HW 4 Math158

Joshua Jansen-Montoya

2022-09-29

Problem 4.1

For the prostate data, fit a model with lpsa as the response and the other variables as predictors. 1. Suppose a new patient with the following values arrives: lcavol = 1.44692, lweight = 3.62301 age = 65, lbph = 0.30010, svi = 0, lcp = -0.79851, gleason = 7, pgg45 = 15. Predict the lpsa for this patient along with an appropriate 95% CI. 2. Repeat the last question for a patient with the same values except that he is age 20. Explain why the CI is wider. 3. For the model of the previous question, remove all the predictors that are not significant at the 5% level. Now recompute the predictions of the previous question. Are the CIs wider or narrower? Which predictions would you prefer? Explain.

Answer 4.1

1. We can predict the lpsa for this patient using the following R-code,

```
library(faraway)
reg <- lm(lpsa ~ lcavol + lweight + age + lbph + svi + lcp + gleason + pgg45, data = prostate)
predict(reg, data.frame(lcavol = 1.44692, lweight = 3.62301, age = 65, lbph = 0.30010, svi = 0, lcp =
```

```
##        1
## 2.389053
```

```
predict(reg,new=data.frame(lcavol = 1.44692, lweight = 3.62301, age = 65, lbph = 0.30010, svi = 0, lc
```

```
##        fit      lwr      upr
## 1 2.389053 2.172437 2.605669
```

We can see that our predicted value for the lpsa for the patient is 2.389053. Looking at our data, we can see that our 95% confidence interval is (2.17, 2.61). 2. Now, repeating the first part with the age 20, we can find that,

```
library(faraway)
  reg <- lm(lpsa ~ lcavol + lweight + age + lbph + svi + lcp + gleason + pgg45, data = prostate)
  predict(reg,new=data.frame(lcavol = 1.44692, lweight = 3.62301, age = 20, lbph = 0.30010, svi = 0, lc
```

```
##        fit      lwr      upr
## 1 3.272726 2.260444 4.285007
```

which gives us a 95% confidence interval of (2.26, 4.29). We can note that our CI is far larger, which is due to the fact that we are extrapolating from the data set, since a majority of the data is for ages greater than 20. 3. We can first find all the predictors that are not significant at a 95% confidence interval as follows,

```
library(faraway)
reg <- lm(lpsa ~ lcavol + lweight + age + lbph + svi + lcp + gleason + pgg45, data = prostate)
summary(reg)
```

```
##
## Call:
```

```
## lm(formula = lpsa ~ lcavol + lweight + age + lbph + svi + lcp +
##     gleason + pgg45, data = prostate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7331 -0.3713 -0.0170  0.4141  1.6381
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.669337   1.296387   0.516  0.60693
## lcavol       0.587022   0.087920   6.677 2.11e-09 ***
## lweight      0.454467   0.170012   2.673  0.00896 **
## age         -0.019637   0.011173  -1.758  0.08229 .
## lbph         0.107054   0.058449   1.832  0.07040 .
## svi          0.766157   0.244309   3.136  0.00233 **
## lcp         -0.105474   0.091013  -1.159  0.24964
## gleason      0.045142   0.157465   0.287  0.77503
## pgg45        0.004525   0.004421   1.024  0.30886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7084 on 88 degrees of freedom
## Multiple R-squared:  0.6548, Adjusted R-squared:  0.6234
## F-statistic: 20.86 on 8 and 88 DF,  p-value: < 2.2e-16
```

Thus, we can see that we can remove everything except for lcavol, lweight, and svi. Doing so, we can check our predictions using the same values as above (removing the non-95% significant values) to find our confidence intervals as follows,

```
library(faraway)
regNew <- lm(lpsa ~ lcavol + lweight + svi, data = prostate)
predict(regNew, data.frame( lcavol = 1.44692, lweight = 3.62301, svi = 0), interval = 'confidence')
```

```
##        fit      lwr      upr
## 1 2.372534 2.197274 2.547794
```

We can see that we get a 95% confidence interval of (2.20, 2.55), which is a more narrow CI than our prior values. We would prefer this model, as since the 95% CI is more narrow and since we have removed the non-significant predictors, we can conclude that we have more trust in our model and thus, we the predictions as well.

