

BST 210 Homework 2

Wenjie Gu

Problem 1

#1(a)

```
dat = read.csv("SCCS2_v12.csv")
dat$bmi = dat$weight / (dat$height/100)^2
age2_gender.tc = lm(tc~age+I(age^2)+gender, data = dat)
age2_gender_bmi.tc = lm(tc~age+I(age^2)+gender+bmi, data =dat)
age2_bmi.tc = lm(tc~age+I(age^2)+bmi, data= dat)
summary(age2_gender.tc)

##
## Call:
## lm(formula = tc ~ age + I(age^2) + gender, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6152 -0.6114 -0.0555  0.5250  4.1205
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.0390448  0.3913591   7.765 4.32e-14 ***
## age          0.0909455  0.0186723   4.871 1.48e-06 ***
## I(age^2)     -0.0007241  0.0002073  -3.494 0.000517 ***
## gender       0.0944912  0.0851936   1.109 0.267882
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9691 on 523 degrees of freedom
## Multiple R-squared:  0.1525, Adjusted R-squared:  0.1476
## F-statistic: 31.36 on 3 and 523 DF,  p-value: < 2.2e-16
summary(age2_bmi.tc)

##
## Call:
## lm(formula = tc ~ age + I(age^2) + bmi, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7195 -0.6100 -0.0762  0.5396  4.1447
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.6352317  0.4136506   6.371 4.13e-10 ***
## age          0.0798443  0.0189545   4.212 2.98e-05 ***
## I(age^2)     -0.0006326  0.0002084  -3.036 0.00252 **
## bmi          0.0305255  0.0102095   2.990 0.00292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.9621 on 523 degrees of freedom
## Multiple R-squared:  0.1648, Adjusted R-squared:  0.16
## F-statistic: 34.39 on 3 and 523 DF,  p-value: < 2.2e-16
summary(age2_gender_bmi.tc)

##
## Call:
## lm(formula = tc ~ age + I(age^2) + gender + bmi, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6861 -0.5977 -0.0903  0.5483  4.1049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.6223982   0.4139550   6.335 5.13e-10 ***
## age          0.0792126   0.0189701   4.176 3.48e-05 ***
## I(age^2)     -0.0006227   0.0002087  -2.984  0.00298 **
## gender        0.0774204   0.0847893   0.913  0.36162
## bmi          0.0298810   0.0102355   2.919  0.00366 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9622 on 522 degrees of freedom
## Multiple R-squared:  0.1661, Adjusted R-squared:  0.1597
## F-statistic: 25.99 on 4 and 522 DF,  p-value: < 2.2e-16
```

If bmi confounds the effect of gender, bmi needs to follow the following conditions:

- 1) bmi needs to be associated with gender in the study base (potentially yes)
- 2) bmi needs to be associated with total cholesterol in the absence of exposure (potentially yes)
- 3) bmi is not a downstream consequence of total cholesterol or gender (potentially yes)

Therefore, it is worth testing if bmi truly confounds the effect of gender.

After adjusting for linear and quadratic age, adding bmi to the linear regression alters the coefficient of gender from 0.0945 to 0.0774, which decreases by 18%. The change in the coefficient of gender is larger than 10%, indicating that bmi confounds the effect of gender.

If gender confounds the effect of bmi, gender needs to qualify the following:

- 1) gender needs to be associated with bmi in the study base (potentially yes)
- 2) gender needs to be associated with total cholesterol in the absence of bmi (no, the gender predictor alone does not tell info about tc)
- 3) gender is not a downstream consequence of total cholesterol or bmi (yes)

According to the conventional definition of confounders, gender is not a confounder for the effect of bmi. Let's take a look at the data:

After adjusting for linear and quadratic age, adding gender to the linear regression alters the coefficient of bmi from 0.0305 to 0.0299, which decreases by 1.9% (< 10%), indicating that gender does not confound the effect of bmi.

#1(b)

```
age2_gender_bmi_emm.tc = lm(tc~age+I(age^2)+bmi+gender+gender*bmi, data = dat)
summary(age2_gender_bmi_emm.tc)
```

```
##
## Call:
## lm(formula = tc ~ age + I(age^2) + bmi + gender + gender * bmi,
##     data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7144 -0.6025 -0.0793  0.5501  4.1128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.4450100  0.4983415   4.906 1.24e-06 ***
## age          0.0790801  0.0189820   4.166 3.63e-05 ***
## I(age^2)     -0.0006209  0.0002088  -2.973  0.00308 **
## bmi          0.0373623  0.0155395   2.404  0.01655 *
## gender       0.3810542  0.4818589   0.791  0.42942
## bmi:gender   -0.0125739  0.0196427  -0.640  0.52237
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9628 on 521 degrees of freedom
## Multiple R-squared:  0.1667, Adjusted R-squared:  0.1587
## F-statistic: 20.85 on 5 and 521 DF,  p-value: < 2.2e-16
```

Since the p-value for the coefficient of bmi&gender interaction term is 0.5224, which is way larger than 0.05, bmi does not serve as an effect modifier of the effect of gender. Gender does not serve as an effect modifier either, because gender and bmi share the same interaction term and being an effect modifier is reciprocal.

