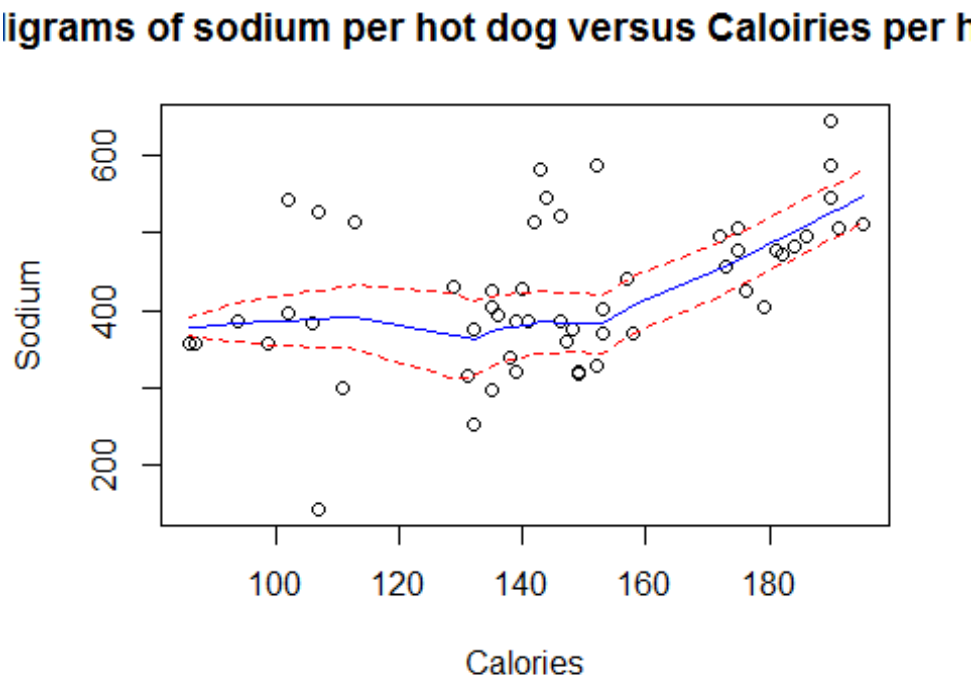
# Task Two

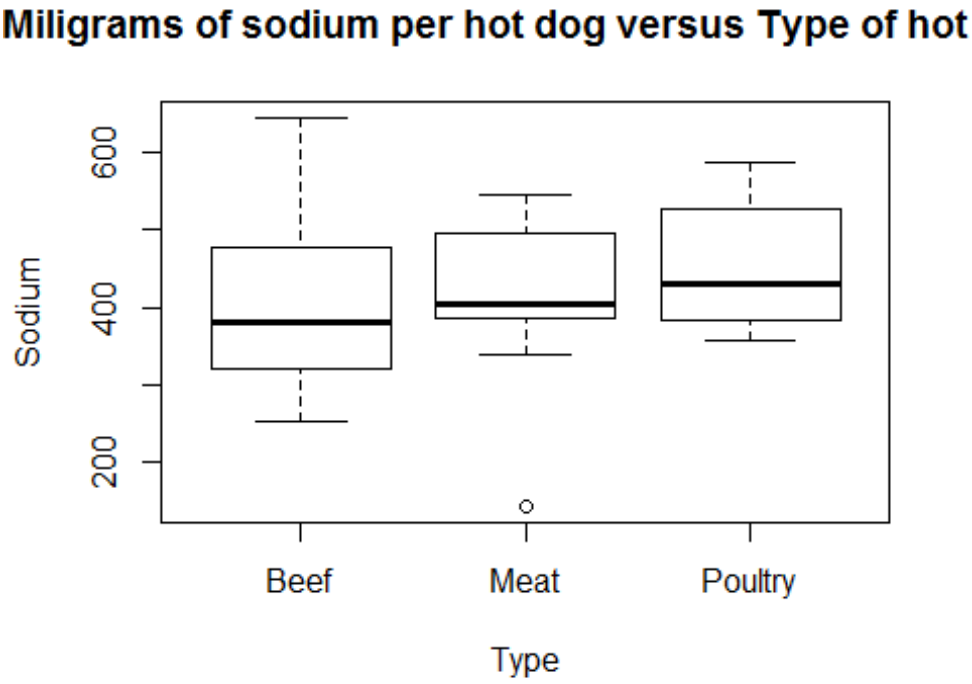
Sang Hyun Kho

11 September 2016

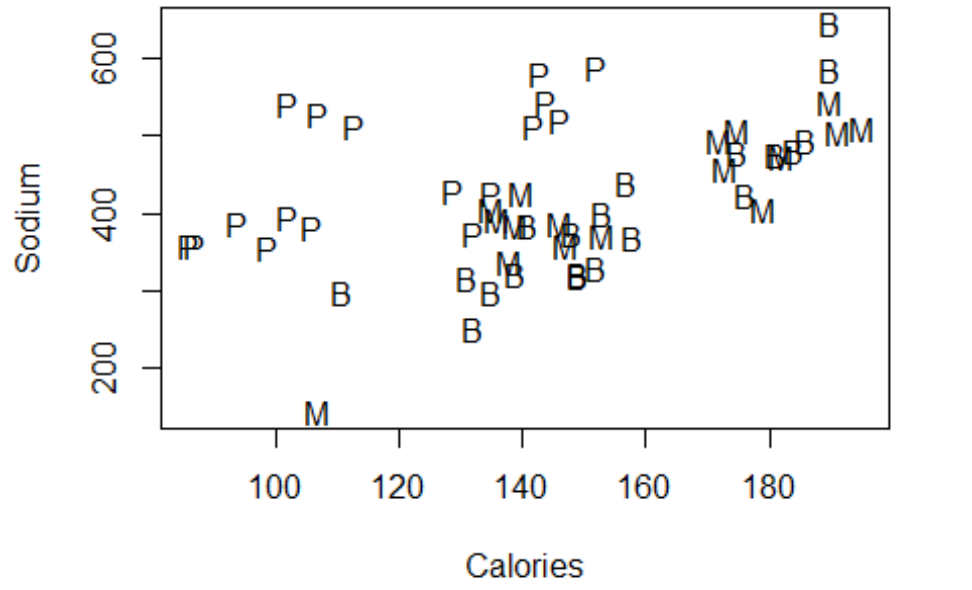
```
library(s20x)
setwd("C:/Users/Buzz/Desktop/Uni/MM3/AssignmentTwo")
hotdogs.df = read.table(file.choose(), header = TRUE) #read text file