Problem 4.2

Using the teengamb data, fit a model with gamble as the response and the other variables as predictors. 1. Predict the amount that a male with average (given these data) status, income and verbal score would gamble along with an appropriate 95% CI. 2. Repeat the prediction for a male with maximal values (for this data) of status, income and verbal score. Which CI is wider and why is this result expected? 3. Fit a model with sqrt(gamble) as the response but with the same predictors. Now predict the response and give a 95% prediction interval for the individual in (a). Take care to give your answer in the original units of the response. 4. Repeat the prediction for the model in (c) for a female with status=20, income=1, verbal = 10. Comment on the credibility of the result.

Answer 4.2

1. We can compute these values as follows,

```r
library(faraway)
reg <- lm(gamble ~ sex + status + income + verbal, data = teengamb)
x <- model.matrix(reg)
(x0 <- apply(x,2,function(x) mean(x)))
```

```
## (Intercept)         sex      status      income      verbal
##   1.0000000   0.4042553  45.2340426   4.6419149   6.6595745
```

```r
predict(reg, data.frame(sex = 0, status = 45.2340426, income = 4.6419149, verbal =  6.6595745 ), inte
```

```
##        fit      lwr      upr
## 1 28.24252 18.78277 37.70227
```

We can see that our predicted gambling output is 28.25 pounds per week with a confidence interval of (18.78, 37.70) as desired. 2. We can compute these values as follows,

```r
library(faraway)
reg <- lm(gamble ~ sex + status + income + verbal, data = teengamb)
x <- model.matrix(reg)
(x0 <- apply(x,2,function(x) max(x)))
```

```
## (Intercept)         sex      status      income      verbal
##           1           1          75          15          10
```

```r
predict(reg, data.frame(sex = 0, status = 75, income = 15, verbal =  10 ), interval = 'confidence')
```

```
##        fit      lwr      upr
## 1 71.30794 42.23237 100.3835
```

We can see that our predicted gambling output is 71.31 pounds per week with a confidence interval of (42.23, 100.38) as desired. The wide CI is with our maximum of each of the predictive intervals since We can note that our CI is far larger, which is due to the fact that we are extrapolating from the data set since we are focusing on the maximums of some of our predictor variables. 3. We can compute these values as follows,

```r
library(faraway)
reg <- lm(sqrt(gamble) ~ sex + status + income + verbal, data = teengamb)
predict(reg, data.frame(sex = 0, status = 45.2340426, income = 4.6419149, verbal =  6.6595745 ), inte
```

```
##        fit      lwr      upr
## 1 4.049523 3.180676 4.918371
```

We can see that the our predicted gambling expenditure per week is 16.40 pounds per week with a confidence interval of (10.11, 24.21) by squaring our output values. 4. We can compute these values as follows,

```r
library(faraway)
reg <- lm(sqrt(gamble) ~ sex + status + income + verbal, data = teengamb)
predict(reg, data.frame(sex = 1, status = 20, income = 1, verbal =  10 ), interval = 'confidence')
```

```
##         fit       lwr      upr
## 1 -2.08648 -4.445937 0.272978
```

Using the model to predict with our given inputs, we can see that our results are not very credible since it is impossible to spend negative money as a gambling expenditure (unless you are talking about winning money).

Problem 4.3

The snail dataset contains percentage water content of the tissues of snails grown under three different levels of relative humidity and two different temperatures. 1. Use the command xtabs(water ~ temp + humid, snail)/4 to produce a table of mean water content for each combination of temperature and humidity. Can you use this table to predict the water content for a temperature of 25 $\circ$ C and a humidity of 60%? Explain.

