Linear Regression IS605 - Assignment 11

Sharon Morris

4/9/2017

Using R's lm function, perform regression analysis and measure the significance of the independent variables for the following two data sets. In the first case, you are evaluating the statement that we hear that Maximum Heart Rate of a person is related to their age by the following equation:

$$MaxHR = 220Age$$

Perform a linear regression analysis fitting the Max Heart Rate to Age using the lm function in R. What is the resulting equation? Is the effect of Age on Max HR significant? What is the significance level? Please also plot the fitted relationship between Max HR and Age.

```
#create data frame
(regData = data.frame(
    Age = c(18, 23, 25, 35, 65, 54, 34, 56, 72, 19, 23, 42, 18, 39, 37),
MaxHR = c(202, 186, 187, 180, 156, 169, 174, 172, 153, 199, 193, 174, 198, 183, 178)
))
```

```
##    Age MaxHR
## 1   18   202
## 2   23   186
## 3   25   187
## 4   35   180
## 5   65   156
## 6   54   169
## 7   34   174
## 8   56   172
## 9   72   153
## 10  19   199
## 11  23   193
## 12  42   174
## 13  18   198
## 14  39   183
## 15  37   178
```

The estimate of the model intercept is 210.0486 The signifiance level is F-statistic: 130 on 1 and 13 DF, p-value: 3.848e-08

```
alli.mod1 <- lm(MaxHR ~ Age, data = regData)
summary(alli.mod1)
```

```
##
## Call:
## lm(formula = MaxHR ~ Age, data = regData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9258 -2.5383  0.3879  3.1867  6.6242
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

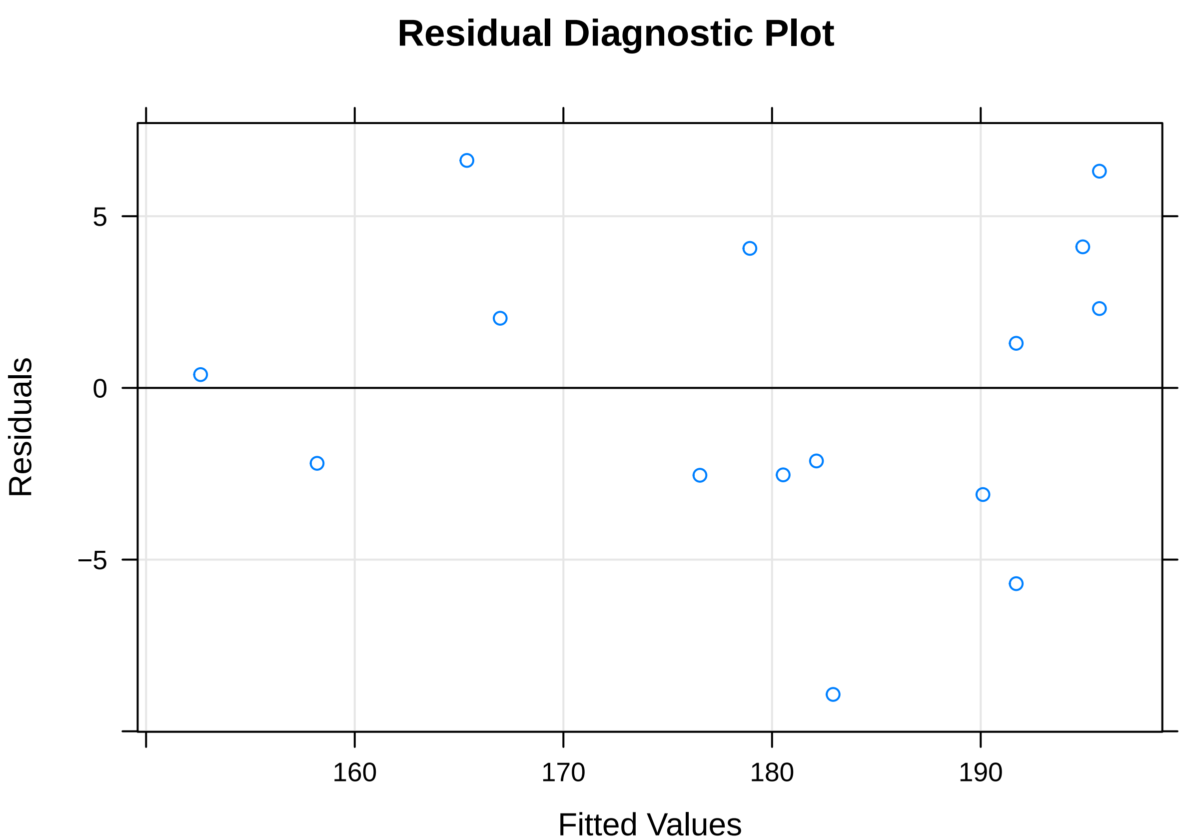
```
## (Intercept) 210.04846    2.86694   73.27  < 2e-16 ***
## Age          -0.79773    0.06996  -11.40 3.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.578 on 13 degrees of freedom
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9021
## F-statistic:   130 on 1 and 13 DF,  p-value: 3.848e-08
```