trendscatter(Sodium ~ Calories, main = "Miligrams of sodium per hot dog versu
s Caloiries per hot dog" , data = hotdogs.df)
```



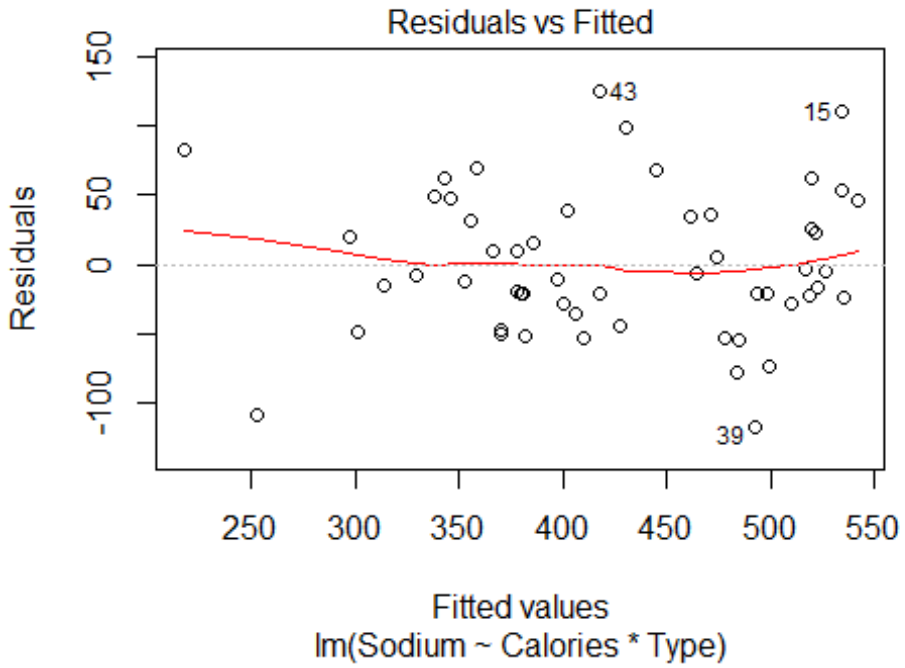

```
#factoring
hotdogs.df$Type = factor(hotdogs.df$Type)
plot(Sodium ~ Type, main = "Miligrams of sodium per hot dog versus Type of ho
t dog", data = hotdogs.df)
```



