R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

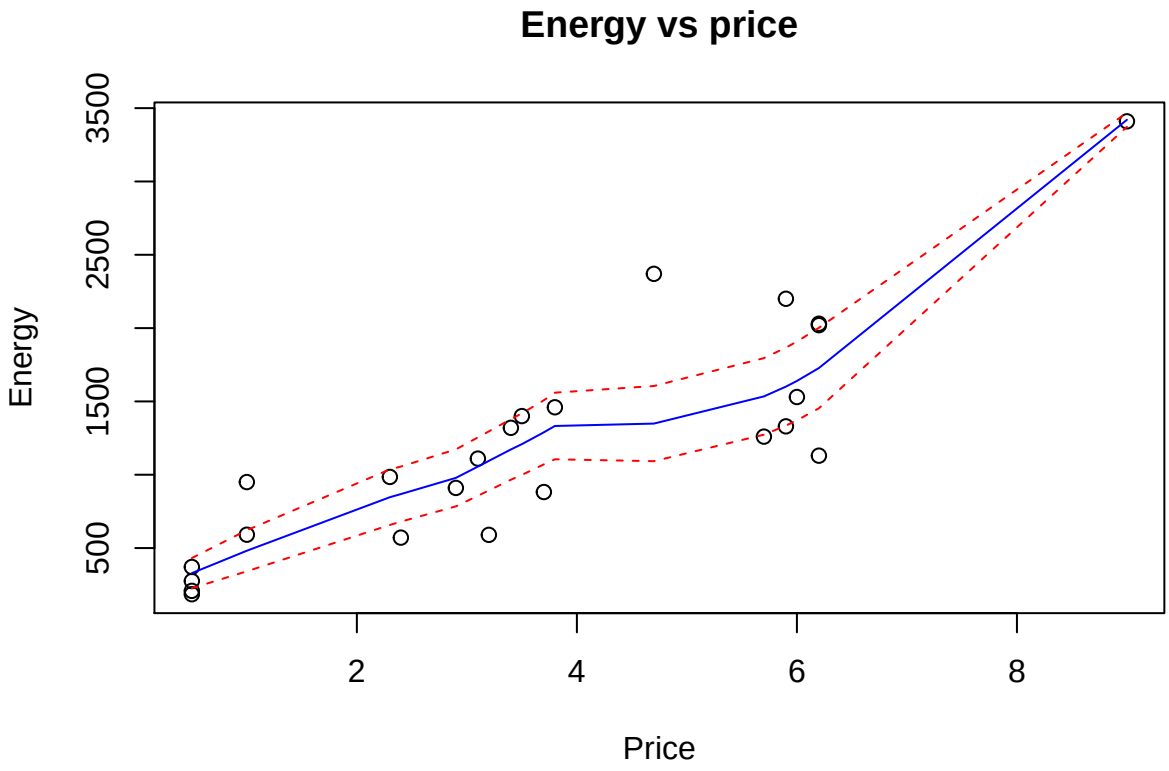
When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

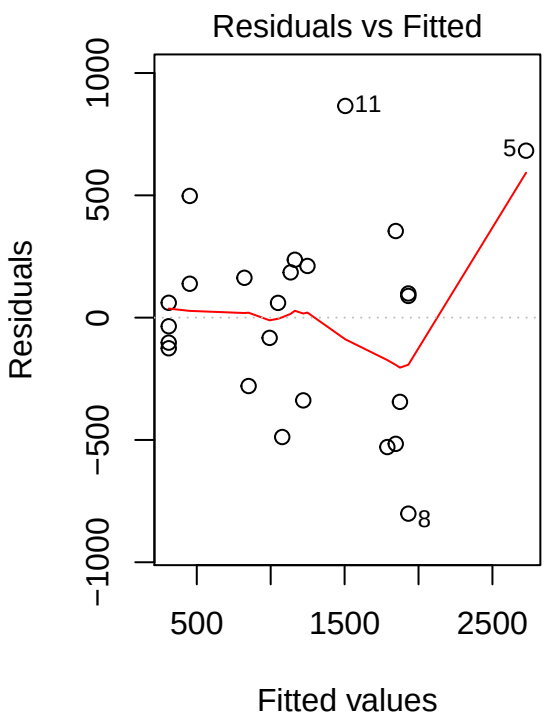
```
library(s20x)
#laptop:/
#setwd("C:/Users/Alex/Documents/GitHub/Engsci211Assignment2")
#home computer:
setwd("I:/GitHub/Engsci 211/Engsci211Assignment2")
energy.df = read.table("McDonalds.txt", header = TRUE)
head(energy.df)
```

