Homework 8

Alex Soupir

November 7, 2019

Packages: HSAUR3, quantreg, TH.data

Collaborators:

Chapter 12

Answer all questions specified on the problem and include a discussion on how your results answered/addressed the question.

Submit your **.rmd** file with the knitted **PDF** (or knitted Word Document saved as a PDF). If you are having trouble with .rmd, let us know and we will help you, but both the .rmd and the PDF are required.

This file can be used as a skeleton document for your code/write up. Please follow the instructions found under Content for Formatting and Guidelines. No code should be in your PDF write-up unless stated otherwise.

For any question asking for plots/graphs, please do as the question asks as well as do the same but using the respective commands in the GGPLOT2 library. (So if the question asks for one plot, your results should have two plots. One produced using the given R-function and one produced from the GGPLOT2 equivalent). This doesn't apply to questions that don't specifically ask for a plot, however I still would encourage you to produce both.

You do not need to include the above statements.

Please do the following problems from the text book R Handbook and stated.

1. Consider the {**clouds**} data from the {**HSAUR3**} package

    a) Review the linear model fitted to this data in Chapter 6 of the text book and report the model and findings.

```
##
## Call:
## lm(formula = clouds_formula, data = clouds)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5259 -1.1486 -0.2704  1.0401  4.3913
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -0.34624    2.78773  -0.124  0.90306
## seedingyes                     15.68293    4.44627   3.527  0.00372 **
## time                           -0.04497    0.02505  -1.795  0.09590 .
## seedingno:sne                   0.41981    0.84453   0.497  0.62742
## seedingyes:sne                 -2.77738    0.92837  -2.992  0.01040 *
## seedingno:cloudcover            0.38786    0.21786   1.780  0.09839 .
## seedingyes:cloudcover          -0.09839    0.11029  -0.892  0.38854
## seedingno:prewetness            4.10834    3.60101   1.141  0.27450
## seedingyes:prewetness           1.55127    2.69287   0.576  0.57441
## seedingno:echomotionstationary  3.15281    1.93253   1.631  0.12677
## seedingyes:echomotionstationary 2.59060    1.81726   1.426  0.17757
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.205 on 13 degrees of freedom
## Multiple R-squared:  0.7158, Adjusted R-squared:  0.4972
## F-statistic: 3.274 on 10 and 13 DF,  p-value: 0.02431
```
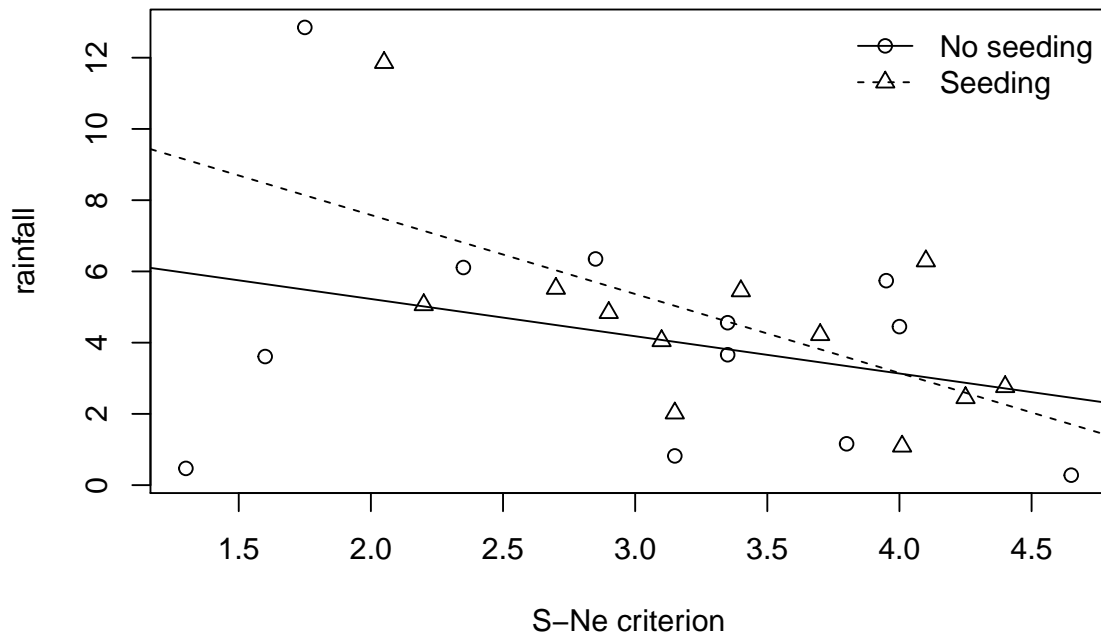