2. Fit a regression model with the water content as the response and the temperature and humidity as predictors. Use this model to predict the water content for a temperature of 25 $\circ$ C and a humidity of 60%? 3. Use this model to predict water content for a temperature of 30 $\circ$ C and a hu midity of 75%? Compare your prediction to the prediction from (a). Discuss the relative merits of these two predictions. 4. The intercept in your model is 52.6%. Give two values of the predictors for which this represents the predicted response. Is your answer unique? Do you think this represents a reasonable prediction? 5. For a temperature of 25 $\circ$ C, what value of humidity would give a predicted response of 80% water content.

Answer 4.3

1. Using the give command, we can see,

```
library(faraway)
xtabs(water ~ temp + humid, snail)/4
```

```
##      humid
## temp    45    75   100
##   20 72.50 81.50 97.00
##   30 69.50 78.25 97.75
```

I would say that we can make a strong prediction, since we can assume that $\frac{69.4-72.5}{10}$ give us a linear coefficient for the impact of each degree C would have on the water content past 20C, and similarly that $\frac{81.50-72.50}{30}$ would give us the impact of each percent increase in humidity would have on the water conent pas 45% humidity, thus giving us an estimation of $72.5 + 15 * \frac{81.50-72.50}{30} + 5 * \frac{69.4-72.5}{10} = 75.45$. 2. We can fit the model and predict the water content for the given values as follows,

```
library(faraway)
reg <- lm(water ~ temp + humid, data = snail)
predict(reg, data.frame(temp = 25, humid = 60 ), interval = 'confidence')
```

```
##        fit      lwr      upr
## 1 76.43681 73.69926 79.17436
```

which gives us prediction of 76.44% water content. 3. We can fit the model and predict the water content for the given values as follows,

```
library(faraway)
reg <- lm(water ~ temp + humid, data = snail)
predict(reg, data.frame(temp = 30, humid = 75 ), interval = 'confidence')
```

```
##        fit     lwr      upr
## 1 82.62248 79.2879 85.95706
```

```
summary(reg)
```

```
##
## Call:
## lm(formula = water ~ temp + humid, data = snail)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.456  -2.915   1.461   3.613   8.749
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 52.61081    6.85346   7.677 1.59e-07 ***
## temp        -0.18333    0.22645  -0.810    0.427
## humid        0.47349    0.05036   9.403 5.63e-09 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.547 on 21 degrees of freedom
## Multiple R-squared:  0.8092, Adjusted R-squared:  0.791
## F-statistic: 44.53 on 2 and 21 DF,  p-value: 2.793e-08
```

We can see that our calculated value for our prediction function does not match up with our prior calculated value, and more specifically, we can see that the earlier calculated value does not even fit into our 95% CI calculated by our prediction model. I would give greater trust to the table of mean water content for each combination of temperature and humidity since this takes the average over all of the 30, 75 pairs, whereas our model accounts for all other points as well when calculating the regression line. 4. Two values of the predictors would be the pair (0,0), and (2.5827196858, 1) that would represent this response. These answers are not unique, and I think that the model provides an accurate prediction for particular values, but not for the two mentioned values. 5. For a temperature of 25C, the humidity that would give a percentage water content of 80% would be 67.5% humidity, which we can calculate since $80 = 25 * -0.18333 + 0.47349x + 52.61081 \implies x = 67.5$ percent humidity.

Problem 4.5

For the fat data used in this chapter, a smaller model using only age, weight, height and abdom was proposed on the grounds that these predictors are either known by the individual or easily measured. 1. Compare this model to the full thirteen-predictor model used earlier in the chapter. Is it justifiable to use the smaller model? 2. Compute a 95% prediction interval for median predictor values and compare to the results to the interval for the full model. Do the intervals differ by a practically important amount? 3. For the smaller model, examine all the observations from case numbers 25 to 50. Which two observations seem particularly anomalous? 4. Recompute the 95% prediction interval for median predictor values after these two anomalous cases have been excluded from the data. Did this make much difference to the outcome?