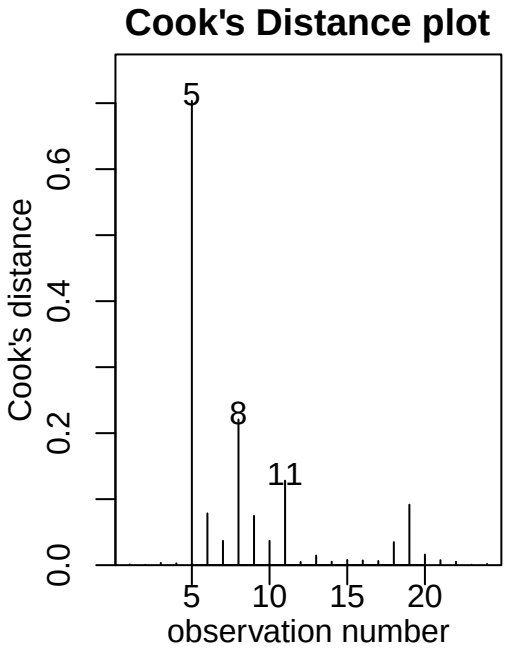
```
##   Energy Price
## 1    910   2.9
## 2   1110   3.1
## 3   2030   6.2
## 4   2020   6.2
## 5   3410   9.0
## 6   1330   5.9
```

```
trendscatter(Energy~Price, main = "Energy vs price", data = energy.df)
```




```
energy.fit = lm(Energy~Price, data = energy.df)
layout20x(1,2)
plot(energy.fit,which=1)
cooks20x(energy.fit)
```





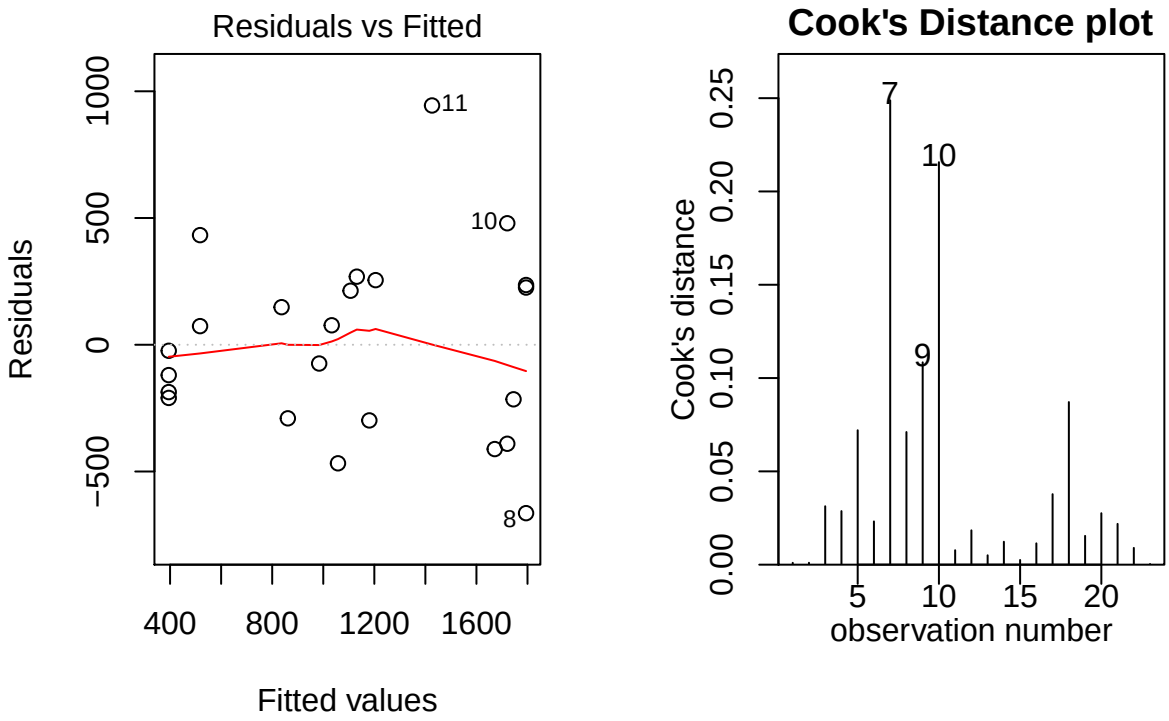

```
summary(energy.fit)
```

```
##
## Call:
## lm(formula = Energy ~ Price, data = energy.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -801.23 -294.28   60.52  191.58  865.33
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   168.11     154.58   1.088    0.289
## Price         284.37      35.67   7.973 6.24e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 402.5 on 22 degrees of freedom
## Multiple R-squared:  0.7429, Adjusted R-squared:  0.7312
## F-statistic: 63.57 on 1 and 22 DF,  p-value: 6.245e-08
```