Problem 2

#2(a)

```
summary(age2_gender_bmi.tc)
```

```
##
## Call:
## lm(formula = tc ~ age + I(age^2) + gender + bmi, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6861 -0.5977 -0.0903  0.5483  4.1049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.6223982  0.4139550   6.335 5.13e-10 ***
## age          0.0792126  0.0189701   4.176 3.48e-05 ***
## I(age^2)     -0.0006227  0.0002087  -2.984  0.00298 **
## gender       0.0774204  0.0847893   0.913  0.36162
## bmi          0.0298810  0.0102355   2.919  0.00366 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.9622 on 522 degrees of freedom
## Multiple R-squared: 0.1661, Adjusted R-squared: 0.1597
## F-statistic: 25.99 on 4 and 522 DF, p-value: < 2.2e-16
```

$$tc = 2.6223 + 0.079 * age - 0.0006227 * age^2 + 0.0774 * gender + 0.02988 * bmi$$

$$tc(\text{man}, 30\text{-yo}, \text{bmi}30) = 2.6223 + 0.079 * 30 - 0.0006227 * (30^2) + 0.0774 * 0 + 0.02988 * 30 = 5.33 \text{ mmol/l}$$

After unit conversion, predicted tc = 5.33 * 38.67 = 206.11 mg/dl.

#2(b)

In R, `rstandard()` gives the internally studentized residuals and `rstudent()` gives the externally studentized residuals.

Raw residuals(`residuals()` in R):

$$e_i = Y_i - \hat{Y}_i$$

Standardized residuals(no specific function in R, but can be calculated by dividing raw residuals by the estimated standard deviation):

$$z_i = \frac{e_i}{s_e} = \frac{e_i}{\sqrt{SSE/(N-p-1)}}$$

Internally studentized residuals(`rstandard()` in R):

$$r_i = \frac{e_i}{SE(e_i)} = \frac{e_i}{\hat{\sigma} \sqrt{1 - H_{ii}}}$$

Externally studentized residuals(`rstudent()` in R):

$$r_{(i)} = \frac{e_i}{\hat{\sigma}_{(i)} \sqrt{1 - H_{ii}}}$$

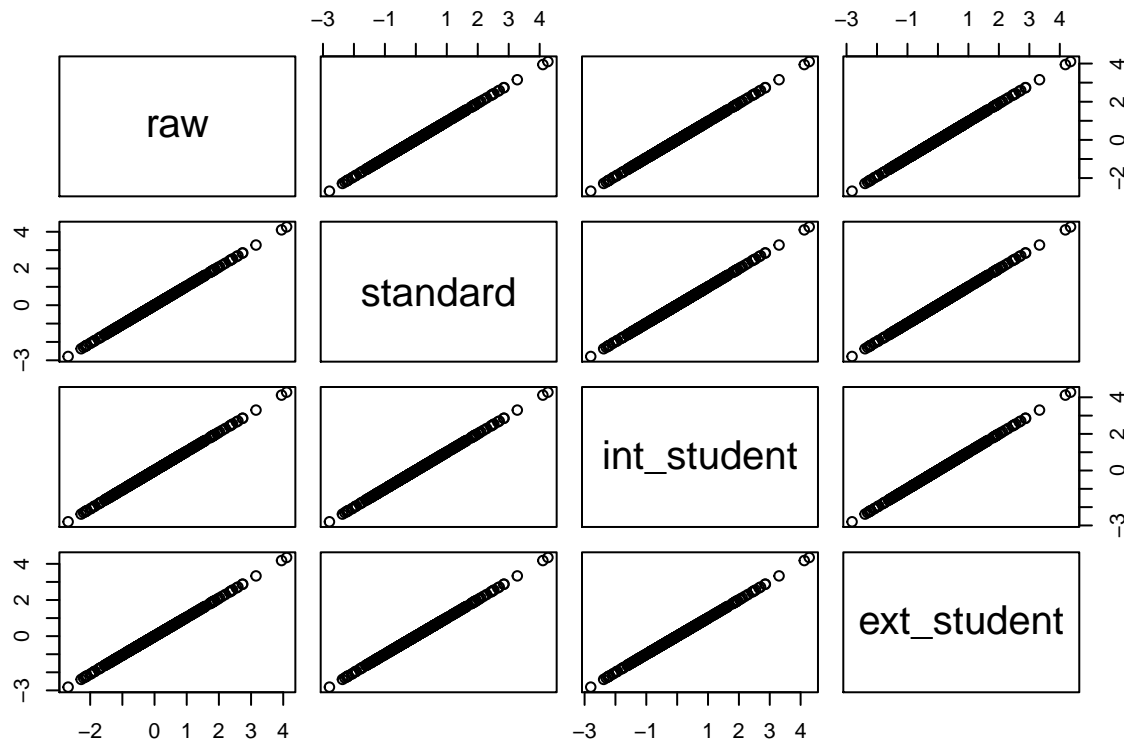
Reference: https://web.stanford.edu/class/stats191/notebooks/Diagnostics_for_multiple_regression.html#Types-of-residuals-and-lab-notes.

