

Week 11 Assignment

Ben Arancibia

November 15, 2014

Using the `lm` function, perform regression analysis and measure independent variables on two datasets.

First Data Set The first data set is heart rate. First create the data set.

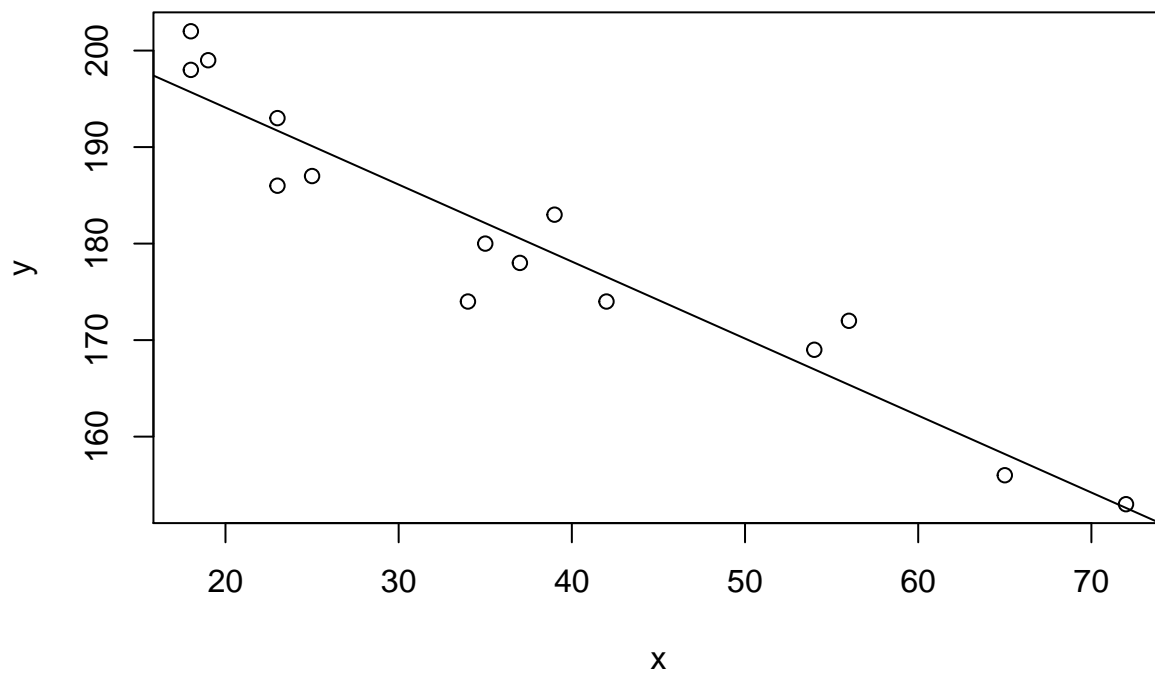
```
x = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37)
y = c(202,186,187,180,156,169,174,172,153,199,193,174,198,183,178)
```

Plot `x` and `y` with regression line and basic values of regression analysis

```
plot(x,y)
lm_age = lm(y ~ x)
lm_age
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##    210.0485     -0.7977
```

```
abline(lm_age)
```



```
summary(lm_age)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.9258 -2.5383  0.3879  3.1867  6.6242
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 210.04846    2.86694   73.27 < 2e-16 ***
## x           -0.79773    0.06996  -11.40 3.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.578 on 13 degrees of freedom
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9021
## F-statistic: 130 on 1 and 13 DF, p-value: 3.848e-08
```

Find that the resulting equation is more like this:

$$MaxHR = -0.7977 + 210.0485$$

As you can see in `summary(lm_age)` you have a hypothesis test calculated by R.

```
summary(lm_age)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.9258 -2.5383  0.3879  3.1867  6.6242
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 210.04846    2.86694   73.27 < 2e-16 ***
## x           -0.79773    0.06996  -11.40 3.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.578 on 13 degrees of freedom
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9021
## F-statistic: 130 on 1 and 13 DF, p-value: 3.848e-08
```