Answer 4.5

1. We can compare the two models by using an anova test as follows,

```r
library(faraway)
regFull <- lm(brozek ~ age + weight + height + neck + chest + abdom + hip + thigh + knee + ankle + bi
regReduce <- lm(brozek ~ age + weight + height + abdom, data=fat)
anova(regReduce, regFull)
```

```
## Analysis of Variance Table
##
## Model 1: brozek ~ age + weight + height + abdom
## Model 2: brozek ~ age + weight + height + neck + chest + abdom + hip +
##     thigh + knee + ankle + biceps + forearm + wrist
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    247 4205.0
## 2    238 3785.1  9     419.9 2.9336 0.002558 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing the two models together, we can see from the anova test that since our p-value is less than 0.05, that we can reject the null hypothesis in favor of the full model. 2. We can compute the 95% prediction interval for the two models for the median predictor values as follow,

```r
library(faraway)
regFull <- lm(brozek ~ age + weight + height + neck + chest + abdom + hip + thigh + knee + ankle + bi
regReduce <- lm(brozek ~ age + weight + height + abdom, data=fat)
```

```r
x <- model.matrix(regFull)
(xFull <- apply(x,2,median))
```

```
## (Intercept)         age      weight      height        neck       chest
##        1.00       43.00      176.50       70.00       38.00       99.65
##       abdom         hip       thigh        knee       ankle      biceps
##       90.95       99.30       59.00       38.50       22.80       32.05
##     forearm       wrist
##       28.70       18.30
```

```r
y <- model.matrix(regReduce)
(xReduce <- apply(y,2,median))
```

```
## (Intercept)         age      weight      height       abdom
##        1.00       43.00      176.50       70.00       90.95
```

```r
predict(regFull, data.frame(t(xFull)), interval = 'confidence')
```

```
##        fit      lwr      upr
## 1 17.49322 16.94426 18.04219
```

```r
predict(regReduce, data.frame(t(xReduce)), interval = 'confidence')
```

```
##        fit      lwr      upr
## 1 17.84028 17.31621 18.36435
```

After computing the two confidence intervals, we can see that the two confidence intervals do not differ by a practically important ammount. 3. We can view all of the observations as follows,

```r
library(faraway)
regFull <- lm(brozek ~ age + weight + height + neck + chest + abdom + hip + thigh + knee + ankle + bic
regReduce <- lm(brozek ~ age + weight + height + abdom, data=fat)
y <- model.matrix(regReduce)
y
```

```
##     (Intercept) age weight height abdom
## 1             1  23 154.25  67.75  85.2
## 2             1  22 173.25  72.25  83.0
## 3             1  22 154.00  66.25  87.9
## 4             1  26 184.75  72.25  86.4
## 5             1  24 184.25  71.25 100.0
## 6             1  24 210.25  74.75  94.4
## 7             1  26 181.00  69.75  90.7
## 8             1  25 176.00  72.50  88.5
## 9             1  25 191.00  74.00  82.5
## 10            1  23 198.25  73.50  88.6
## 11            1  26 186.25  74.50  83.6
## 12            1  27 216.00  76.00  90.9
## 13            1  32 180.50  69.50  91.6
## 14            1  30 205.25  71.25 101.8
## 15            1  35 187.75  69.50  96.4
## 16            1  35 162.75  66.00  92.8
## 17            1  34 195.75  71.00  96.4
## 18            1  32 209.25  71.00  97.5
## 19            1  28 183.75  67.75  89.6
## 20            1  33 211.75  73.50 100.5
## 21            1  28 179.00  68.00  95.9
```