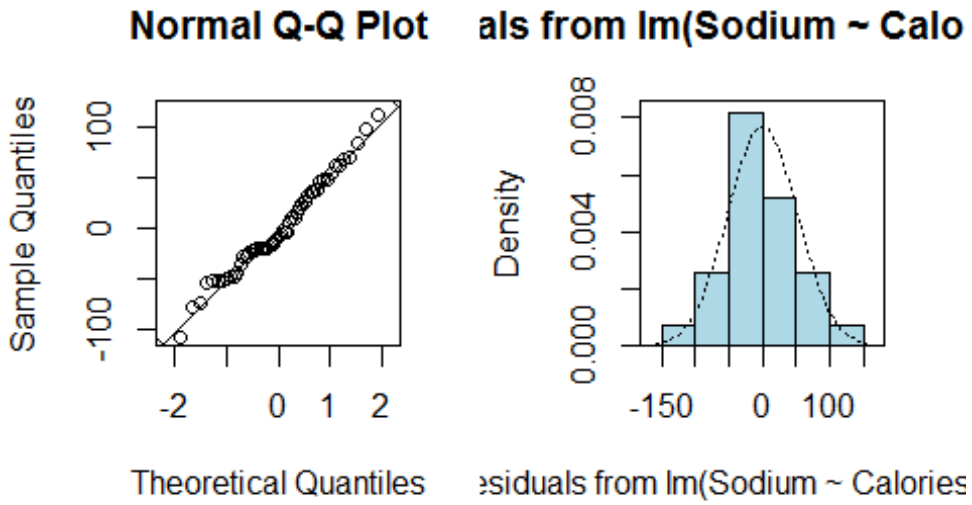

```r
plot(Sodium ~ Calories, pch = substr(Type, 1, 1) , data = hotdogs.df) # P=poultry, B=beef and M=meat
```



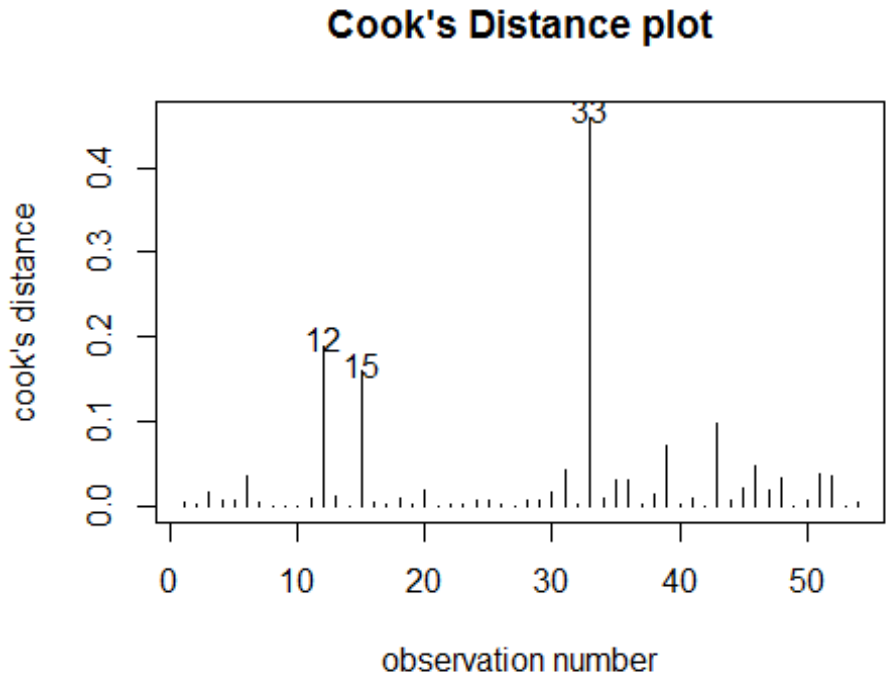
```
hotdogs.fit = lm(Sodium ~ Calories * Type, data = hotdogs.df)
plot(hotdogs.fit, which = 1)
```




```
normcheck(hotdogs.fit)
```




```
cooks20x(hotdogs.fit)
```



```r
summary(hotdogs.fit)
```

```
##
## Call:
## lm(formula = Sodium ~ Calories * Type, data = hotdogs.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -116.916  -28.180   -8.961   35.798  124.694
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -228.3313    87.5770  -2.607  0.01213 *
## Calories                4.0133     0.5529   7.259 2.96e-09 ***
## TypeMeat              137.1460   123.3106   1.112  0.27159
## TypePoultry           391.9615   114.0463   3.437  0.00122 **
## Calories:TypeMeat      -0.8016     0.7733  -1.037  0.30511
## Calories:TypePoultry   -1.5263     0.8195  -1.862  0.06868 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.57 on 48 degrees of freedom
## Multiple R-squared:  0.7065, Adjusted R-squared:  0.6759
## F-statistic: 23.11 on 5 and 48 DF,  p-value: 9.698e-12
```

```r
hotdogs.fit2 <- lm(Sodium ~ Calories * Type, data = hotdogs.df[-33, ]) # dele
ting 33rd row data
summary(hotdogs.fit2) # Check for effect of dropping observation 33
```

