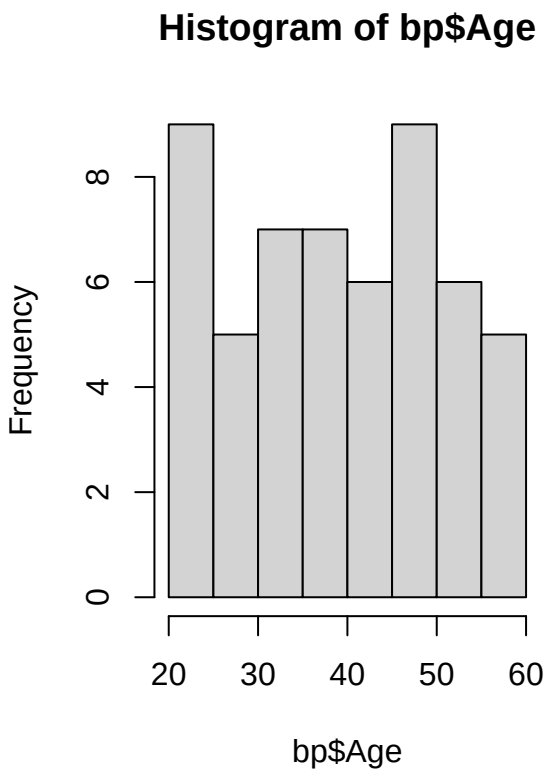
```
bp<-read.table("../DataSets/BloodPressure.txt", header=T)
summary(bp)
```

```
##       Age              BP
##  Min.   :20.00   Min.   : 63.00
##  1st Qu.:30.25   1st Qu.: 71.00
##  Median :40.00   Median : 77.00
##  Mean   :39.57   Mean   : 79.11
##  3rd Qu.:49.00   3rd Qu.: 85.75
##  Max.   :59.00   Max.   :109.00
```

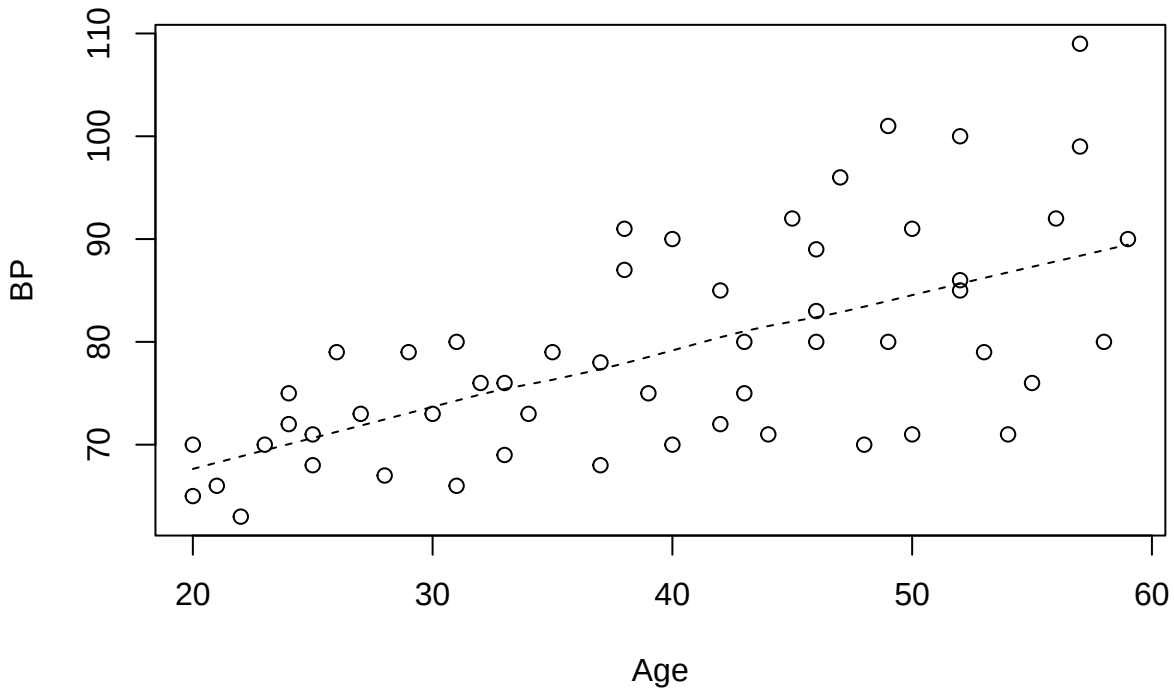
1.1 Plot the data and comment

```
par(mfrow=c(1, 2))
hist(bp$BP)
hist(bp$Age)
```