#2(c)

```
standardized_res = function(model,n,p){
  raw = residuals(model)
  SSE = sum(raw^2)
  res = raw/(sqrt(SSE/(n-p-1)))
  return(res)
}

standard = standardized_res(age2_gender_bmi.tc, 527, 4)

raw = residuals(age2_gender_bmi.tc)
standard = standardized_res(age2_gender_bmi.tc, 527, 4)
int_student = rstandard(age2_gender_bmi.tc)
ext_student = rstudent(age2_gender_bmi.tc)
residuals = data.frame(raw,standard,int_student,ext_student)
plot(residuals)
```



```
cor(residuals, method = "pearson")
```

```
##               raw  standard int_student ext_student
## raw           1.0000000 1.0000000  0.9999964  0.9999888
## standard      1.0000000 1.0000000  0.9999964  0.9999888
## int_student   0.9999964 0.9999964  1.0000000  0.9999911
## ext_student   0.9999888 0.9999888  0.9999911  1.0000000
```

Pearson correlation coefficient between raw and standardized residuals is 1 which makes sense because standardized residuals are directly derived from raw residuals. Pearson correlation coefficients between other pairs of residuals are just very close to 1, which indicates strong linear association between either pair of residuals, but not exactly 1 due to slight difference in the ways they're calculated.

For this SCCS2 example, we can use any residual but the scaled residuals (standardized and studentized) are preferred.

#2(d)

```
summary(raw)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.6861 -0.5977 -0.0903  0.0000  0.5483  4.1049
```

```
summary(standard)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.79153 -0.62113 -0.09385  0.00000  0.56978  4.26604
```

```
summary(ext_student)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.818656 -0.623985 -0.094891  0.000411  0.571104  4.351831
```

```
summary(int_student)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -2.800091 -0.624351 -0.094982 -0.000089  0.571473  4.278930
```

The mean raw residuals is zero.

The mean standardized residuals is also zero.

The mean ext_studentized residuals is 0.00041, which is close to zero.

The mean int_studentized residuals is -0.000089, which is close to zero.

Assuming the regression line is the “best” fit, then the sum of residuals should be zero, thus the raw and standardized residuals means always equal zero in the “best” fit model.

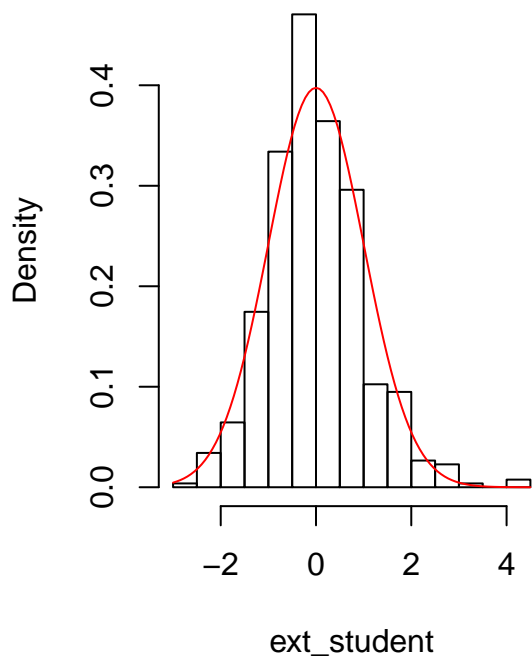
While the internal and external studentized residuals should have non-zero means since each residual is scaled with respect to a specific term associated with each observation.