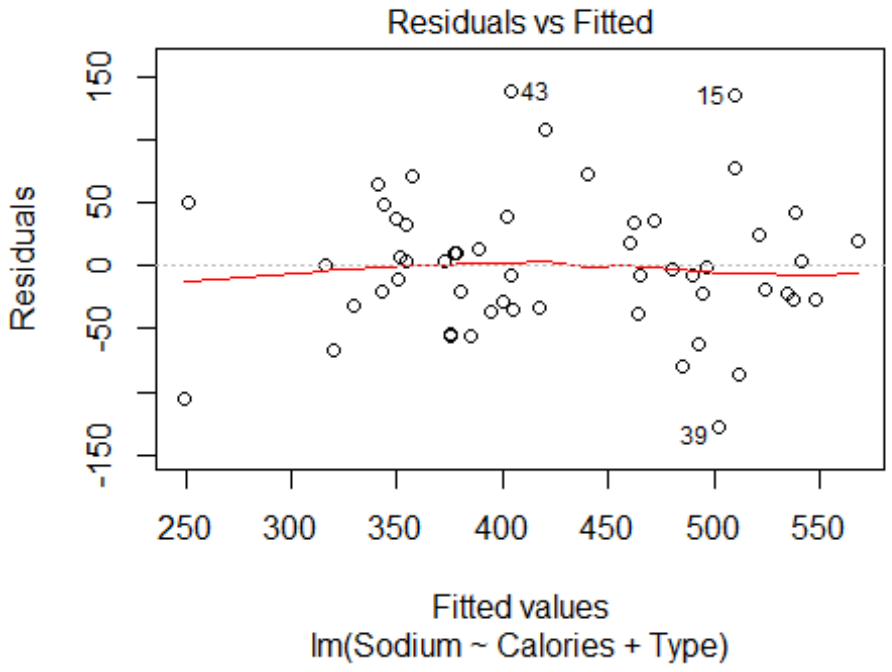
```
##
## Call:
## lm(formula = Sodium ~ Calories * Type, data = hotdogs.df[-33,
##     ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -116.916  -39.217   -4.735   36.154  124.694
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -228.3313    82.9644  -2.752 0.008386 **
## Calories                4.0133     0.5238   7.662 8.17e-10 ***
## TypeMeat              275.2313   128.7861   2.137 0.037822 *
## TypePoultry           391.9615   108.0395   3.628 0.000702 ***
## Calories:TypeMeat      -1.6124     0.7987  -2.019 0.049242 *
## Calories:TypePoultry   -1.5263     0.7764  -1.966 0.055237 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 51.69 on 47 degrees of freedom
## Multiple R-squared:  0.6911, Adjusted R-squared:  0.6583
## F-statistic: 21.03 on 5 and 47 DF,  p-value: 5.593e-11
```