Histogram of bp$BP
Frequency
0
2
4
6
8
10
12
60
70
80
90
100
bp$BP


Histogram of bp$Age
Frequency
0
2
4
6
8
20
30
40
50
60
bp$Age

```
plot(BP~Age, data=bp)
lines(lowess(bp$Age, bp$BP), lty=2)
```

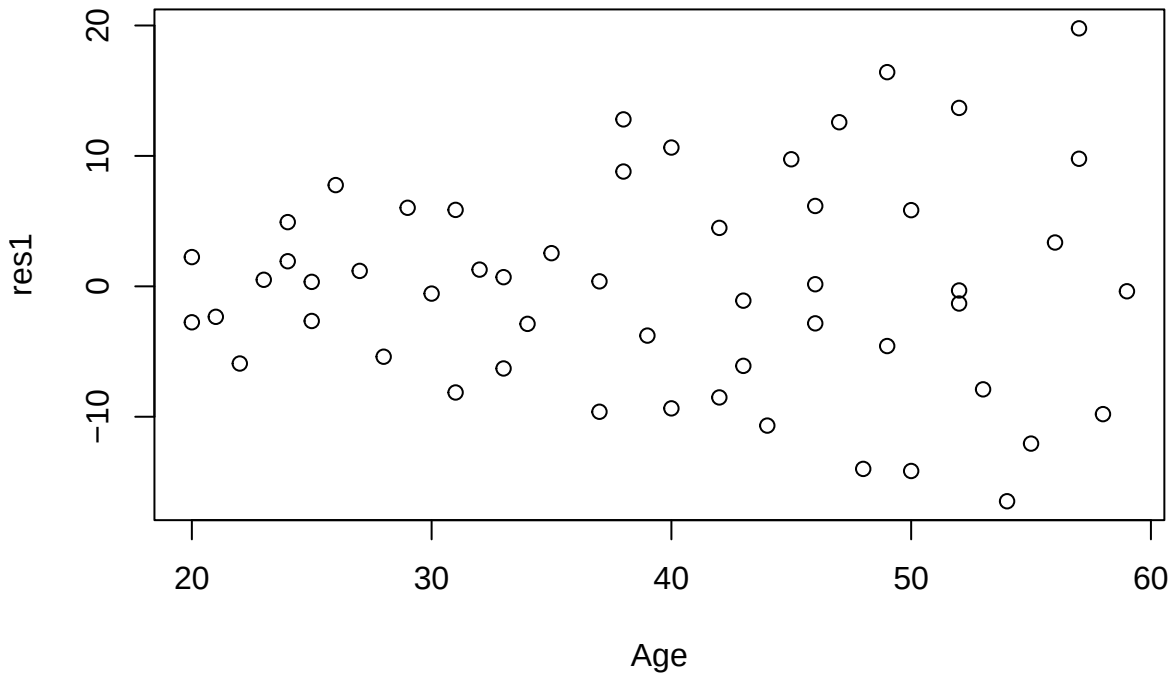


1.2 Fit regression, and save the residuals.

```
bp.ols1<-lm(BP~Age, data=bp)
bp$res1<-bp.ols1$resi
```

1.3 Plot the residuals and comment

```
plot(res1~Age, data=bp)
```



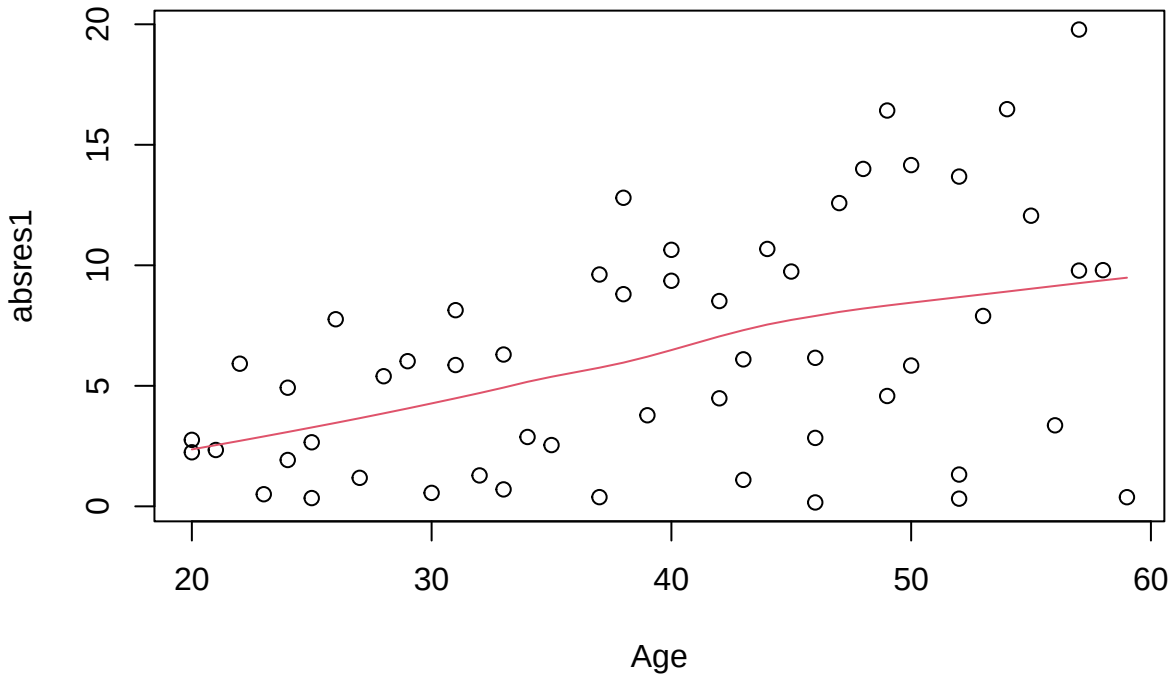
1.4 Find the absolute value of the residuals

```
bp$absres1 <- abs(bp$res1)
head(bp, 3)
```

```
##   Age BP      res1  absres1
## 1  27 73  1.182239 1.182239
## 2  21 66 -2.337576 2.337576
## 3  22 63 -5.917607 5.917607
```

