Lab Exercise 1 - Problem 3

Spencer Moon

10/15/2017

Before building the regression model, the data was loaded with the following code:

```
library(tidyverse)

labdata <- read_tsv("labdata.txt")
```

Part A

Below is the linear model for the lab data:

```
reg <- lm(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8, labdata)
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8, data = labdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.7138  -7.3129  -0.1718   7.4281  23.8909
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.58565    5.10223   3.447 0.000629 ***
## x1           1.91936    0.05492  34.951  < 2e-16 ***
## x2           0.89747    0.08389  10.699  < 2e-16 ***
## x3           1.07895    0.08370  12.890  < 2e-16 ***
## x4           0.23834    0.08759   2.721 0.006798 **
## x5           0.10141    0.03725   2.723 0.006766 **
## x6           0.29608    0.15153   1.954 0.051421 .
## x7          -0.06268    0.15824  -0.396 0.692262
## x8          -0.01515    0.15846  -0.096 0.923860
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.01 on 391 degrees of freedom
## Multiple R-squared:  0.8113, Adjusted R-squared:  0.8074
## F-statistic: 210.1 on 8 and 391 DF,  p-value: < 2.2e-16
```

Part B

```
attach(labdata)
plot(x1, y)
```
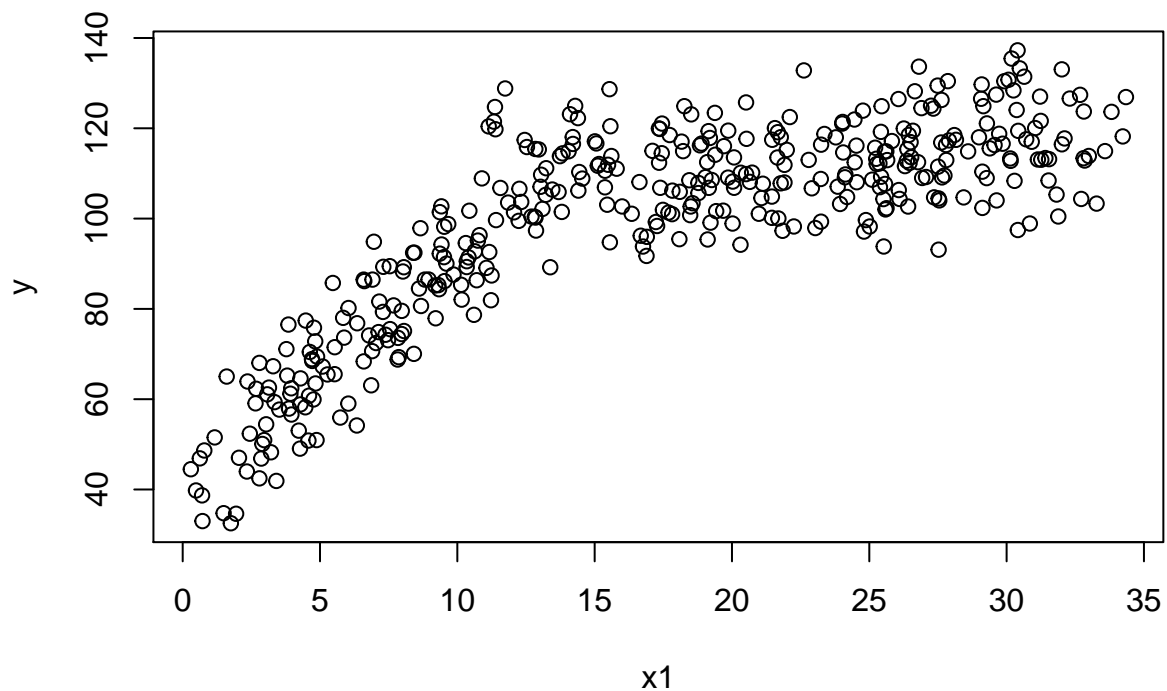
```
#plot(x2, y)
#plot(x3, y)
#plot(x4, y)
#plot(x5, y)
#plot(x6, y)
#plot(x7, y)
#plot(x8, y)
```

Variable x1 can be used in a piecewise regression model as the scatter plot of x1 against y clearly shows a kink.

Part C

```
avg = mean(x1)
avg
```

```
## [1] 17.19417
```

```
reg.piece <- lm(y ~ (x1<avg)*x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8, labdata)
summary(reg.piece)
```

```
##
## Call:
## lm(formula = y ~ (x1 < avg) * x1 + x2 + x3 + x4 + x5 + x6 + x7 +
##     x8, data = labdata)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.3914  -1.3793  -0.1569   1.3062  14.5014
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      61.403705   2.254025  27.242   <2e-16 ***
## x1 < avgTRUE    -57.094813   1.444014 -39.539   <2e-16 ***
## x1                0.517863   0.051891   9.980   <2e-16 ***
## x2                0.989106   0.029237  33.831   <2e-16 ***
## x3                1.032202   0.029060  35.520   <2e-16 ***
## x4                0.018861   0.030815   0.612    0.541
## x5               -0.017325   0.013135  -1.319    0.188
## x6               -0.006076   0.052914  -0.115    0.909
## x7               -0.053892   0.054904  -0.982    0.327
## x8               -0.038638   0.055426  -0.697    0.486
## x1 < avgTRUE:x1   4.097539   0.078416  52.254   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.475 on 389 degrees of freedom
## Multiple R-squared:  0.9774, Adjusted R-squared:  0.9768
## F-statistic:  1682 on 10 and 389 DF,  p-value: < 2.2e-16
```

Above summary shows that the piecewise regression model of the lab data is indeed better because the R-squared value is signficantly greater than the original regression model.