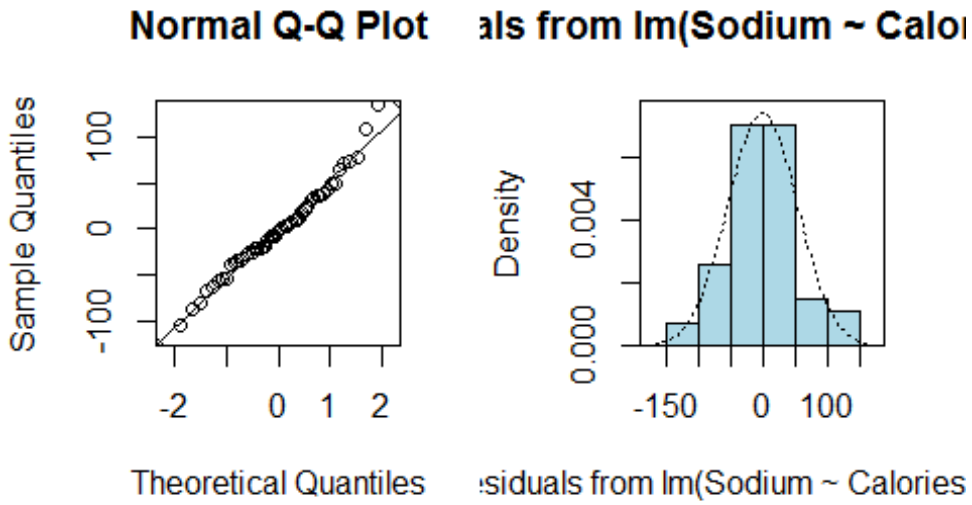
```
anova(hotdogs.fit2)
```

```
## Analysis of Variance Table
##
## Response: Sodium
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Calories       1  77952   77952 29.1705 2.143e-06 ***
## Type           2 188368   94184 35.2446 4.460e-10 ***
## Calories:Type  2  14714    7357  2.7531   0.07402 .
## Residuals     47 125598    2672
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

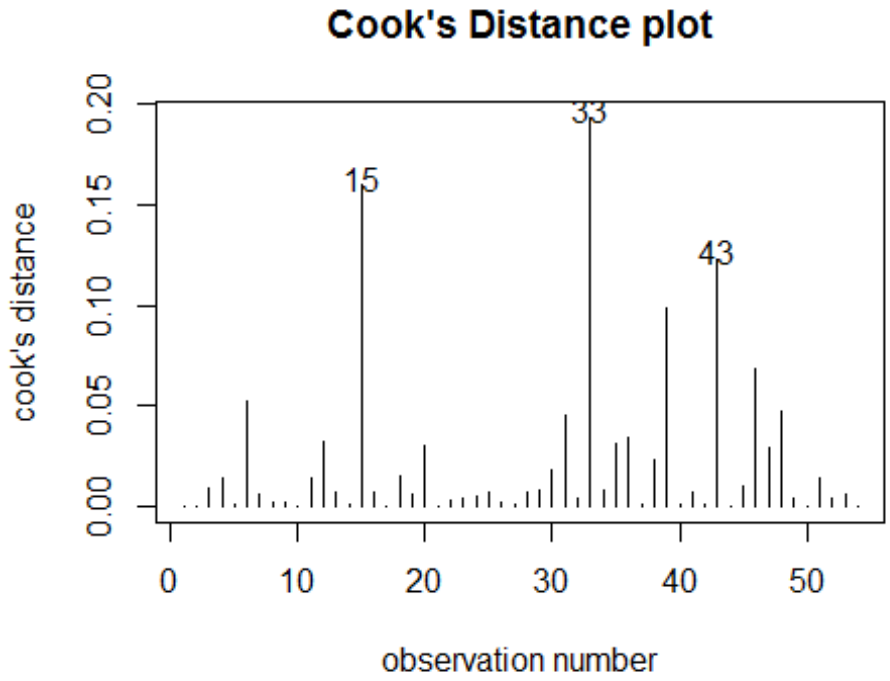
```
# Drop interaction -> Parallel lines model
hotdogs.fit3 = lm(Sodium ~ Calories + Type, data = hotdogs.df)

plot(hotdogs.fit3, which = 1)
```




```
normcheck(hotdogs.fit3)
```




```
cooks20x(hotdogs.fit3)
```



```r
anova(hotdogs.fit3)
```

```
## Analysis of Variance Table
##
## Response: Sodium
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Calories   1 106270  106270  34.654 3.281e-07 ***
## Type       2 227386  113693  37.074 1.336e-10 ***
## Residuals 50 153331    3067
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(hotdogs.fit3)
```