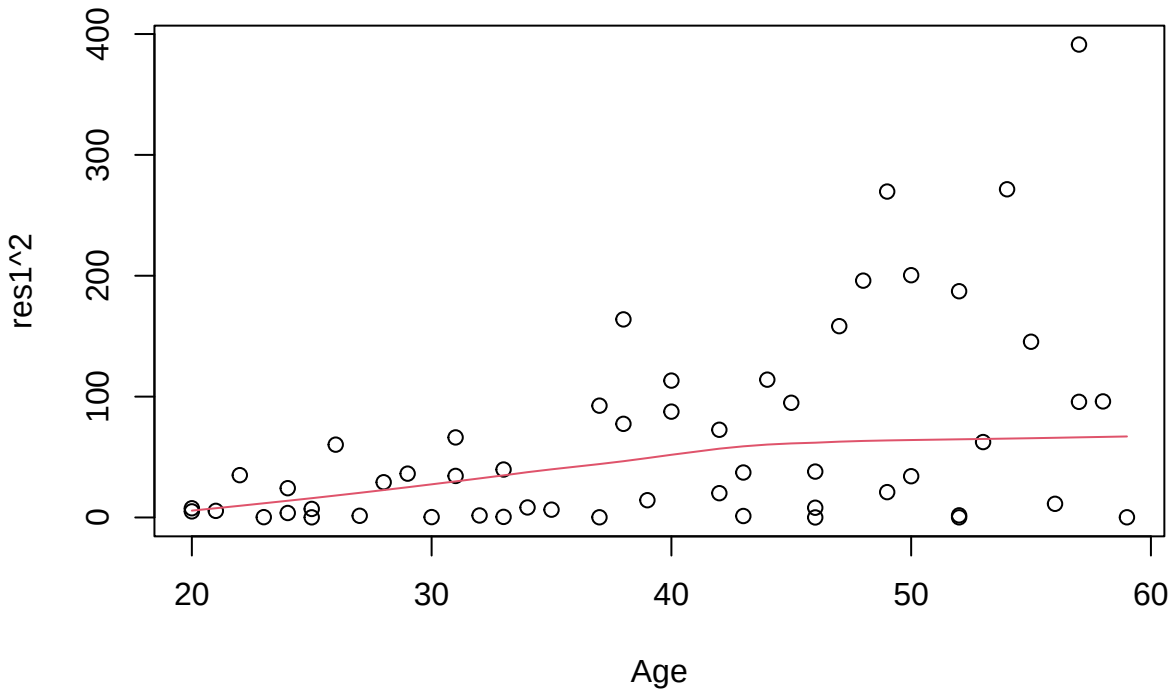
1.5 Predictor vs the absoluate value of the residuals

```
plot(absres1~Age, data=bp)

# Plot the "locally weighted scatterplot smoothing" (Lowess) line to illustrate the pattern.
lines(lowess(bp$absres1~bp$Age), col=2)
```



1.6 Predictor vs the square of the residuals

```
plot(res1^2~Age, data=bp)

# Plot the "locally weighted scatterplot smoothing" (Lowess) line to illustrate the pattern.
lines(lowess(bp$res1^2~bp$Age), col=2)
```



1.7 Use Age to predict s_1 in a linear regression model. Save the predicted values ($\hat{s_1}$).

```
bp$shat1<-lm(absres1~Age, data=bp)$fitted
```