```r
require(lattice)
```

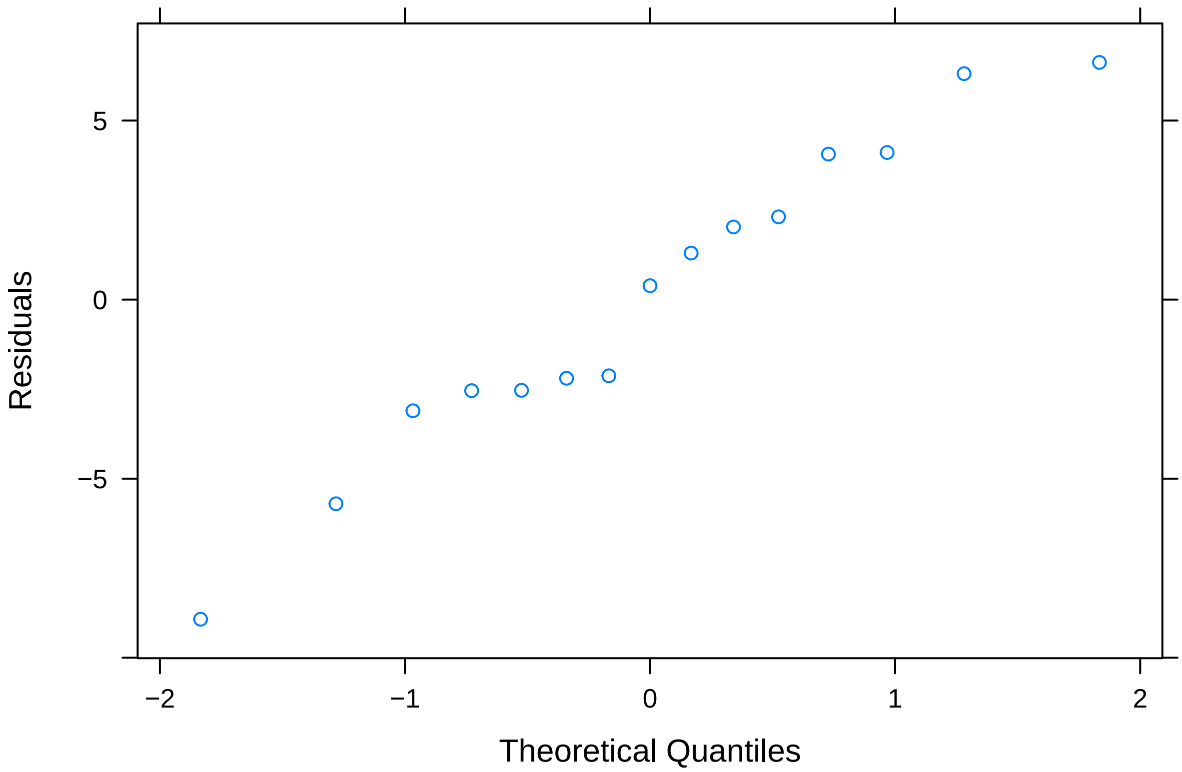
```
## Loading required package: lattice
```

```
## Warning: package 'lattice' was built under R version 3.3.2
```

```r
#Plot the residuals
xyplot(resid(alli.mod1) ~ fitted(alli.mod1),
  xlab = "Fitted Values",
  ylab = "Residuals",
  main = "Residual Diagnostic Plot",
  panel = function(x, y, ...)
  {
    panel.grid(h = -1, v = -1)
    panel.abline(h = 0)
    panel.xyplot(x, y, ...)
  }
)
```



```r
#The function resid extracts the model residuals from the fitted model object
qqmath( ~ resid(alli.mod1),
  xlab = "Theoretical Quantiles",
  ylab = "Residuals"
)
```



Using the Auto data set from Assignment 5 perform a Linear Regression analysis using mpg as the dependent variable and the other 4 (displacement, horse- power, weight, acceleration) as independent variables. What is the final linear regression fit equation? Which of the 4 independent variables have a significant impact on mpg? What are their corresponding significance levels? What are the standard errors on each of the coefficients?