It does not look like it is significant.

```
es = residuals(lm_age)
b1 = (coef(lm_age))[['x']]
s = sqrt(sum(es^2) / (13)) #plugging in n-2 instead of 13 gives me an error
```

```
SE = s/sqrt(sum((x-mean(x))^2))
t = (b1 - (-1))/SE
pt(t, 13, lower.tail=FALSE)
```

```
## [1] 0.006310157
```

Not significant.

Auto Data Set

Perform a linear regression analysis using mpg as the dependent variable and the other 4 as independent variables. First import the data.

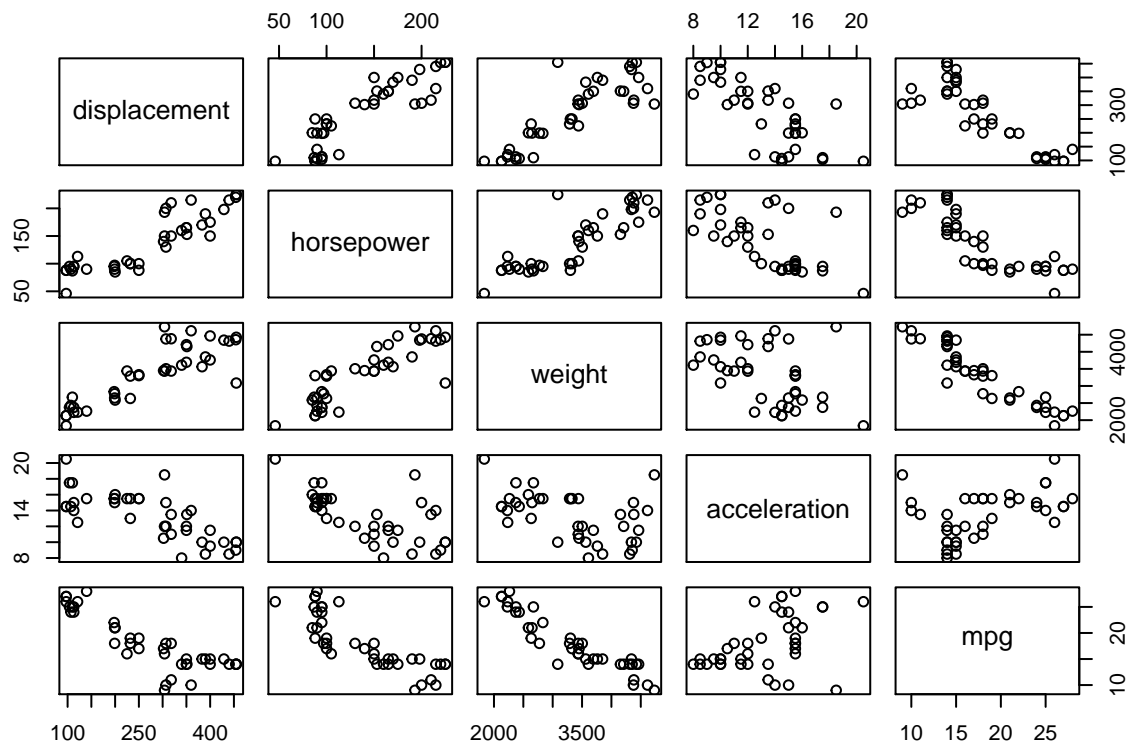
```
data <- read.table("/Users/bcarancibia/CUNY_IS_605/assign11/auto-mpg.data")
names(data) <- c("displacement", "horsepower", "weight", "acceleration", "mpg")
```

Based on the first take a random 40 points from the data set.

```
sub_new <- data[1:40,]
```