1.8 Compute the weights w_1.

```
bp$w1<-1/(bp$shat1)^2
```

1.9 Weigthted Least Regression

```
bp.wls1<-lm(BP~Age, weights=w1, data=bp)
```

1.10 Compare the results from OLS and WLS

```
summary(bp.ols1)
```

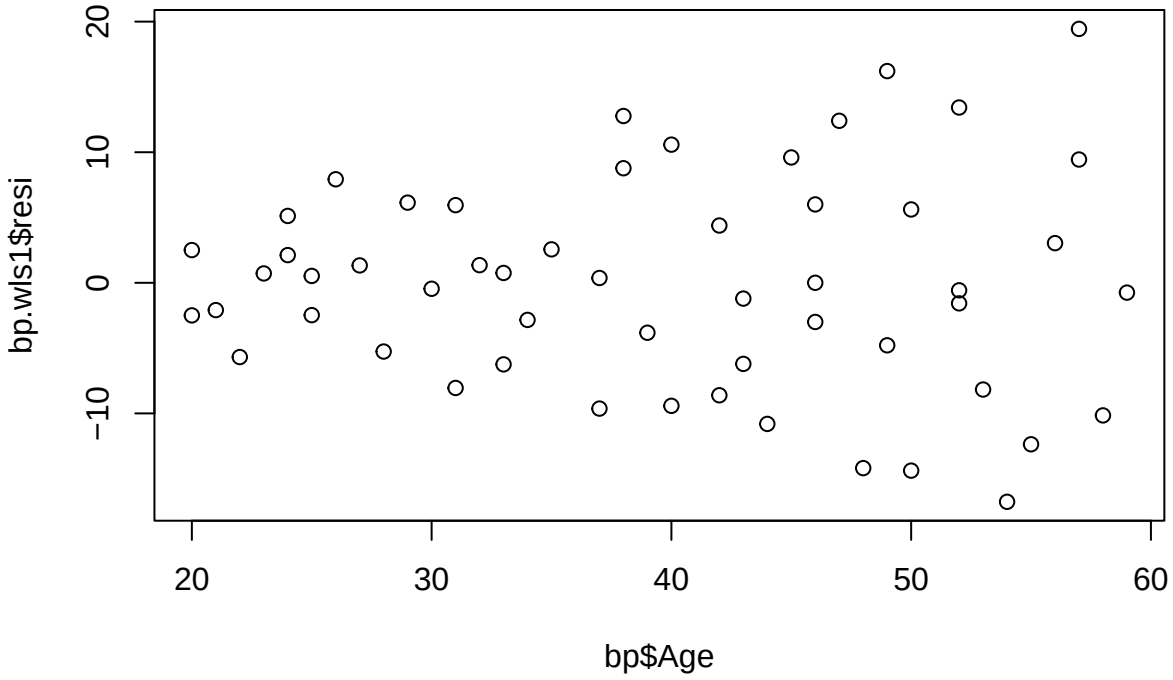
```
##
## Call:
## lm(formula = BP ~ Age, data = bp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.4786  -5.7877  -0.0784   5.6117  19.7813
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 56.15693    3.99367  14.061  < 2e-16 ***
## Age          0.58003    0.09695   5.983 2.05e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.146 on 52 degrees of freedom
## Multiple R-squared:  0.4077, Adjusted R-squared:  0.3963
## F-statistic: 35.79 on 1 and 52 DF,  p-value: 2.05e-07
```