Above is the model summary as well as the plot of the relationship between the suitability criterion and rainfall with and without seeding. The summary shows that a model with seeding is better able to predict the rainfall than a model with no cloud seeding. Seeding clouds and suitability criterion also shows significance to the model meaning that the effect of seeding clouds is not standalone but also depends on the suitability criterion for determining the rainfall.

b) Fit a median regression model.

```
##                                    coefficients        lower bd        upper bd
## (Intercept)                         -0.39510353   -2.032259e+00    1.234196e+01
## seedingyes                           9.28416250    4.632247e+00    2.478669e+01
## time                                -0.02682160   -7.150623e-02   -2.068740e-02
## seedingno:sne                        0.36860476   -1.090559e+00    1.196003e+00
## seedingyes:sne                      -1.33267160   -6.025488e+00   -1.177594e+00
## seedingno:cloudcover                 0.20691306    1.818597e-02    1.043587e+00
## seedingyes:cloudcover               -0.06071068   -3.426312e-01    2.468352e-01
## seedingno:prewetness                 5.22263667   -9.255066e+00    1.156672e+01
## seedingyes:prewetness                2.01808261   -1.797693e+308   1.797693e+308
## seedingno:echomotionstationary       2.13502276   -4.986951e-01    1.103820e+01
## seedingyes:echomotionstationary      2.78255068   -1.797693e+308   1.797693e+308
```
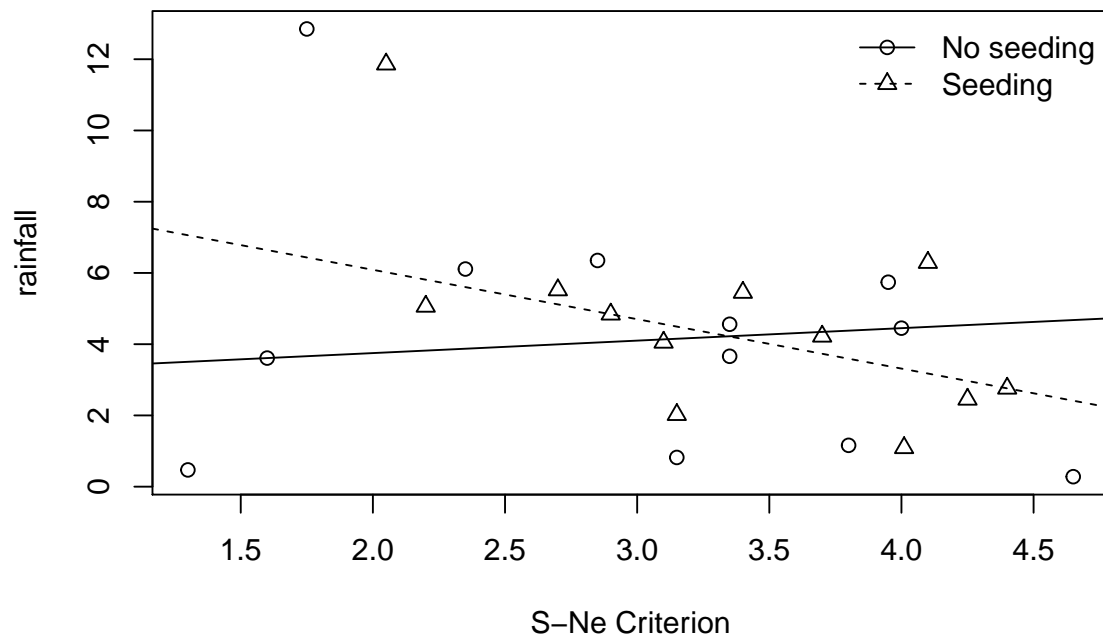
A median regression model was fit using *tau = 0.5* on the same formula from the book that was used in *Part A*. The plot was created the same was as *Part A* as well but using *rq* instead of *lm* and adding *tau*.