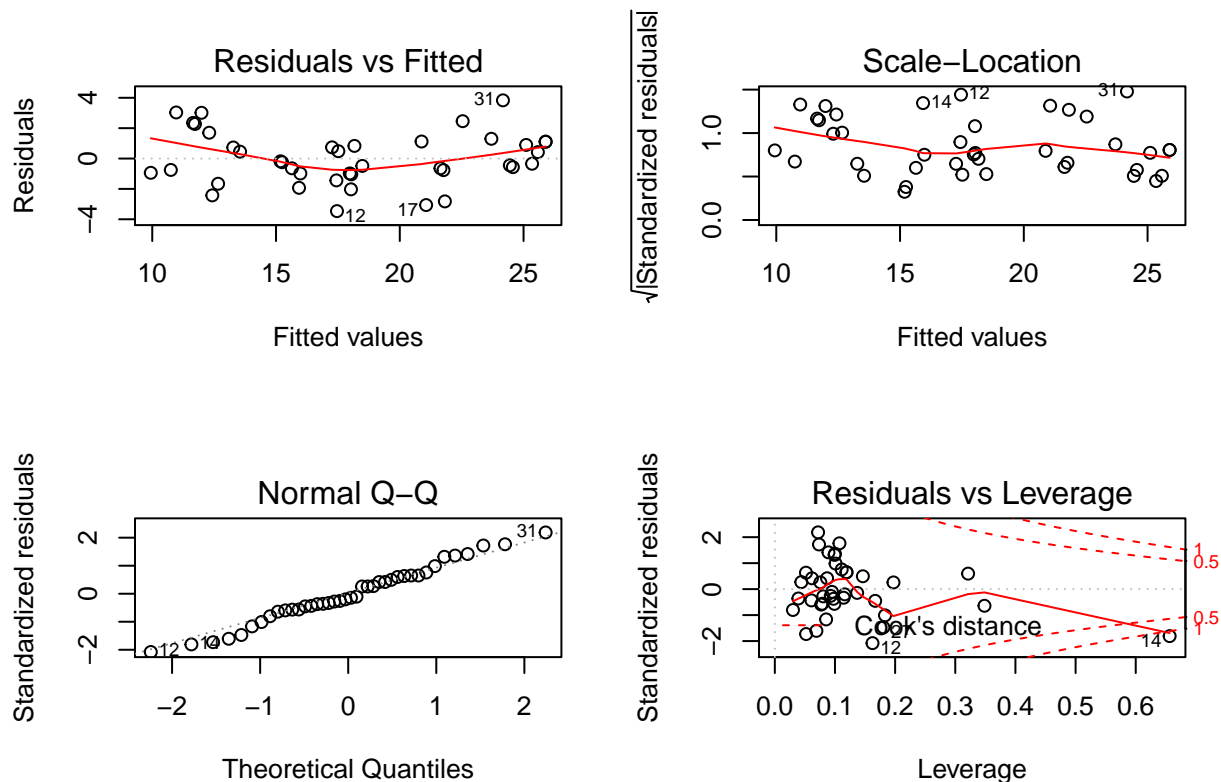
Plot mpg vs the four other variables for the subset.

```
pairs(sub_new)
```



```
auto_sub <- lm(sub_new$mpg ~ sub_new$displacement + sub_new$horsepower + sub_new$weight +
               sub_new$acceleration, data=sub_new)
```

```
layout(matrix(c(1,2,3,4),2,2))
plot(auto_sub)
```



```
summary(auto_sub)
```

```
##
## Call:
## lm(formula = sub_new$mpg ~ sub_new$displacement + sub_new$horsepower +
##     sub_new$weight + sub_new$acceleration, data = sub_new)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4627 -0.9808 -0.2936  1.1086  3.8319
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    41.6275702   2.8843945   14.432 2.65e-16 ***
## sub_new$displacement -0.0197082   0.0076703   -2.569 0.014606 *
## sub_new$horsepower  -0.0073270   0.0141243   -0.519 0.607200
## sub_new$weight      -0.0037062   0.0008949   -4.141 0.000207 ***
## sub_new$acceleration -0.3645195   0.1757330   -2.074 0.045472 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.82 on 35 degrees of freedom
## Multiple R-squared:  0.8933, Adjusted R-squared:  0.8811
## F-statistic: 73.22 on 4 and 35 DF, p-value: < 2.2e-16
```