```
summary(bp.wls1)
```

```
##
## Call:
## lm(formula = BP ~ Age, data = bp, weights = w1)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0230 -0.9939 -0.0327  0.9250  2.2008
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 55.56577    2.52092  22.042  < 2e-16 ***
## Age          0.59634    0.07924   7.526 7.19e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.213 on 52 degrees of freedom
## Multiple R-squared:  0.5214, Adjusted R-squared:  0.5122
## F-statistic: 56.64 on 1 and 52 DF,  p-value: 7.187e-10
```

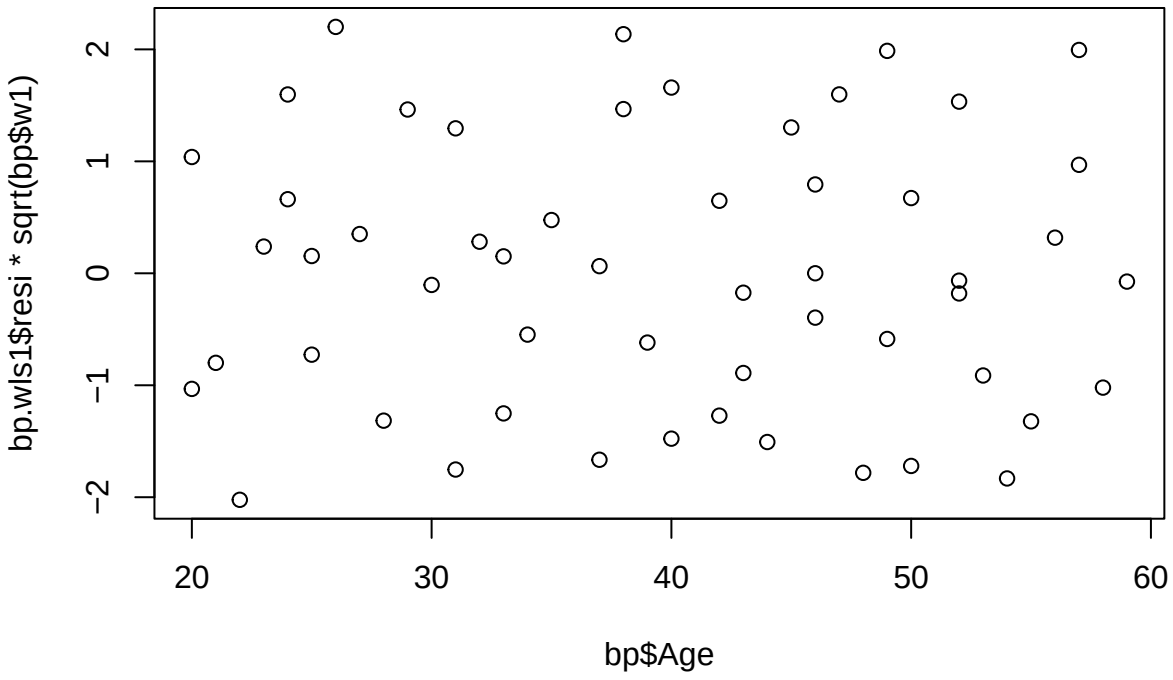
1.11 Plot the residuals from WLS and comment

```
plot(bp$Age, bp.wls1$resi)
```