#2(e)

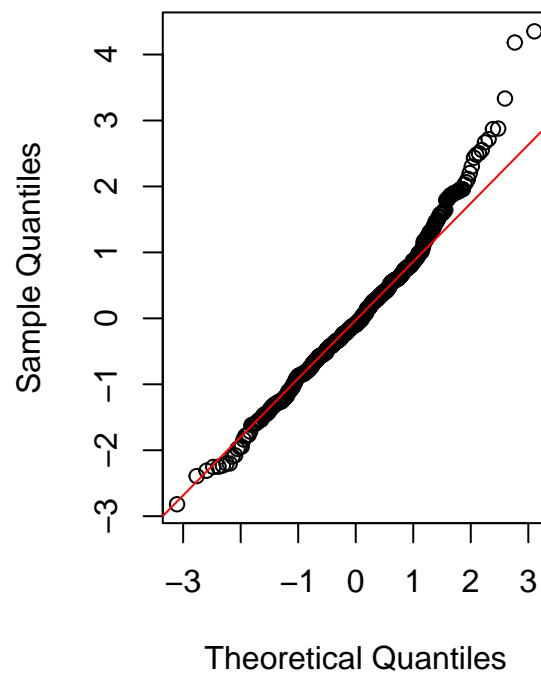
```
par(mfrow = c(1,2))
hist(ext_student, probability = T)
m = mean(ext_student)
std = sqrt(var(ext_student))
curve(dnorm(x, mean = m, sd = std), col = "red", add = T)

qqnorm(ext_student)
qqline(ext_student, col = "red")
```

Histogram of ext_student



Normal Q-Q Plot



According to the histogram, the median is slightly to the left of zero and the distribution is slightly right skewed.

The normal qq plot of the external studentized residuals has a longer right tail deviating up from the reference normal qq line.

Therefore, the external studentized residuals are not normally distributed.

#2(f)

```
dat$stdres = standard
resid2 = dat[dat$stdres > 2, c("age", "gender", "bmi", "tc", "stdres")]
resid3 = dat[dat$stdres > 3, c("age", "gender", "bmi", "tc", "stdres")]
resid2
```

```
##          age gender      bmi   tc   stdres
## 43  20.79398      0 21.56679 7.07 2.520523
## 66  48.34497      0 20.34894 7.97 2.458422
## 92  48.31485      0 21.87004 7.59 2.016860
## 96  40.42163      0 22.99086 7.82 2.417567
## 103 54.57906      0 29.83798 9.93 4.102769
## 149 28.74196      0 23.27265 7.47 2.483763
## 191 39.44422      0 24.41728 8.10 2.694207
## 243 53.38809      0 28.77855 7.91 2.051174
## 256 57.27584      0 23.05327 8.35 2.644643
## 284 41.52772      0 25.11651 8.33 2.849190
## 387 63.32649      1 23.42319 8.02 2.183908
## 409 57.07050      1 24.02658 8.11 2.286241
## 451 46.19028      1 25.94075 9.91 4.266038
## 512 58.08898      1 29.40531 8.81 2.838760
## 514 67.12663      1 24.61937 9.10 3.277174
## 516 51.52088      1 31.25248 8.08 2.097499
```

```
resid3
```

```
##          age gender      bmi   tc   stdres
## 103 54.57906      0 29.83798 9.93 4.102769
## 451 46.19028      1 25.94075 9.91 4.266038
## 514 67.12663      1 24.61937 9.10 3.277174
```

```
predict(age2_gender_bmi.tc)[c(103,451,514)]
```

```
##          103          451          514
## 5.982244 5.805144 5.946646
```

There are three individuals having standardized residuals >3 . For each of the cases, the predicted tc value is much lower than the true tc. For subject 103, the predicted tc is 5.98, while his real tc is 9.93. For subject 451, the predicted tc is 5.81, whereas her real tc is 9.91. And for subject 514, the predicted tc is 5.95 while her real tc is 9.10. The high residuals of these three observations are resulting from the large difference between the predicted and true values.

Problem 3

#3(a)

```
hat_val = hat(model.matrix(age2_gender_bmi.tc))
sum(hat_val < 0)
```

```
## [1] 0
```

```
mean(hat_val)
```

```
## [1] 0.009487666
```