c) Compare the two results.

```
##
## Mean squared error of linear model on not seeding data:

## [1] 10.22228

##
## Mean squared error of linear model on seeding data:

## [1] 4.292417

##
## Mean squared error of median model on not seeding data:

## [1] 12.24009

##
## Mean squared error of median model on seeding data:

## [1] 4.834842

##
## Coefficients of linear and median models using clouds_formula from book:

##                                  Linear        Median
## (Intercept)                  -0.34624093   -0.39510353
## seedingyes                   15.68293481    9.28416250
## time                         -0.04497427   -0.02682160
```

```
## seedingno:sne                     0.41981393  0.36860476
## seedingyes:sne                    -2.77737613 -1.33267160
## seedingno:cloudcover              0.38786207  0.20691306
## seedingyes:cloudcover             -0.09839285 -0.06071068
## seedingno:prewetness              4.10834188  5.22263667
## seedingyes:prewetness             1.55127493  2.01808261
## seedingno:echomotionstationary    3.15281358  2.13502276
## seedingyes:echomotionstationary   2.59059513  2.78255068
```

Comparing the plots produced in parts *A* and *B*, the largest difference in the change in the *No seeding* line where it went from having a negative slope when using the linear model to having a positive slope using the median regression. This could be due to the extreme point at about (1.7, 13) doesn't have as much 'pull' in the model anymore. The slope of the seeding model from linear to median increased slightly. The linear model produced a lower MSE for both the data created with cloud seeding and without cloud seeding. Additionally, the coefficients for both models were placed into a table similarly to in lecture. The seeding coefficient for when clouds are seeded is rather different between the linear model and the median model.