The final regression fit looks to be:

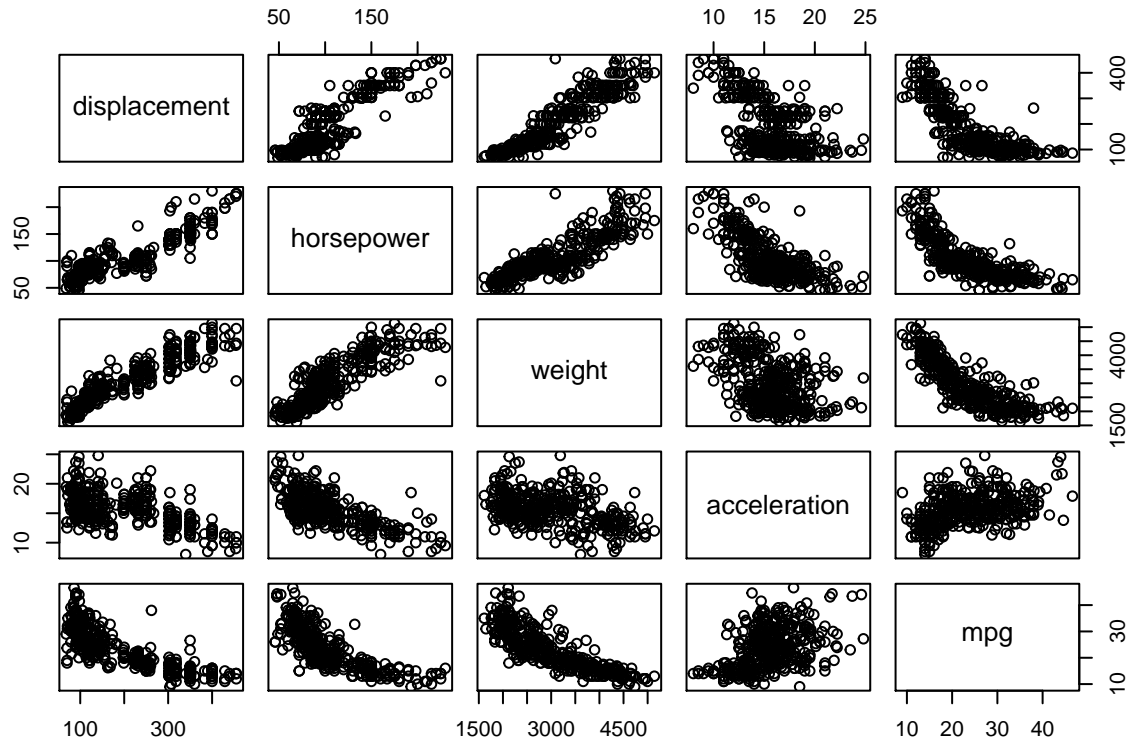
has a significant impact on mpg

The corresponding significance levels are:

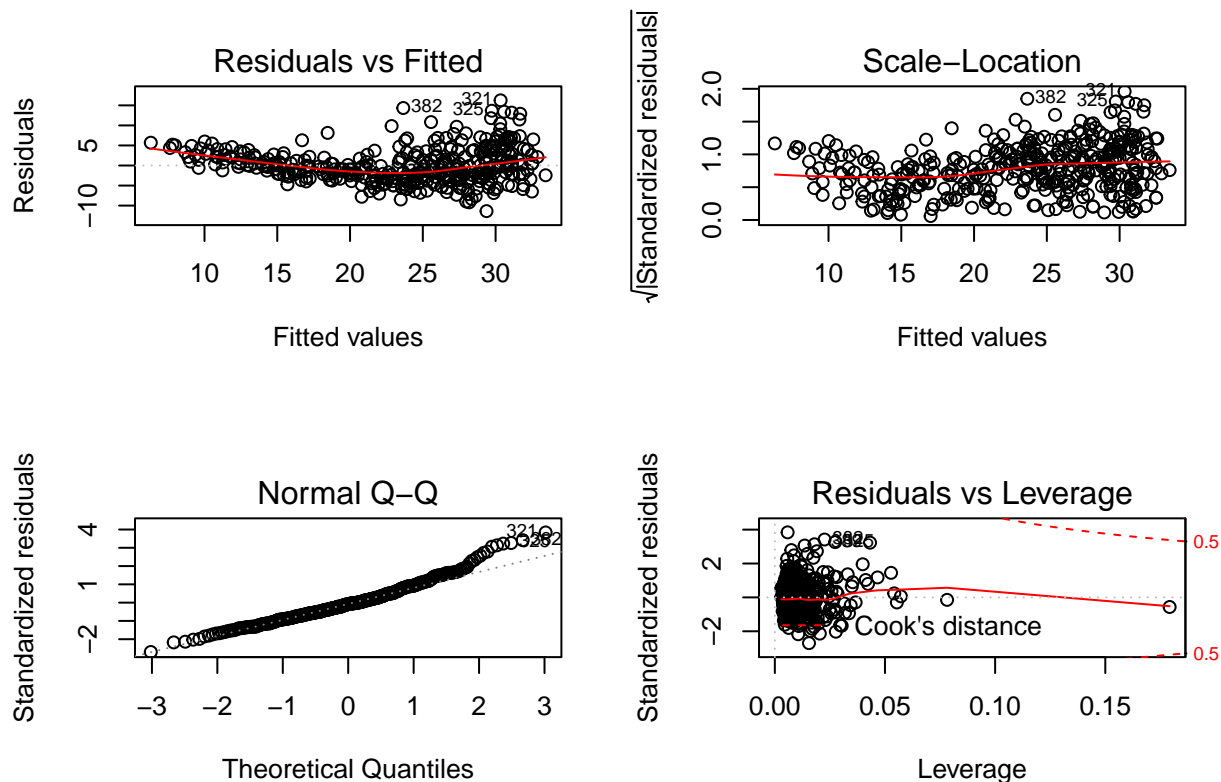
Standard errors for each of the coefficients:

Plot mpg vs the four other variables.

```
pairs(data)
```



```
auto <- lm(data$mpg ~ data$displacement + data$horsepower  
           + data$weight+ data$acceleration, data=data)  
  
layout(matrix(c(1,2,3,4),2,2))  
plot(auto)
```



```
summary(auto)
```

```
##
## Call:
## lm(formula = data$mpg ~ data$displacement + data$horsepower +
##     data$weight + data$acceleration, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.378  -2.793  -0.333   2.193  16.256
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   45.2511397   2.4560447  18.424 < 2e-16 ***
## data$displacement -0.0060009   0.0067093  -0.894  0.37166
## data$horsepower  -0.0436077   0.0165735  -2.631  0.00885 **
## data$weight      -0.0052805   0.0008109  -6.512  2.3e-10 ***
## data$acceleration -0.0231480   0.1256012  -0.184  0.85388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.247 on 387 degrees of freedom
## Multiple R-squared:  0.707, Adjusted R-squared:  0.704
## F-statistic: 233.4 on 4 and 387 DF, p-value: < 2.2e-16
```

The final regression fit looks to be:

has a significant impact on mpg

The corresponding significance levels are:
Standard errors for each of the coefficients:
Do it using 95% confidence intervals.