```
## 22             1   28 200.50  69.75  98.8
## 23             1   31 140.25  68.25  76.4
## 24             1   32 148.75  70.00  80.0
## 25             1   28 151.25  67.75  76.3
## 26             1   27 159.25  71.50  79.7
## 27             1   34 131.50  67.50  74.6
## 28             1   31 148.00  67.50  88.7
## 29             1   27 133.25  64.75  73.9
## 30             1   29 160.75  69.00  83.5
## 31             1   32 182.00  73.75  88.7
## 32             1   29 160.25  71.25  84.5
## 33             1   27 168.00  71.25  79.1
## 34             1   41 218.50  71.00 100.5
## 35             1   41 247.25  73.50 115.6
## 36             1   49 191.75  65.00 113.1
## 37             1   40 202.25  70.00 100.9
## 38             1   50 196.75  68.25  98.8
## 39             1   46 363.15  72.25 148.1
## 40             1   50 203.00  67.00 108.1
## 41             1   45 262.75  68.75 126.2
## 42             1   44 205.00  29.50 104.3
## 43             1   48 217.00  70.00 111.2
## 44             1   41 212.00  71.50 104.3
## 45             1   39 125.25  68.00  76.0
## 46             1   43 164.25  73.25  81.5
## 47             1   40 133.50  67.50  73.7
## 48             1   39 148.50  71.25  79.5
## 49             1   45 135.75  68.50  83.4
## 50             1   47 127.50  66.75  70.4
## 51             1   47 158.25  72.25  86.7
## 52             1   40 139.25  69.00  77.9
## 53             1   51 137.25  67.75  82.0
## 54             1   49 152.75  73.50  79.6
## 55             1   42 136.25  67.50  77.6
## 56             1   54 198.00  72.00 100.0
## 57             1   58 181.50  68.00  99.8
## 58             1   62 201.25  69.50 104.2
## 59             1   54 202.50  70.75 105.3
## 60             1   61 179.75  65.75  98.3
## 61             1   62 216.00  73.25 104.8
## 62             1   56 178.75  68.50  94.7
## 63             1   54 193.25  70.25 102.4
## 64             1   61 178.00  67.00  99.7
## 65             1   57 205.50  70.00 105.5
## 66             1   55 183.50  67.50 100.3
## 67             1   54 151.50  70.75  83.9
## 68             1   55 154.75  71.50  86.6
## 69             1   54 155.25  69.25  78.4
## 70             1   55 156.75  71.50  84.6
## 71             1   62 167.50  71.50  91.5
## 72             1   55 146.75  68.75  82.8
## 73             1   56 160.75  73.75  82.9
## 74             1   55 125.00  64.00  76.0
## 75             1   61 143.00  65.75  83.3
```

```
## 76              1  61 148.25  67.50  81.8
## 77              1  57 162.50  69.50  78.8
## 78              1  69 177.75  68.50  95.0
## 79              1  81 161.25  70.25  95.4
## 80              1  66 171.25  69.25  98.6
## 81              1  67 163.75  67.75  95.8
## 82              1  64 150.25  67.25  89.0
## 83              1  64 190.25  72.75  97.8
## 84              1  70 170.75  70.00  94.9
## 85              1  72 168.00  69.25  99.8
## 86              1  67 167.00  67.50  89.7
## 87              1  72 157.75  67.25  88.1
## 88              1  64 160.00  65.75  90.9
## 89              1  46 176.75  72.50  86.0
## 90              1  48 176.00  73.00  86.5
## 91              1  46 177.00  70.00  95.6
## 92              1  44 179.75  69.50  93.2
## 93              1  47 165.25  70.50  83.1
## 94              1  46 192.50  71.75  97.5
## 95              1  47 184.25  74.50  88.8
## 96              1  53 224.50  77.75  99.2
## 97              1  38 188.75  73.25  91.6
## 98              1  50 162.50  66.50  86.7
## 99              1  46 156.50  68.25  88.2
## 100             1  47 197.00  72.00  94.0
## 101             1  49 198.50  73.50  95.0
## 102             1  48 173.75  72.00  92.0
## 103             1  41 172.75  71.25  89.2
## 104             1  49 196.75  73.75  95.5
## 105             1  43 177.00  69.25  98.6
## 106             1  43 165.50  68.50  87.3
## 107             1  43 200.25  73.50 102.8
## 108             1  52 203.25  74.25 101.6
## 109             1  43 194.00  75.50  88.7
## 110             1  40 168.50  69.25  92.3
## 111             1  43 170.75  68.50  90.6
## 112             1  43 183.25  70.00 105.0
## 113             1  47 178.25  70.00  95.0
## 114             1  42 163.00  70.25  89.6
## 115             1  48 175.25  71.75  92.4
## 116             1  40 158.00  69.25  86.6
## 117             1  48 177.25  72.75  90.0
## 118             1  51 179.00  72.00  90.0
## 119             1  40 191.00  74.00  92.4
## 120             1  44 187.50  72.25  87.5
## 121             1  52 206.50  74.50  99.2
## 122             1  44 185.25  71.50  98.1
## 123             1  40 160.25  68.75  83.3
## 124             1  47 151.50  66.75  86.1
## 125             1  50 161.00  66.50  84.1
## 126             1  46 167.00  67.00  89.9
## 127             1  42 177.50  68.75  92.1
## 128             1  43 152.25  67.75  78.0
## 129             1  40 192.25  73.25  93.5
```