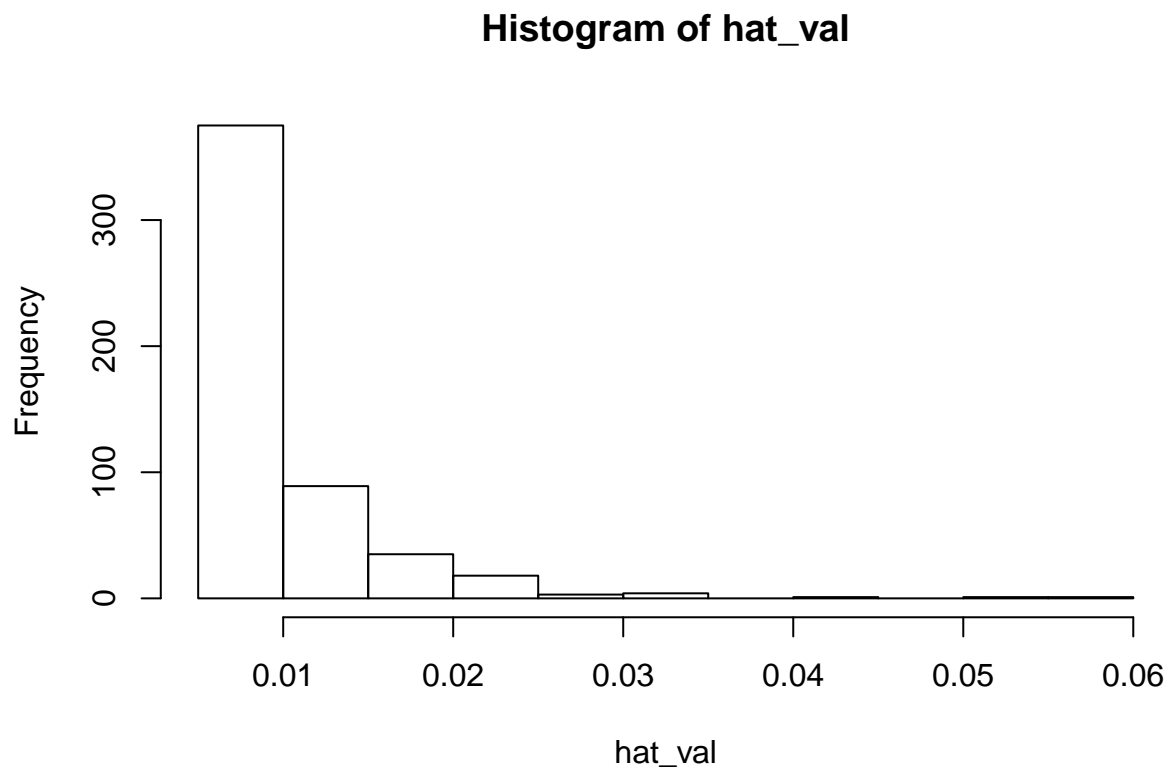
```
p = 4  
n = nrow(dat)  
(p+1)/n
```

```
## [1] 0.009487666
```

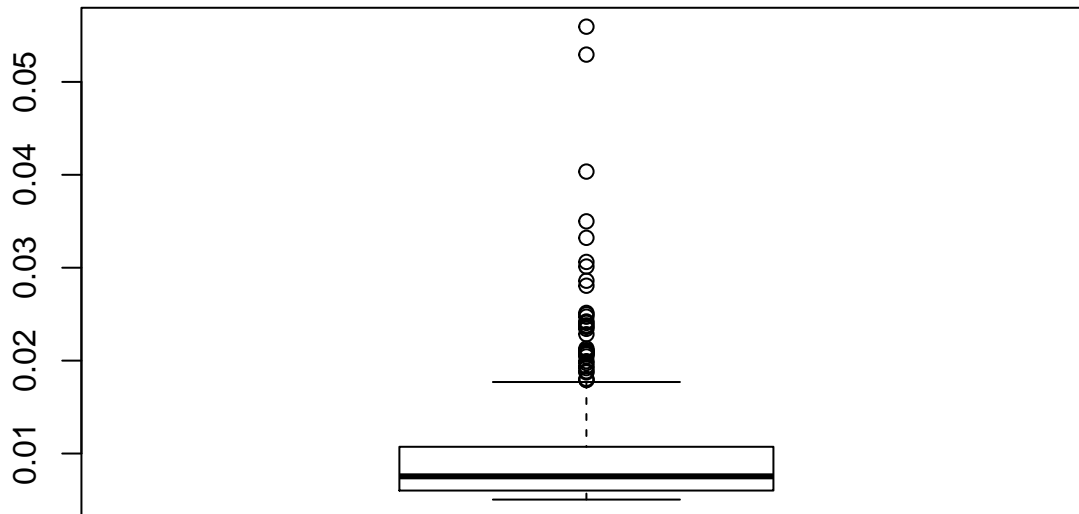
After counting for negative elements, the function returns zero, which means no negative element in the hat values list. The mean value of `hat_val` is 0.00949, where $(p+1)/n$ also equals 0.00949. Therefore, in this case, all hat values are positive and average out to equal $(p+1)/n$.