As we can see the data is quite heavily skewed, and, the observation at position 5 is heavily influential with a cooks distance of 0.7, which is greater than 0.4.

```
energy.fit2 <- lm(Energy~Price, data = energy.df[-5,])
layout20x(1,2)
plot(energy.fit2,which=1)
```

```
cooks20x(energy.fit2)
```

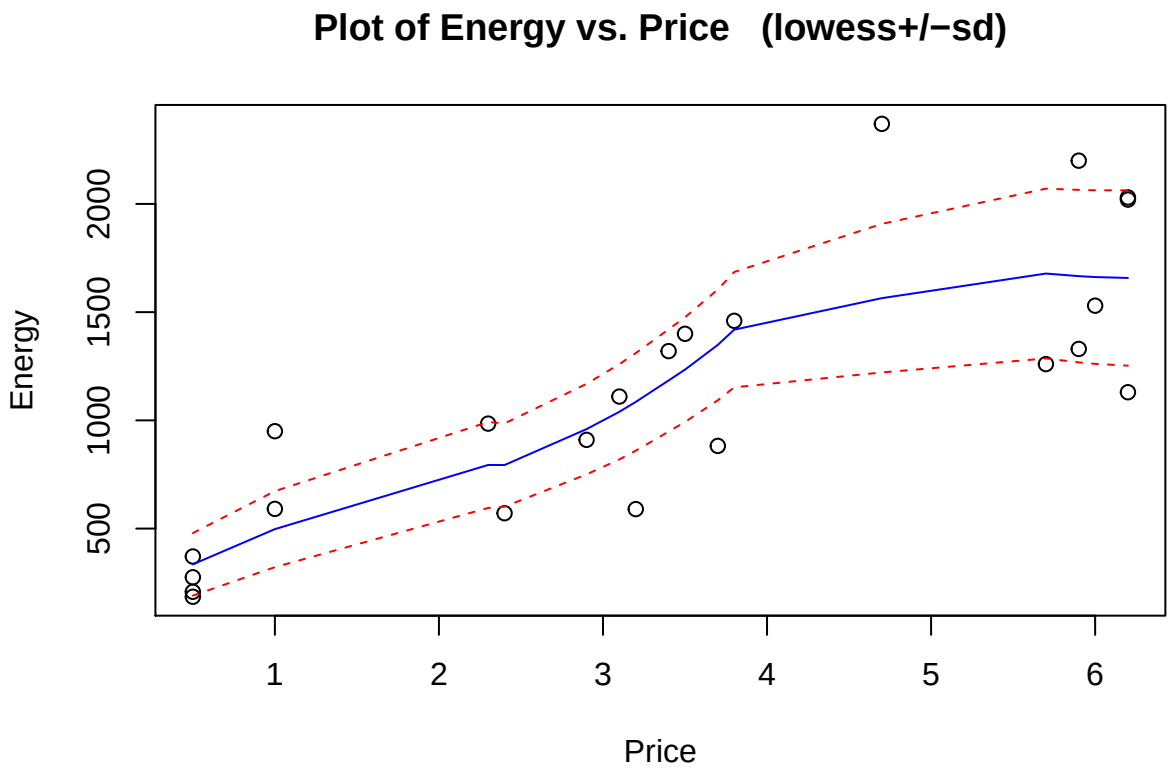



```
summary(energy.fit2)
```

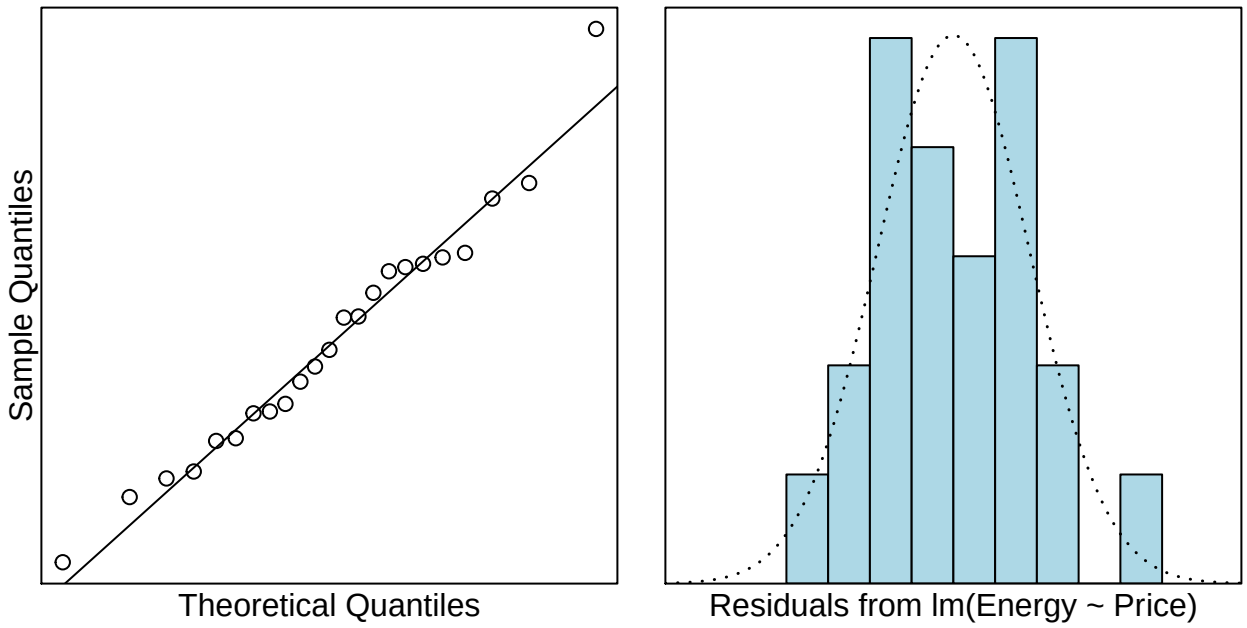
```
##
## Call:
## lm(formula = Energy ~ Price, data = energy.df[-5, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -664.34 -252.76  -23.76  230.66  943.97
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   271.99     151.53   1.795   0.0871 .
## Price         245.54      37.79   6.497 1.95e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 373.6 on 21 degrees of freedom
## Multiple R-squared:  0.6678, Adjusted R-squared:  0.652
## F-statistic: 42.21 on 1 and 21 DF,  p-value: 1.946e-06
```

As we can see, the cooks distance plot no longer shows any influential observations above 0.4. Similarly, we can see that by removing the observation the estimate for price shifts from 284.37 to 245.54, which is a change of 38.83, which is a change of more than 35.57, which is the standard error, which implies the observation at 5 should be removed.

```
trendscatter(Energy~Price, data = energy.df[-5,])
```




```
normcheck(energy.fit2)
```



As we can see the sample with the influential observation removed, the distribution is normal so our assumption of normality is correct.

```
summary(energy.fit2)
```

```
##
## Call:
## lm(formula = Energy ~ Price, data = energy.df[-5, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -664.34 -252.76  -23.76  230.66  943.97
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   271.99     151.53   1.795   0.0871 .
## Price         245.54      37.79   6.497 1.95e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 373.6 on 21 degrees of freedom
## Multiple R-squared:  0.6678, Adjusted R-squared:  0.652
## F-statistic: 42.21 on 1 and 21 DF,  p-value: 1.946e-06
```

```
pred.df = data.frame
```