

Multiple linear regression

Amira Ibrahim

October 24, 2019

Attach data

```
library(readr)
Multiple_Linear_regression <- read.csv(file.choose(),header = TRUE)
attach(Multiple_Linear_regression)

names(Multiple_Linear_regression)

## [1] "Treatment"      "HDL_LEVEL"      "SEX"
## [5] "dummy_treatment"
## [5] "dummy_sex"
```

check normality :

```
library(moments)
skewness(HDL_LEVEL)
```

```
## [1] 0.2515518
```

accepted range from -1 to +1

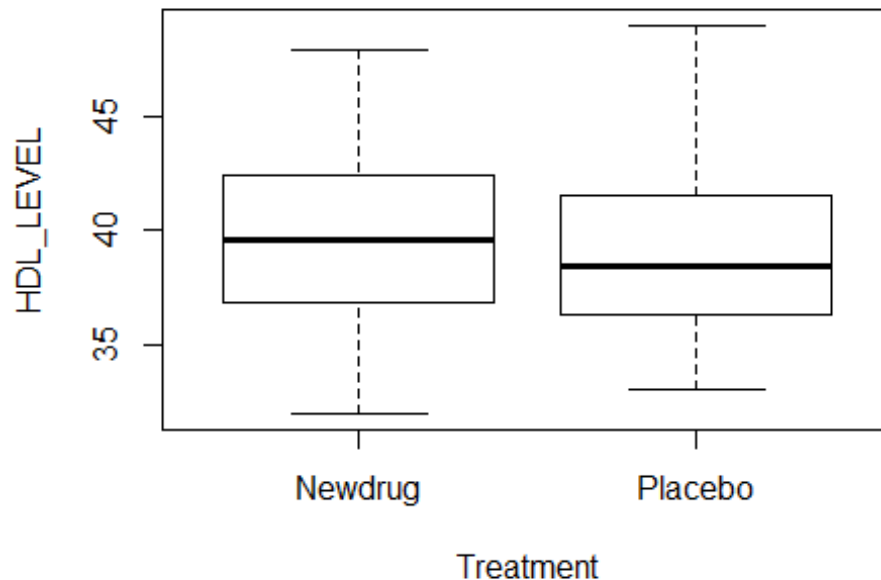
```
kurtosis(HDL_LEVEL)
```

```
## [1] 2.523675
```

accepted range from -2 to +2 may to +3

```
boxplot(HDL_LEVEL~Treatment , main="Relation between type of treatment & HDL
level")
```

Relation between type of treatment & HDL level



visually, data is

normally distributed

```
shapiro.test(HDL_LEVEL)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  HDL_LEVEL  
## W = 0.98428, p-value = 0.2814
```

p-value > 0.05 , Fail to reject H_0 , data is normally distributed

test : H_0 :mean of HDL level in new drug = mean of HDL level in placebo

```
t.test(HDL_LEVEL~Treatment , mu =0 , paired= F , alternative="two.sided"  
,conf.int=TRUE, conf.level=0.95)
```

```
##  
##  Welch Two Sample t-test  
##  
## data:  HDL_LEVEL by Treatment  
## t = 0.62698, df = 97.681, p-value = 0.5321  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
##  -1.016751  1.955911  
## sample estimates:
```

```
## mean in group Newdrug mean in group Placebo
##          39.56499          39.09541
```

p-value > 0.05 , Fail to reject H0 , No significant difference in mean between two types of treatment.

Check correlation :

```
cor(dummy_treatment, HDL_LEVEL)

## [1] 0.06320816

# +ve very weak correlation
cor((dummy_treatment+dummy_sex), HDL_LEVEL)

## [1] 0.264165

# +ve weak correlation
```

building regressin model

```
mlm1 <- lm(HDL_LEVEL~dummy_treatment , data = Multiple_Linear_regression)
summary(mlm1)

##
## Call:
## lm(formula = HDL_LEVEL ~ dummy_treatment, data =
Multiple_Linear_regression)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.532 -2.720 -0.291  2.545  9.849
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    39.0954     0.5296  73.822  <2e-16 ***
## dummy_treatment  0.4696     0.7490   0.627   0.532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.745 on 98 degrees of freedom
## Multiple R-squared:  0.003995,    Adjusted R-squared:  -0.006168
## F-statistic: 0.3931 on 1 and 98 DF,  p-value: 0.5321
```

test if sex is a cofounder

```
mlm2 <- lm(HDL_LEVEL~dummy_treatment+dummy_sex , data =
Multiple_Linear_regression)
summary(mlm2)
```

```
##
## Call:
## lm(formula = HDL_LEVEL ~ dummy_treatment + dummy_sex, data =
Multiple_Linear_regression)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.1004 -2.3789 -0.3158  2.7165  8.7120
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    38.5555     0.4985  77.344 < 2e-16 ***
## dummy_treatment  0.5776     0.6848   0.843  0.401
## dummy_sex       5.3995     1.1964   4.513 1.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.422 on 97 degrees of freedom
## Multiple R-squared:  0.1768, Adjusted R-squared:  0.1599
## F-statistic: 10.42 on 2 and 97 DF, p-value: 7.963e-05

(0.5776 - 0.4696)*100

## [1] 10.8
```