```r
auto <- as.data.frame(read.table("auto-mpg.data", header = FALSE, as.is = TRUE))
colnames(auto) <- c("displacement", "horsepower", "weight", "acceleration", "mpg")
head(auto)
```

```
##   displacement horsepower weight acceleration mpg
## 1          307        130   3504         12.0  18
## 2          350        165   3693         11.5  15
## 3          318        150   3436         11.0  18
## 4          304        150   3433         12.0  16
## 5          302        140   3449         10.5  17
## 6          429        198   4341         10.0  15
```

```r
autoLm = lm(formula = mpg ~ displacement + horsepower + weight + acceleration,
       data = auto)
(autoLmSum <- summary(autoLm))
```

```
##
## Call:
## lm(formula = mpg ~ displacement + horsepower + weight + acceleration,
```

```
##     data = auto)
##
## 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 ***
## displacement -0.0060009  0.0067093  -0.894  0.37166
## horsepower   -0.0436077  0.0165735  -2.631  0.00885 **
## weight       -0.0052805  0.0008109  -6.512  2.3e-10 ***
## 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 intercept and coefficients - the

```r
autoLmSum$coefficients[,1]
```

```
##  (Intercept) displacement   horsepower       weight acceleration
## 45.251139699 -0.006000871 -0.043607731 -0.005280508 -0.023147999
```

From the results, the liner regression fit equation is: mpg = 45.251139699 + -0.006000871 * displacement + -0.043607731 * horsepower - -0.005280508 * weight - -0.023147999 * acceleration – weight has a significant impact on mpg

Take the entire data set (all 392 points) and perform linear regression and measure the 95% confidence intervals.

```r
#Examine at 95% confidence interval
confint(autoLm, level = .95)
```

```
##                     2.5 %       97.5 %
## (Intercept)  40.422278855 50.080000544
## displacement -0.019192122  0.007190380
## horsepower   -0.076193029 -0.011022433
## weight       -0.006874738 -0.003686277
## acceleration -0.270094049  0.223798050
```

First take any random 40 data points from the entire auto data sample and perform the linear regression fit and measure the 95% confidence intervals.

```r
autoSample <- auto[sample(1:nrow(auto), 40,
    replace=FALSE),]
head(autoSample)
```

```
##     displacement horsepower weight acceleration  mpg
## 271        151.0         85   2855         17.6 23.8
## 44         400.0        175   5140         12.0 13.0
## 58          97.5         80   2126         17.0 25.0
## 192        200.0         81   3012         17.6 24.0
## 119        114.0         91   2582         14.0 20.0
## 55          97.0         60   1834         19.0 27.0
```

```r
autoSampLm = lm(formula = mpg ~ displacement + horsepower + weight + acceleration,
       data = autoSample)
(autoSampSum <- summary(autoSampLm))
```

```
##
## Call:
## lm(formula = mpg ~ displacement + horsepower + weight + acceleration,
##     data = autoSample)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0244 -1.8915 -0.3693  1.4328  8.9366
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.78612    7.02369   4.810 2.84e-05 ***
## displacement -0.03268    0.01539  -2.123   0.0409 *
## horsepower    0.04188    0.05070   0.826   0.4144
## weight       -0.00366    0.00211  -1.735   0.0916 .
## acceleration  0.11454    0.33278   0.344   0.7328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.241 on 35 degrees of freedom
## Multiple R-squared:  0.7372, Adjusted R-squared:  0.7071
## F-statistic: 24.54 on 4 and 35 DF,  p-value: 9.708e-10
```

```r
#Examine at 95% confidence interval
confint(autoSampLm, level = .95)
```

```
##                     2.5 %        97.5 %
## (Intercept)  19.527281600 48.0449587709
## displacement -0.063917591 -0.0014369500
## horsepower   -0.061041216  0.1447947965
## weight       -0.007943611  0.0006229602
## acceleration -0.561041658  0.7901293002
```

Please report the resulting fit equation, their significance values and confidence intervals for each of the two runs.

The p=value of the 392 records is larger than the p-value fo the sample.

```r
autoLmSum  #entire data set
```

```
##
## Call:
## lm(formula = mpg ~ displacement + horsepower + weight + acceleration,
##     data = auto)
##
## 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 ***
```

```
## displacement -0.0060009  0.0067093  -0.894  0.37166
## horsepower   -0.0436077  0.0165735  -2.631  0.00885 **
## weight       -0.0052805  0.0008109  -6.512  2.3e-10 ***
## 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
```

```r
autoSampSum #sample summary
```

```
##
## Call:
## lm(formula = mpg ~ displacement + horsepower + weight + acceleration,
##     data = autoSample)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0244 -1.8915 -0.3693  1.4328  8.9366
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.78612    7.02369   4.810 2.84e-05 ***
## displacement -0.03268    0.01539  -2.123   0.0409 *
## horsepower    0.04188    0.05070   0.826   0.4144
## weight       -0.00366    0.00211  -1.735   0.0916 .
## acceleration  0.11454    0.33278   0.344   0.7328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.241 on 35 degrees of freedom
## Multiple R-squared:  0.7372, Adjusted R-squared:  0.7071
## F-statistic: 24.54 on 4 and 35 DF,  p-value: 9.708e-10
```

Reference https://www.r-bloggers.com/simple-linear-regression-2/ https://www.r-bloggers.com/r-tutorial-series-multiple-linear-regression/