#3(b)

```
dat$hat = hat_val  
hist(hat_val)
```



```
boxplot(hat_val)
```

```
lev1 = dat[dat$hat > 2*(p+1)/n, c("age", "gender", "bmi", "tc", "hat")]
head(lev1)
```

```
##      age gender      bmi   tc      hat
## 22  18.55989     0 24.61002 4.40 0.02130464
## 44  17.88912     0 18.99863 4.95 0.02088954
## 101 70.64203     0 21.43580 4.80 0.02089109
## 163 78.28063     0 24.52592 4.76 0.05294147
## 171 54.45585     0 39.52477 6.18 0.02806296
## 177 69.31691     0 31.04451 5.65 0.02068107
```

```
lev2 = dat[dat$hat > 4*(p+1)/n, c("age", "gender", "bmi", "tc", "hat")]
lev2
```

```
##      age gender      bmi   tc      hat
## 163 78.28063     0 24.52592 4.76 0.05294147
## 242 78.74332     0 23.12740 6.42 0.05594461
## 475 67.48528     1 40.49367 5.90 0.04034805
```

```
attach(dat)
mean(age)
```

```
## [1] 43.48676
```

```
mean(bmi)
```

```
## [1] 24.16128
```

```
detach(dat)
```

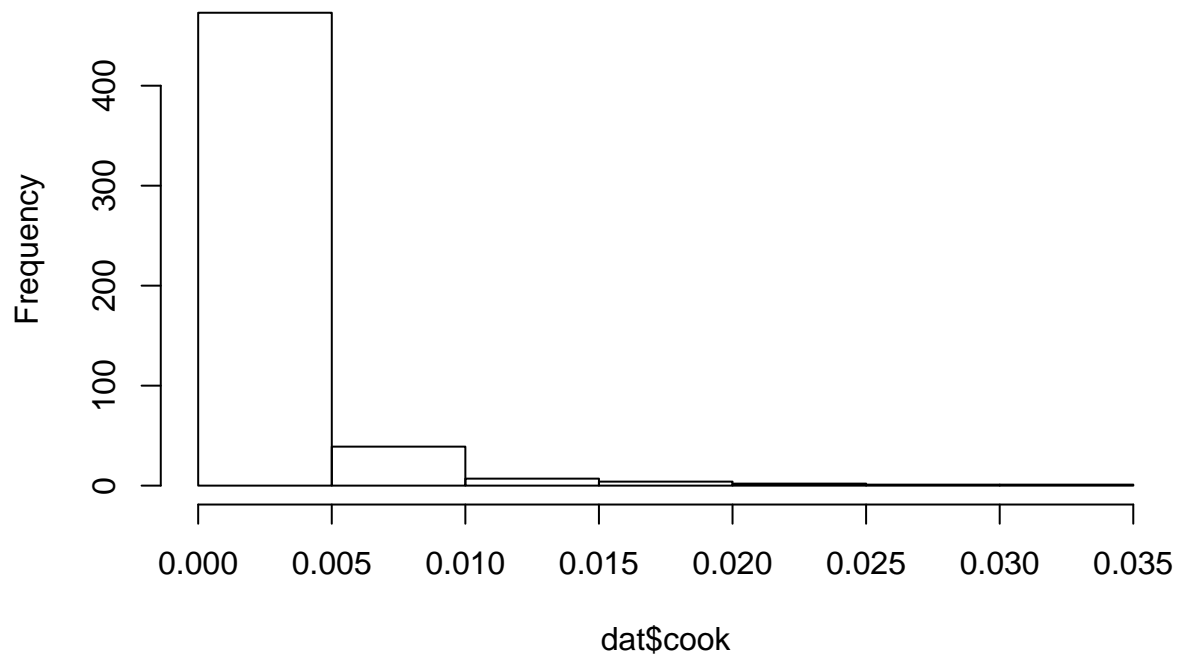
A total of 3 individuals have leverage $> 4(p+1)/n$. Leverage measures how far observation i 's covariates are from the overall covariate average. For the individuals with high leverage, they are of older age (> 78 yo), much larger than average age of 43. The other individual with high leverage has extremely high bmi value ($\text{bmi} = 40.49$), much higher than average bmi of 24.16.