there is 10.8 % change in coefficient , Sex is confounder

```
mlm3 <- lm(HDL_LEVEL~dummy_treatment+dummy_sex+(dummy_treatment*dummy_sex) ,
data=Multiple_Linear_regression)
summary(mlm3)

##
## Call:
## lm(formula = HDL_LEVEL ~ dummy_treatment + dummy_sex + (dummy_treatment *
##      dummy_sex), data = Multiple_Linear_regression)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.3426 -2.0625 -0.2795  2.5988  8.4698
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    38.3079     0.4987  76.812 < 2e-16 ***
## dummy_treatment  1.0673     0.7015   1.522  0.1314
## dummy_sex       7.8753     1.5771   4.994 2.65e-06 ***
## dummy_treatment:dummy_sex -5.5033     2.3513  -2.341  0.0213 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.346 on 96 degrees of freedom
```

```
## Multiple R-squared:  0.2213, Adjusted R-squared:  0.1969
## F-statistic: 9.093 on 3 and 96 DF,  p-value: 2.345e-05
```

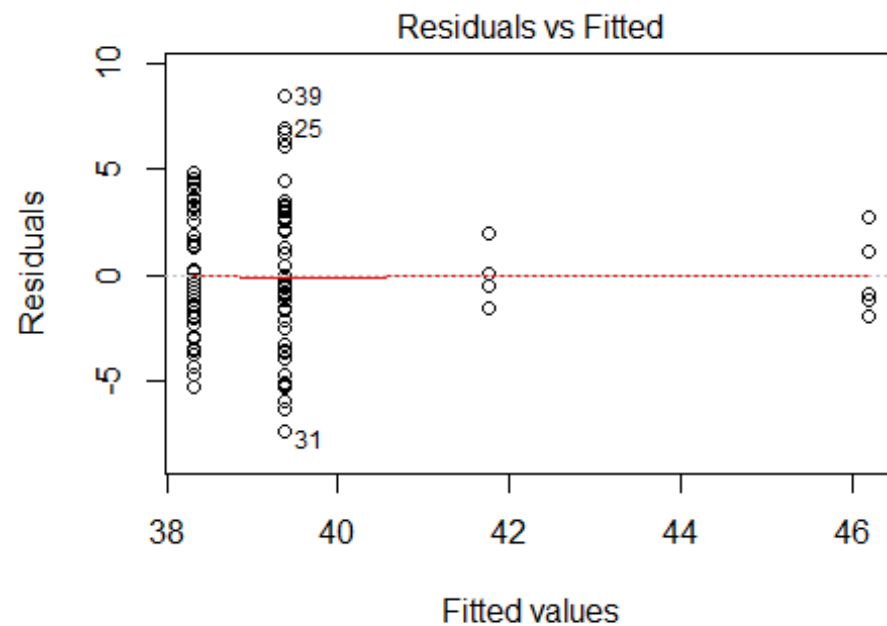
$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + b_3X_1X_2$$

HDL level in Male (NEW Drug) = $38.3079 + (1.06731) + (7.87531) + (-5.503311)$

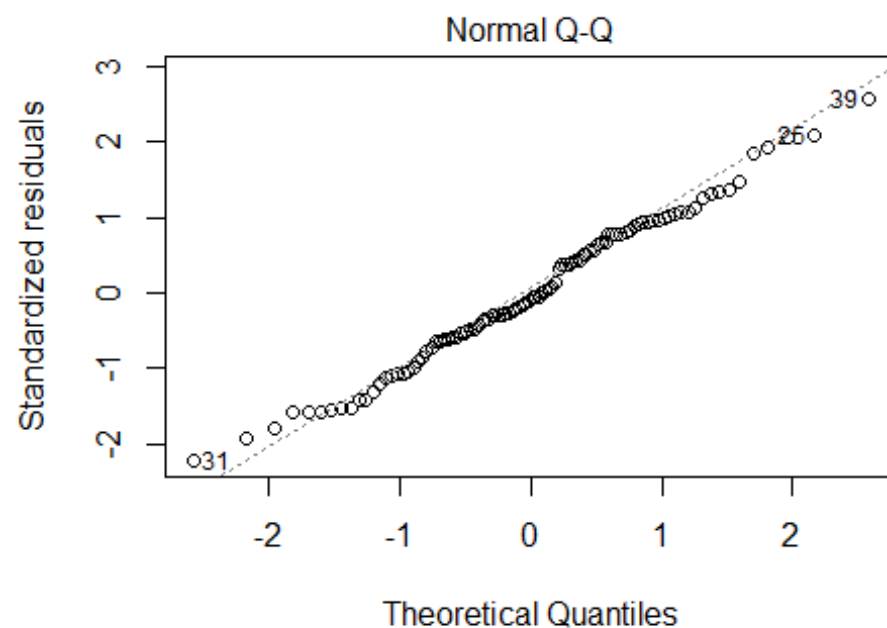
HDL level in Female (New Drug) = $38.3079 + (1.06731) + (7.87530) + (-5.503310)$