```
## 130            1  42 165.25  69.75  87.0
## 131            1  49 171.75  71.50  90.1
## 132            1  40 171.25  70.50  90.3
## 133            1  47 197.00  73.25  99.8
## 134            1  50 157.00  66.75  89.4
## 135            1  41 168.25  69.50  87.2
## 136            1  44 186.00  69.75 101.1
## 137            1  39 166.75  70.75  86.1
## 138            1  43 187.75  74.00  98.6
## 139            1  40 168.25  71.25  88.5
## 140            1  49 212.75  75.00 106.6
## 141            1  40 176.75  71.00  93.1
## 142            1  40 173.25  69.50  93.0
## 143            1  52 167.00  67.75  91.0
## 144            1  23 159.75  72.25  77.1
## 145            1  23 188.15  77.50  85.3
## 146            1  24 156.00  70.75  81.9
## 147            1  24 208.50  72.75  99.1
## 148            1  25 206.50  69.75 100.5
## 149            1  25 143.75  72.50  76.5
## 150            1  26 223.00  70.25 106.8
## 151            1  26 152.25  69.00  77.6
## 152            1  26 241.75  74.50 102.9
## 153            1  27 146.00  72.25  72.8
## 154            1  27 156.75  67.25  88.2
## 155            1  27 200.25  73.50 100.1
## 156            1  28 171.50  75.25  83.5
## 157            1  28 205.75  69.00 105.0
## 158            1  28 182.50  72.25  90.8
## 159            1  30 136.50  68.75  76.6
## 160            1  31 177.25  71.50  92.4
## 161            1  31 151.25  72.25  81.2
## 162            1  33 196.00  73.00  95.6
## 163            1  33 184.25  68.75  92.1
## 164            1  34 140.00  70.50  83.4
## 165            1  34 218.75  72.00 106.0
## 166            1  35 217.00  73.75  95.1
## 167            1  35 166.25  68.00  90.4
## 168            1  35 224.75  72.25 100.4
## 169            1  35 228.25  69.50 115.9
## 170            1  35 172.75  69.50  90.8
## 171            1  35 152.25  67.75  81.9
## 172            1  35 125.75  65.50  75.0
## 173            1  35 177.25  71.00  90.3
## 174            1  36 176.25  71.50  90.3
## 175            1  36 226.75  71.75 108.8
## 176            1  37 145.25  69.25  79.4
## 177            1  37 151.00  67.00  83.2
## 178            1  37 241.25  71.50 110.3
## 179            1  38 187.25  69.25  92.7
## 180            1  39 234.75  74.50 104.5
## 181            1  39 219.25  74.25 104.6
## 182            1  40 118.50  68.00  69.4
## 183            1  40 145.75  67.25  83.6
```