2. Reanalyze the {**bodyfat**} data from the {**TH.data**} package.

   a) Compare the regression tree approach from chapter 9 of the textbook to median regression and summarize the different findings.
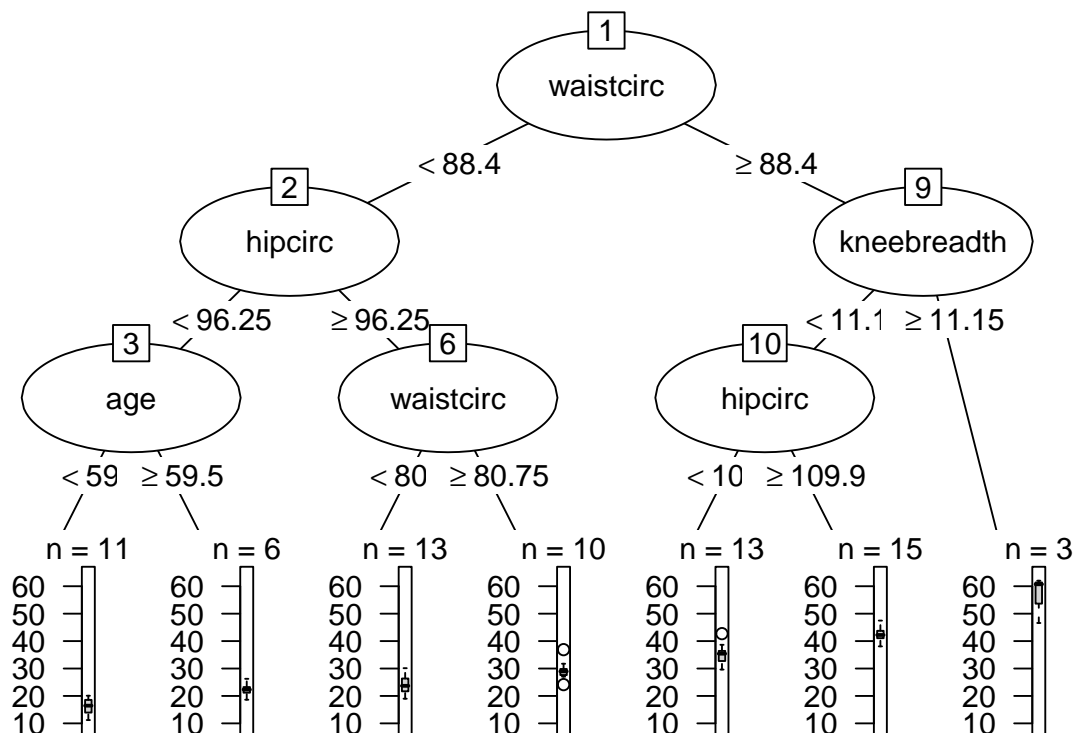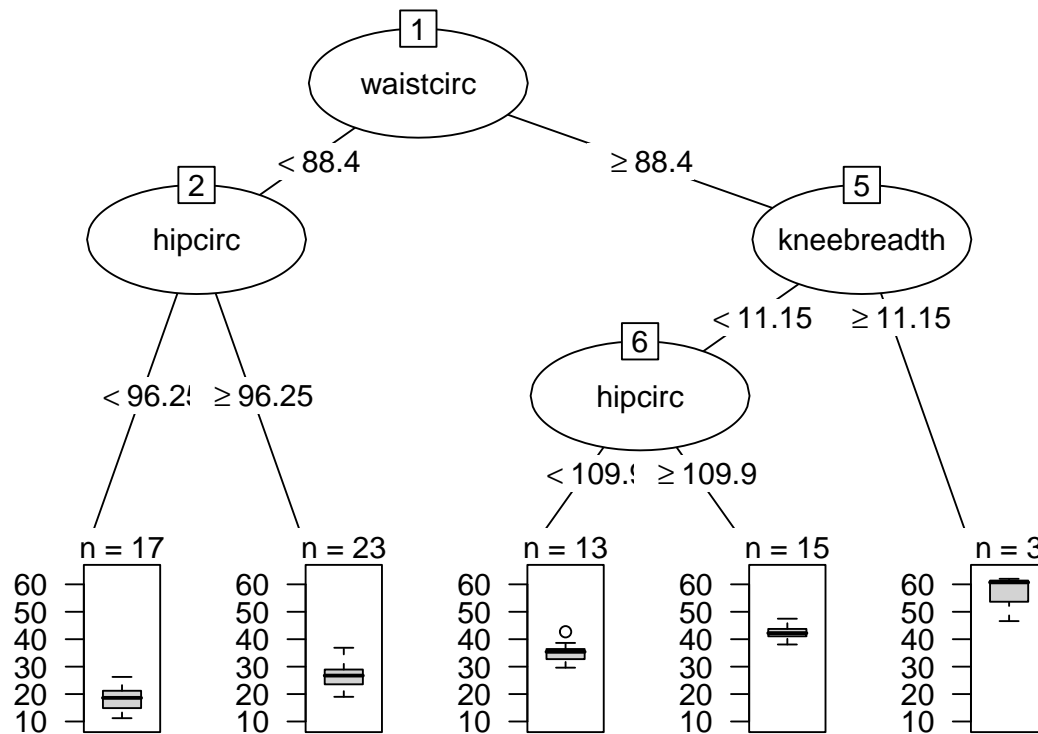
```
## MSE of tree using Chapter 9 equation without pruning:

## [1] 10.1705

## MSE of tree using Chapter 9 equation with pruning:

## [1] 14.575

## MSE of median regression of original equation:

## [1] 15.0245
```

**Comparing the regression tree's, both unpruned and pruned, shows that the unpruned
tree has lower mean squared error in the predicted DEXfat. A median regression was
also created and the mean squared error was calculated. The mean squared error of the
median regression was greater than the pruned tree putting it at the worst approach of
the 3 at predicting the DEXfat response variable.**

   b) Choose one independent variable. For the relationship between this variable and DEXfat, create
linear regression quantile models for the 25%, 50% and 75% quantiles. Plot DEXfat vs that
independent variable and plot the lines from the models on the graph.

3. Consider {**db**} data from the lecture notes (package {**gamlss.data**}). Refit the additive quantile
regression models presented ({**rqssmod**}) with varying values of λ (lambda) in {**qss**}. How do the
estimated quantile curves change?

4. Read the paper by Koenker and Hallock (2001), posted on D2L. Write a one page summary of the
paper. This should include but not be limited to introduction, motivation, case study considered and
findings.

Resources Used:

- rdocumentation.org

- theanalysisfactor.com
-