Engsci 211 Task 2: Energy vs price for McDonalds

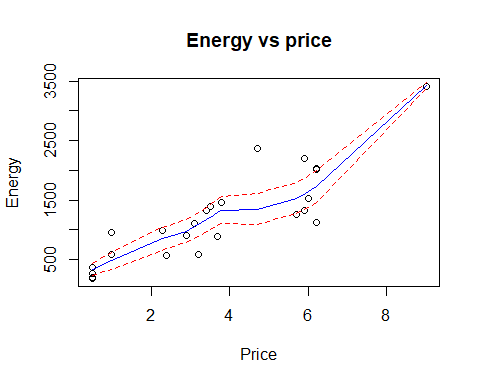
Alex Verkerk

library(s20x)  
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

### Exploratory Analysis

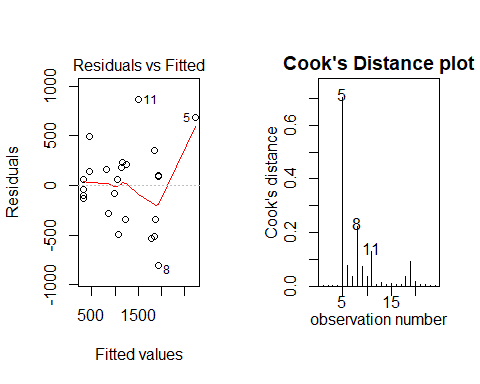
trendscatter(Energy~Price, main = "Energy vs price", data = energy.df)



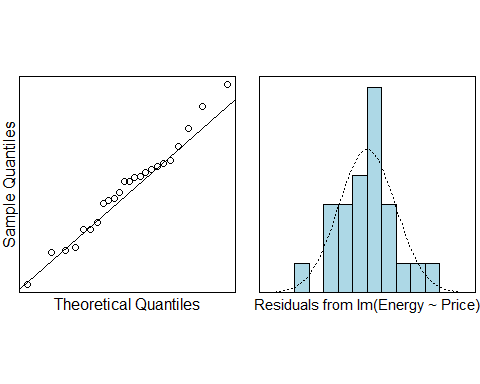
There appears to be a positive linear relationship between the price of McDonalds meals and energy in kilojoules you can expect from the meal. The scatter is approximately constant.

##### Original EOV and normality checks:

energy.fit = lm(Energy~Price, data = energy.df)  
layout20x(1,2)  
plot(energy.fit,which=1)#Equality of variance check with a trend line  
cooks20x(energy.fit)



normcheck(energy.fit)



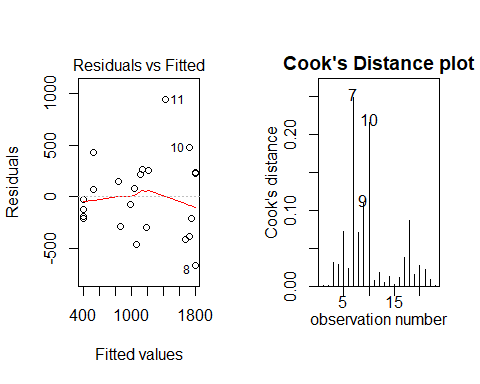
summary(energy.fit)

##   
## 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

The observation at position 5 is heavily influential with a cooks distance of 0.7, which is greater than 0.4 and suggests removal.

##### EOV & normality checks with observation 5 removed:

energy.fit2 <- lm(Energy~Price, data = energy.df[-5,])  
layout20x(1,2)  
plot(energy.fit2,which=1)#Equality of variance check with a trend line  
cooks20x(energy.fit2)

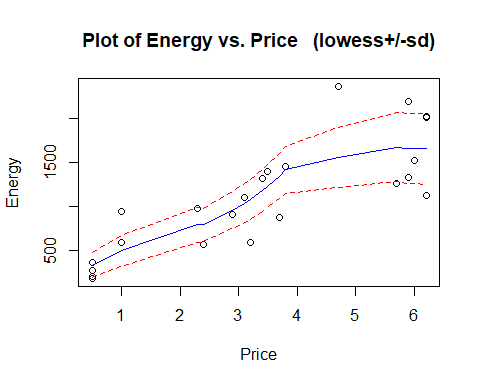


summary(energy.fit2)

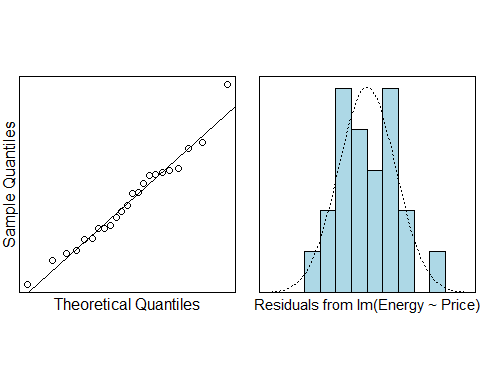
##   
## Call:  
## lm(formula = Energy ~ Price, data = energy.df[-5, ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -664.34 -252.76 -23.76 230.66 943.97   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 271.99 151.53 1.795 0.0871 .   
## Price 245.54 37.79 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 35.67, which is the standard error, which implies observation 5 should be removed.

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

 The previous positive linear relationship appears roughly maintained, with an increase in scatter toward the right hand side of the graph. This scatter isn’t particularly concerning as it is likely just due to lack of data points at the $5+ range.

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.

confint(energy.fit2)

## 2.5 % 97.5 %  
## (Intercept) -43.13239 587.1131  
## Price 166.94820 324.1320

pred.df = data.frame(Price = c(1,5.90))  
predict(energy.fit2, pred.df, interval = "prediction")

## fit lwr upr  
## 1 517.5304 -298.9690 1334.030  
## 2 1720.6770 903.7747 2537.579

### Methods and Assumption Checks

A scatter plot of price vs energy of McDonalds meals shows a roughly linear association, with approximately constant scatter, and so we fitted a linear model. We assume independant observations.

The equality of variance assumption for our model appears to unsatisfied, bending quite dramatically. The assumption of Normality appears satisfied but looks slightly abnormal.

We observe an influential observation (observation 5) in the Cooks plot. Upon removal of observation 5, the coefficient changes by more than 1 standard error and thus should be removed. (For the coefficient changes from 284.37 to 245.54, which is a drop of 38.83 compared to a standard error of 37.79.) After this our assumption of Equality of Variance appears to be satisfied, alongside our assumption of Normality.

The model we fitted is ${\tt Energy\_i} = \beta\_0+\beta\_1\times price\_i + \epsilon\_i$ where .

The multiple value of our model is fairly low (less than 80%), which means the overall variability of our sample is not fully accounted for.

### Executive summary

This data was collected to find a relationship between energy content of McDonalds food and its price, in order to judge the expected energy gained from purchasing food for $5.90.

We found that as price increased, the energy content of the food also increased. Discounting negative energy contents, a McDonalds meal will begin with up to 587kJ of energy, and We estimate that for every additional dollar spent, the mean number of kilojoules will increase by somewhere between 167 to 324kJ.

Our model only explains 66.8% of the variation between price and energy values of meals at McDonalds, which means our model is likely not a reasonable model for prediction.

If we were to use our model predict the energy you may expect for $5.90, you can get between 903.8 and 2537.6 kJ of energy. This may not be a useful prediction, and may only be useful in addition to other models.