```
## 184           1  40 159.25  69.75  86.8
## 185           1  40 170.50  74.25  90.4
## 186           1  40 167.50  71.50  83.7
## 187           1  41 232.75  74.25 109.3
## 188           1  41 210.50  72.00  98.9
## 189           1  41 202.25  72.50  98.0
## 190           1  41 185.00  68.25 101.2
## 191           1  41 153.00  69.25  80.6
## 192           1  42 244.25  76.00 113.7
## 193           1  42 193.50  70.50  94.1
## 194           1  42 224.75  74.75 105.7
## 195           1  42 162.75  72.75  85.6
## 196           1  42 180.00  68.25  96.6
## 197           1  42 156.25  69.00  86.0
## 198           1  42 168.00  71.50  89.7
## 199           1  42 167.25  72.75  78.0
## 200           1  43 170.75  67.50  89.7
## 201           1  43 178.25  70.25  89.2
## 202           1  43 150.00  69.25  85.7
## 203           1  43 200.50  71.50 103.1
## 204           1  44 184.00  74.00  89.1
## 205           1  44 223.00  69.75 113.9
## 206           1  44 208.75  73.00  96.3
## 207           1  44 166.00  65.50  93.9
## 208           1  47 195.00  72.50 101.3
## 209           1  47 160.50  70.25  83.9
## 210           1  47 159.75  70.75  84.4
## 211           1  49 140.50  68.00  79.4
## 212           1  49 216.25  74.50 104.0
## 213           1  49 168.25  71.75  89.7
## 214           1  50 194.75  70.75  97.6
## 215           1  50 172.75  73.00  87.6
## 216           1  51 219.00  64.00 122.1
## 217           1  51 149.25  69.75  81.1
## 218           1  51 154.50  70.00  81.5
## 219           1  52 199.25  71.75 100.0
## 220           1  53 154.50  69.25  88.7
## 221           1  54 153.25  70.50  91.8
## 222           1  54 230.00  72.25 110.4
## 223           1  54 161.75  67.50  87.6
## 224           1  55 142.25  67.25  82.8
## 225           1  55 179.75  68.75  95.3
## 226           1  55 126.50  66.75  78.2
## 227           1  55 169.50  68.25  91.1
## 228           1  55 198.50  74.25  96.7
## 229           1  56 174.50  69.50  89.4
## 230           1  56 167.75  68.50  93.0
## 231           1  57 147.75  65.75  86.4
## 232           1  57 182.25  71.75  96.7
## 233           1  58 175.50  71.50  88.1
## 234           1  58 161.75  67.25  94.9
## 235           1  60 157.75  67.50  93.3
## 236           1  62 168.75  67.50  95.6
## 237           1  62 191.50  72.25  98.2
```

```
## 238           1  63 219.15  69.50 113.8
## 239           1  64 155.25  69.50  82.8
## 240           1  65 189.75  65.75 100.5
## 241           1  65 127.50  65.75  79.7
## 242           1  65 224.50  68.25 118.0
## 243           1  66 234.25  72.00 109.0
## 244           1  67 227.75  72.75 113.4
## 245           1  67 199.50  68.50 106.1
## 246           1  68 155.50  69.25  84.3
## 247           1  69 215.50  70.50 107.6
## 248           1  70 134.25  67.00  83.6
## 249           1  72 201.00  69.75 105.0
## 250           1  72 186.75  66.00 111.5
## 251           1  72 190.75  70.50 101.3
## 252           1  74 207.50  70.00 108.5
## attr(,"assign")
## [1] 0 1 2 3 4
```

The observations that seem anamolous are observations 39 and 42. 4. We can remove the observations as follows,

```
library(faraway)
regFull <- lm(brozek ~ age + weight + height + neck + chest + abdom + hip + thigh + knee + ankle + bic
y <- model.matrix(regReduce)
(xReduce <- apply(y,2,median))
```

```
## (Intercept)         age      weight      height       abdom
##        1.00       43.00      176.50       70.00       90.95
```

```
predict(regReduce, data.frame(t(xReduce)), interval = 'confidence')
```

```
##        fit      lwr      upr
## 1 17.84028 17.31621 18.36435
```

We can note that removing these values does not really change our outcomes with our CI in a meaningful way for the reduced model.