Problem 4

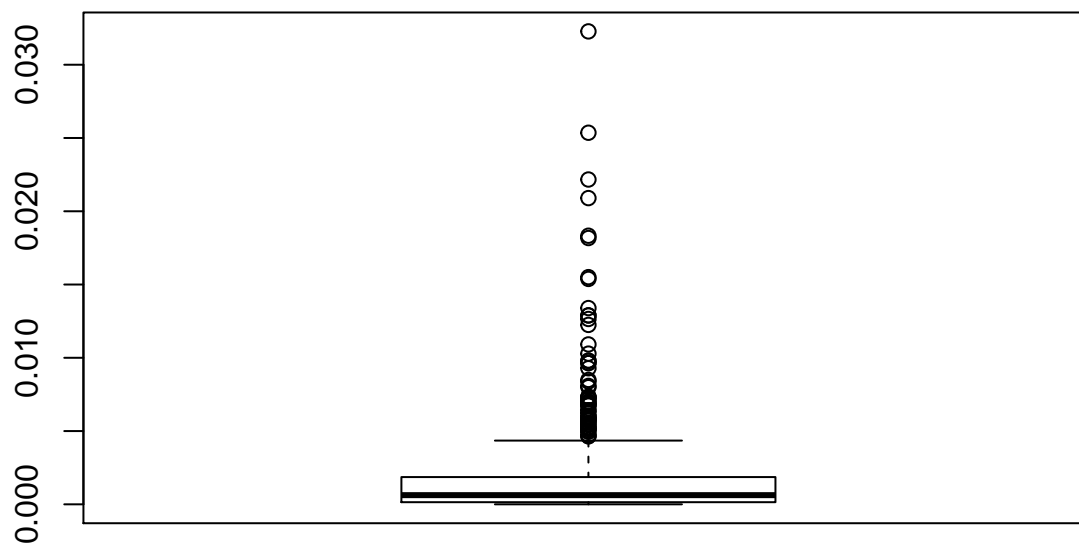
#4(a)

```
dat$cook = cooks.distance(age2_gender_bmi.tc)
hist(dat$cook)
```

Histogram of dat\$cook



```
boxplot(dat$cook)
```



```
sum(dat$cook <= 0)
```

```
## [1] 0
```

No non-positive cook's distance in this example.

#4(b)

```

cook1 = dat[dat$cook > 4/n, c("age", "gender", "bmi", "tc", "cook")]
cook2 = dat[dat$cook > 12/n, c("age", "gender", "bmi", "tc", "cook")]
head(cook1)

```

```

##      age gender      bmi   tc      cook
## 25  63.53456      0 16.40226 3.78 0.012648252
## 43  20.79398      0 21.56679 7.07 0.018331201
## 66  48.34497      0 20.34894 7.97 0.009814110
## 93  24.73922      0 20.78217 2.54 0.009640282
## 99  68.68720      0 25.80639 3.90 0.013392039
## 103 54.57906      0 29.83798 9.93 0.025358620

```

```

cook2

```

```

##      age gender      bmi   tc      cook
## 103 54.57906      0 29.83798 9.93 0.02535862
## 514 67.12663      1 24.61937 9.10 0.03227442

```

There are two observations with high cook's distance ($>4/n$), this is probably due to their high total cholesterol level though their age and bmi are not too high, which could potentially have high influence on the regression model.

#4(c)

```

dat$dffit = dffits(age2_gender_bmi.tc)
high_dfit = dat[abs(dat$dffit)>(2*sqrt((p+1)/n)), c("age", "gender", "bmi", "tc", "dfit")]
high_dfit = high_dfit[order(-abs(high_dfit$dffit)),]
high_dfit

```

```

##      age gender      bmi   tc      dfit
## 514 67.12663      1 24.61937 9.10 0.4055829
## 103 54.57906      0 29.83798 9.93 0.3616622
## 451 46.19028      1 25.94075 9.91 0.3385871
## 468 32.07666      1 36.42918 3.72 -0.3242795
## 43  20.79398      0 21.56679 7.07 0.3043415
## 455 70.35729      1 30.48458 7.90 0.3021814
## 459 38.62560      1 33.87477 3.70 -0.2795034
## 477 69.21835      1 37.40035 4.94 -0.2776338
## 99  68.68720      0 25.80639 3.90 -0.2596078
## 512 58.08898      1 29.40531 8.81 0.2556269
## 498 56.22450      1 15.88121 7.48 0.2545604
## 25  63.53456      0 16.40226 3.78 -0.2521480
## 163 78.28063      0 24.52592 4.76 -0.2474526
## 273 73.29227      0 19.81879 4.43 -0.2337445
## 422 36.58864      1 37.53626 4.54 -0.2269843
## 66  48.34497      0 20.34894 7.97 0.2226094
## 387 63.32649      1 23.42319 8.02 0.2219687
## 204 59.16496      0 27.67561 3.27 -0.2209856
## 93  24.73922      0 20.78217 2.54 -0.2205392
## 284 41.52772      0 25.11651 8.33 0.2169576
## 430 58.23135      1 18.95034 7.56 0.2065473
## 516 51.52088      1 31.25248 8.08 0.2053538
## 256 57.27584      0 23.05327 8.35 0.2022190
## 191 39.44422      0 24.41728 8.10 0.2011222

```

Yes, there are a few individuals that have high influence (

$$|DFFITs_i| > 2\sqrt{(p+1)/n}$$

). After sorting by dffits values, we found that the first two observations with the highest dffits values are the ones observed from #4(b) with high cook's distances.