visualize the assumption

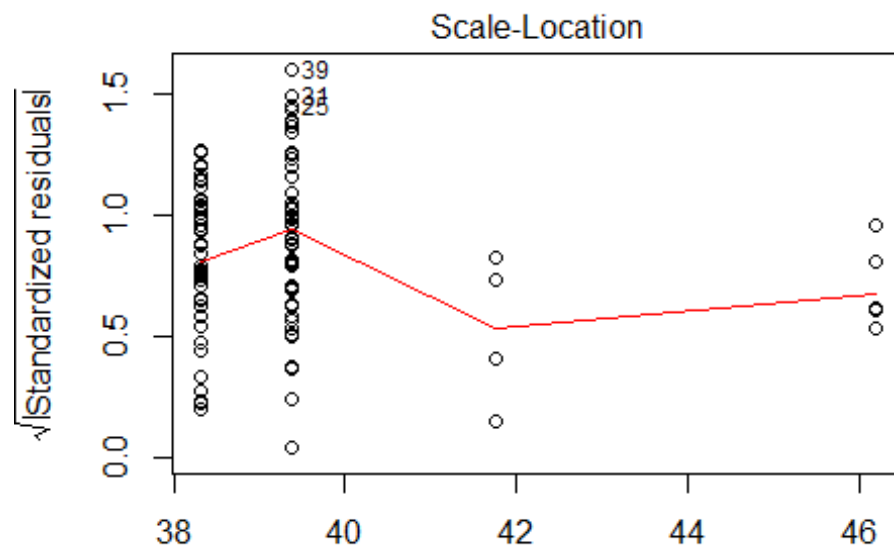
```
plot(mlm3)
```



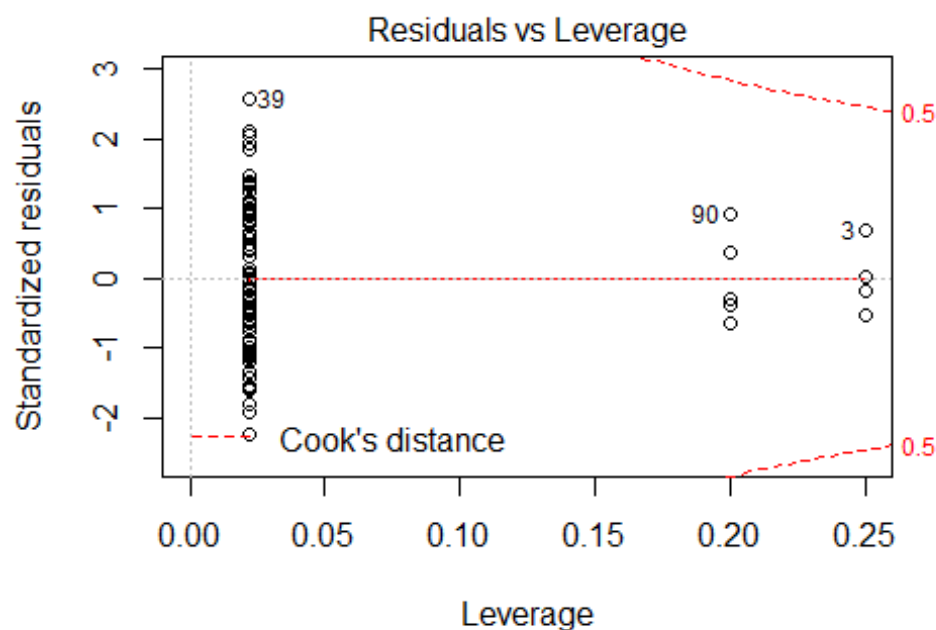
$L_LEVEL \sim \text{dummy_treatment} + \text{dummy_sex} + (\text{dummy_treatment} * \text{di})$



$L_LEVEL \sim \text{dummy_treatment} + \text{dummy_sex} + (\text{dummy_treatment} * \text{di})$



Fitted values
 $L_LEVEL \sim \text{dummy_treatment} + \text{dummy_sex} + (\text{dummy_treatment} * \text{dummy_sex})$



Leverage
 $L_LEVEL \sim \text{dummy_treatment} + \text{dummy_sex} + (\text{dummy_treatment} * \text{dummy_sex})$