1.12 Plot (WLS residuals)*sqrt(weights) and comment

```
plot(bp$Age, bp.wls1$resi*sqrt(bp$w1))
```



1.13 Use Iteratively re-Weighted Least Square (IWLS) to improve the weight estimation.

```
bp$resi2<-bp.wls1$res
bp$shat2<-lm(abs(resi2)~Age, data=bp)$fitted
bp$w2<-1/(bp$shat2)^2
bp.wls2<-lm(BP~Age, weights=w2, data=bp)
confint(bp.ols1)
```

```
##                 2.5 %     97.5 %
## (Intercept) 48.1430367 64.1708221
## Age          0.3854841  0.7745775
```

```
confint(bp.wls1)
```

```
##                 2.5 %     97.5 %
## (Intercept) 50.507175 60.6243577
## Age          0.437339  0.7553445
```

```
confint(bp.wls2)
```

```
##                 2.5 %     97.5 %
## (Intercept) 50.5088675 60.6164046
```

```
## Age          0.4374644  0.7553992
```

2 Machine Speed (Homework)

Refer to Ch.11, Problem 11.7 on p.473. This will be part of your next homework. Note that you can skip part (b). If you work on it for extra credit, please read text p.118, 119 for Breusch-Pagan test.

3 Ridge Regression: Body Fat Example revisited (Table 7.1. Optional)

We first worked on this example in Chapter 7. See Rlab3_Multiple1.pdf for exploratory analysis and OLS multiple regression.

```
bfdata <- read.table("../DataSets/CH07TA01.txt", header = F)
colnames(bfdata) <- c("triceps", "thigh", "midarm", "bodyfat")
head(bfdata, 3)
```

```
##   triceps thigh midarm bodyfat
## 1    19.5  43.1   29.1    11.9
## 2    24.7  49.8   28.2    22.8
## 3    30.7  51.9   37.0    18.7
```

```
bfreg1<-lm(bodyfat~triceps+thigh+midarm, data=bfdata)
bfreg1
```

```
##
## Call:
## lm(formula = bodyfat ~ triceps + thigh + midarm, data = bfdata)
##
## Coefficients:
## (Intercept)      triceps        thigh       midarm
##     117.085        4.334       -2.857       -2.186
```

```
library(MASS)  # lm.ridge() function
library(car)   # vif() function
```

```
## Loading required package: carData
```

```
vif(bfreg1)
```

```
##  triceps    thigh   midarm
## 708.8429 564.3434 104.6060
```

- Ridge regression should be applied to standardized data, especially the predictors, so that the predictors will be on the scale.
- Function `lm.ridge()` fits a linear model by ridge regression. It will scale the predictors, fit ridge regression with different ridge parameter value, and transform the results back the original scale of the predictors. I.e., users can use the original data directly.