Problem 5

#5(a)

```
dat[(dat$stdres > 3) & (dat$cook > 12/n) & (dat$hat > 4*(p+1)/n), c("age", "gender", "bmi", "tc", "hat", "cook", "stdres")]

## [1] age    gender bmi    tc    hat    cook    stdres
## <0 rows> (or 0-length row.names)

dat[(dat$stdres > 3) & (dat$cook > 12/n), c("age", "gender", "bmi", "tc", "hat", "cook", "stdres")]

##          age gender      bmi    tc      hat      cook    stdres
## 103 54.57906      0 29.83798 9.93 0.007421154 0.02535862 4.102769
## 514 67.12663      1 24.61937 9.10 0.014590264 0.03227442 3.277174

dat[(dat$stdres > 3) & (dat$hat > 4*(p+1)/n), c("age", "gender", "bmi", "tc", "hat", "cook", "stdres")]

## [1] age    gender bmi    tc    hat    cook    stdres
## <0 rows> (or 0-length row.names)
```

There is no individual meeting all of the three conditions (high standardized residual, high cook's distance and high leverage). There are two individuals having both high standardized residual and high cook's distance. There is no overlap between high standardized residual and high leverage.

#5(b)

```
dat[(dat$stdres > 3) | (dat$cook > 12/n) | (dat$hat > 4*(p+1)/n), c("age", "gender", "bmi", "tc", "hat", "cook", "stdres")]

##          age gender      bmi    tc      hat      cook    stdres
## 103 54.57906      0 29.83798 9.93 0.007421154 0.02535862 4.102769
## 163 78.28063      0 24.52592 4.76 0.052941469 0.012244316 -1.0184298
## 242 78.74332      0 23.12740 6.42 0.055944608 0.007234465 0.7591132
## 451 46.19028      1 25.94075 9.91 0.006016941 0.022166501 4.2660375
## 475 67.48528      1 40.49367 5.90 0.040348053 0.002552495 -0.5397217
## 514 67.12663      1 24.61937 9.10 0.014590264 0.032274418 3.2771738

newdat = dat[!((dat$stdres > 3) | (dat$cook > 12/n) | (dat$hat > 4*(p+1)/n)), c("age", "gender", "bmi", "tc", "hat", "cook", "stdres")]
nrow(newdat)

## [1] 521

newmodel = lm(tc ~ age + I(age^2) + gender + bmi, data = newdat)
summary(newmodel)

##
## Call:
## lm(formula = tc ~ age + I(age^2) + gender + bmi, data = newdat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.65432 -0.57138 -0.07401  0.56057  2.78131
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.7164836  0.4078872   6.660 7.03e-11 ***
## age          0.0762966  0.0189559   4.025 6.55e-05 ***
## I(age^2)     -0.0006011  0.0002105  -2.856 0.00447 **
## gender       0.0607624  0.0816723   0.744 0.45723
## bmi         0.0288282  0.0099503   2.897 0.00392 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9221 on 516 degrees of freedom
## Multiple R-squared:  0.166, Adjusted R-squared:  0.1595
## F-statistic: 25.68 on 4 and 516 DF,  p-value: < 2.2e-16
summary(age2_gender_bmi.tc)
```

```
##
## Call:
## lm(formula = tc ~ age + I(age^2) + gender + bmi, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6861 -0.5977 -0.0903  0.5483  4.1049
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.6223982  0.4139550   6.335 5.13e-10 ***
## age          0.0792126  0.0189701   4.176 3.48e-05 ***
## I(age^2)     -0.0006227  0.0002087  -2.984 0.00298 **
## gender       0.0774204  0.0847893   0.913 0.36162
## bmi         0.0298810  0.0102355   2.919 0.00366 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9622 on 522 degrees of freedom
## Multiple R-squared:  0.1661, Adjusted R-squared:  0.1597
## F-statistic: 25.99 on 4 and 522 DF,  p-value: < 2.2e-16
```

After eliminating 6 observations with standardized residual >3 , cook's distance $>12/n$, or leverage $>4(p+1)/n$ and fitting regression model to the new data, we found that the coefficient of age changes from 0.0792 to 0.0763(-3.7%), the coefficient for quadratic age from -0.0006227 to -0.0006011 (+3.4%), the coefficient for gender from 0.0774 to 0.0608(-21%), and the coefficient for bmi from 0.0299 to 0.0288(-3.7%). Only the change of coefficient for gender is significant. The overall findings do not change much.