```
##
## Call:
## lm(formula = Sodium ~ Calories + Type, data = hotdogs.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -127.409  -30.569   -2.556   33.427  137.985
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -113.2838    53.3043  -2.125   0.0385 *
## Calories       3.2798     0.3305   9.922 2.09e-13 ***
## TypeMeat      11.2925    18.2783   0.618   0.5395
## TypePoultry  182.7615    22.1856   8.238 7.15e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.38 on 50 degrees of freedom
## Multiple R-squared:  0.6851, Adjusted R-squared:  0.6663
## F-statistic: 36.27 on 3 and 50 DF,  p-value: 1.356e-12
```

```r
confint(hotdogs.fit3)
```

```
##                   2.5 %     97.5 %
## (Intercept) -220.348659  -6.218980
## Calories       2.615862   3.943703
## TypeMeat     -25.420599  48.005644
## TypePoultry  138.200304 227.322627
```

```r
# To compare Meat and Poultry, we need to rotate levels of the factor.
# Create new factor variable within the hotdogs.df dataframe:
hotdogs.df$Typefac2 = factor(hotdogs.df$Type, levels = c('Meat','Beef','Poult
ry'))

# Refit and get output
```

```
hotdogs.fit4 = lm(Sodium ~ Calories + Typefac2, data = hotdogs.df)
summary(hotdogs.fit4)$coef
```

```
##                     Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)     -101.991297 54.1515808 -1.883441 6.546333e-02
## Calories           3.279782  0.3305457  9.922326 2.088697e-13
## Typefac2Beef     -11.292522 18.2783376 -0.617809 5.395051e-01
## Typefac2Poultry  171.468943 23.1318549  7.412676 1.359599e-09
```

```
confint(hotdogs.fit4)
```

```
##                       2.5 %     97.5 %
## (Intercept)     -210.757948   6.775354
## Calories           2.615862   3.943703
## Typefac2Beef     -48.005644  25.420599
## Typefac2Poultry  125.007245 217.930641
```

```
predSodium.df = data.frame(Calories = c(425), Type = factor(c("Poultry")))
predSodium.df
```

```
##   Calories    Type
## 1      425 Poultry
```

```
predict(hotdogs.fit3, predSodium.df, interval = "confidence")
```

```
##        fit      lwr      upr
## 1 1463.385 1258.287 1668.483
```

## Method and Assumption Checks

To estimate the effect on sodium content, we first fitted the model with explanatory variables Type, Calories and their interaction.

We have random sample of hot dogs, so the independence assumption appears satisfied.

The equality of variance and normality assumptions appear to be satisfied by the original model. The observation 33 showed up of concern in the Cooks plot, and deleting it, changed either coefficient by more than 1 standard error, so we took out the observation 33. The interaction term was not significant (p-value = `0.07402`). The model was refitted with the interaction term removed. Again, the equality of variance and normality assumptions appear to be satisfied.

Statistical significance of a coefficient is equivalent to the confidence interval NOT containing zero (Confidence intervals for Meat and Beef hot dogs contain zero).

The final model is

$Sodium_i = \beta_0 + (\beta_1 \text{ x } Calories_i) + (\beta_2 \times D_{2i}) + (\beta_3 \times D_{3i} \times Calories) + \varepsilon_i$ where $D_{2i}, D_{3i} = 1\ if\ hot\ dog\ i\ is\ Meat\ or\ Poultry, otherwise\ 0\ and\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

Here Beef is our baseline.

The final model was also refitted with Meat as the baseline.

## Executive Summary

We were interested in comparing the effectiveness of three types of hot dogs on sodium content. We also wanted to see how this was affected by calories per hot dog.

We found that the effects of the type of hot dogs are the same regardless of calories and the effect of calories is the same regardless of the type of hot dog.

In particular, Poultry hot dog contains the significantly large sodium content than the other two types.

For the same type of hot dog, we estimate that the expected sodium content increases by between 2.6 and 3.9 milligrams for each additional 1 calorie.

Using our model, we predict that the sodium content is between 1258 and 1668 milligrams for a hot dog made from poultry containing 425 calories.

For hot dogs with the same calories, we estimate that the expected sodium content for:

- Poultry hot dogs contain between 138 and 227 milligrams of sodium higher than Beef hot dogs.
- Poultry hot dogs contain between 125 and 218 milligrams of sodium higher than Meat hot dogs.

Our model explains almost 69% of the variation in sodium content.