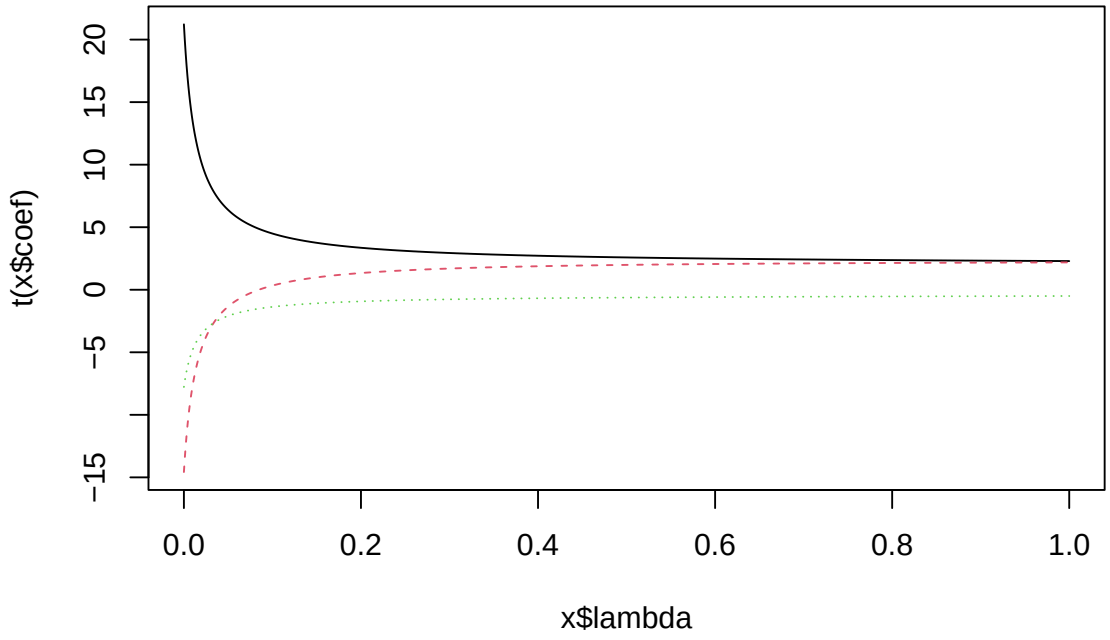
  ```
  bfRidgeReg <- lm.ridge(bodyfat ~ triceps + thigh + midarm,
                         data=bfdata, lambda = seq(0, 1, by = 0.1))
  bfRidgeReg
  ```

  ```
  ##                  triceps       thigh     midarm
  ## 0.0 117.084695 4.3340920 -2.85684794 -2.1860603
  ## 0.1   4.267704 0.9177169  0.06534957 -0.3852088
  ```

```
## 0.2  -3.311810 0.6853025  0.26182628 -0.2618545
## 0.3  -6.029223 0.6000024  0.33237735 -0.2160242
## 0.4  -7.403425 0.5553531  0.36814444 -0.1916269
## 0.5  -8.217704 0.5276560  0.38941512 -0.1761768
## 0.6  -8.745505 0.5086382  0.40327165 -0.1653148
## 0.7  -9.107195 0.4946579  0.41283154 -0.1571221
## 0.8  -9.364034 0.4838616  0.41968127 -0.1506218
## 0.9  -9.550486 0.4752067  0.42471335 -0.1452643
## 1.0  -9.687428 0.4680613  0.42846838 -0.1407166
```

- Ridge trace plot.

```
# Refit with a finer grid
bfRidgeReg <- lm.ridge(bodyfat ~ triceps + thigh + midarm,
                       data=bfdata, lambda = seq(0, 1, by = 0.001))
plot(bfRidgeReg)
```



- Ridge parameter estimation

```
select(bfRidgeReg)
```

```
## modified HKB estimator is 0.008505093
## modified L-W estimator is 0.3098511
## smallest value of GCV  at 0.019
```

```
lm.ridge(bodyfat ~ triceps + thigh + midarm, data=bfdata, lambda = 0.019)
```

```
##               triceps      thigh     midarm
## 43.8401126  2.1174933 -0.9597309 -1.0180612
```

```
lm.ridge(bodyfat ~ triceps + thigh + midarm, data=bfdata, lambda = 0.31)
```

```
##               triceps      thigh     midarm
## -6.2065973  0.5943282  0.3369886 -0.2129466
```

- Reproduce the results from textbook, p.435. Note that `lambda` in R is (c in the textbook)*n.

```
lm.ridge(bodyfat ~ triceps + thigh + midarm, data=bfdata, lambda = 0.02*20)
```

```
##               triceps      thigh     midarm
## -7.4034254  0.5553531  0.3681444 -0.1916269
```

—— This is the end of Lab 7. ——