STATS 201/8 Assignment 1

Chunkai Meng / 857058772

Due Date: 3pm, Thursday 4th August 2016

## Loading required package: s20x

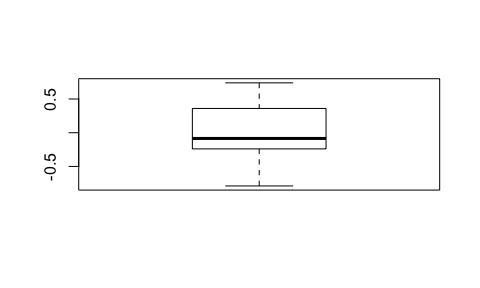
# Question 1

## Question of interest/goal of the study

We are interested in seeing if a fuel treatment improves the fuel consumption of vehicles.

## Read in and inspect the data:

Fuel.df=read.table("FuelConsumption.txt", header=T)  
Fuel.diff=Fuel.df$modified-Fuel.df$standard  
boxplot(Fuel.diff)



\*\* WRITE A COMMENT HERE

## Using the t.test function

t.test(Fuel.diff,var.equal=TRUE)

##   
## One Sample t-test  
##   
## data: Fuel.diff  
## t = 0.36177, df = 19, p-value = 0.7215  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## -0.1674911 0.2374911  
## sample estimates:  
## mean of x   
## 0.035

# Why 2-sample t-test is not appropriate

\*\* WRITE ANSWER HERE

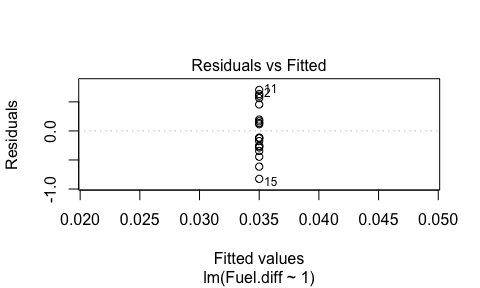
# Manual calculation in R

Formulae: , and 95% confidence interval

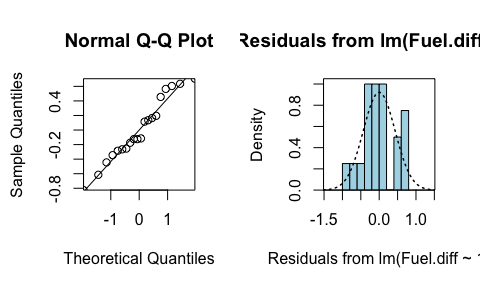
NOTES: The R code mean(y) calculates , sd(y) calculates the standard deviation of , and qt(0.975,25) gives the multiplier. The standard error, is calculated by the standard deviation of divided by the square root of of the sample size, .

# Using the lm function

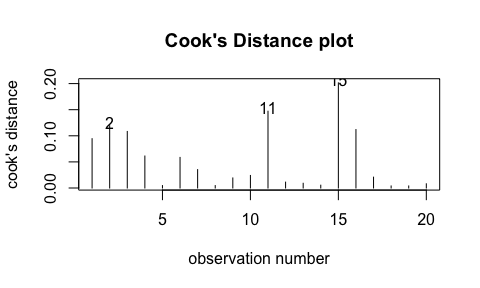
Fuel.lm=lm(Fuel.diff~1)  
plot(Fuel.lm,which=1)



normcheck(Fuel.lm)



cooks20x(Fuel.lm)



summary(Fuel.lm)

##   
## Call:  
## lm(formula = Fuel.diff ~ 1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.825 -0.270 -0.120 0.260 0.705   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.03500 0.09675 0.362 0.722  
##   
## Residual standard error: 0.4327 on 19 degrees of freedom

confint(Fuel.lm)

## 2.5 % 97.5 %  
## (Intercept) -0.1674911 0.2374911

## Note:

You should get exactly the same results from the manual calculations and using the linear model function as the t.test function. Doing this was to giving you practice using some R code.

# Method and Assumption Checks

As the data has two measurements on each vehicle, we have applied a paired sample t-test (i.e., a one-sample t-test applied to the differences within each vehicle.) There is no reason to suspect lack of independence between vehicles and no problem with residuals or any overly influential observations.

The estimated coefficient for the true fuel difference was not significantly different from zero (p-value 0.72). So, our preferred model is: where

# Executive Summary

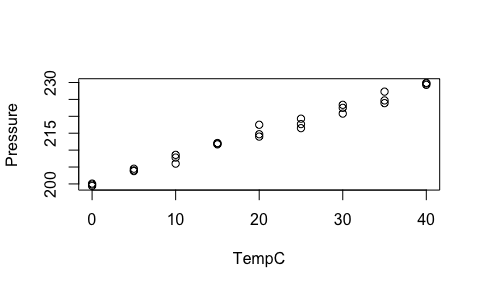
# Question 2

# Question of interest/goal of the study

\*\* WRITE HERE

# Read in and inspect the data:

Air.df=read.table("Pressure.txt", header=T)  
plot(Pressure~TempC,data=Air.df)



\*\* WRITE COMMENTS HERE

# Fit model and do checks

# Estimate the expected pressure at 50 degrees C

# Create plot with superimposed lines

# Method and Assumption Checks

# Executive Summary

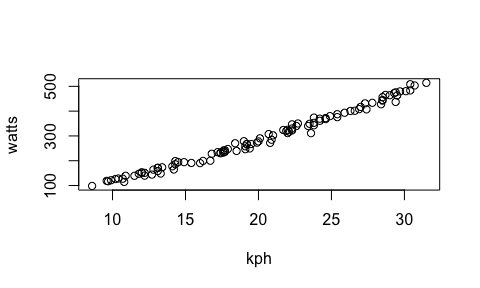
# Question 3

# Question of interest/goal of the study

\*\* WRITE HERE

# Read in and inspect the data:

Ebike.df=read.table("CyclePower.txt", header=T)  
plot(watts~kph,data=Ebike.df)



\*\* WRITE COMMENTS HERE

# Model fitting and checks.

# Prediction of power consumption at 20 kph

# Method and Assumption Checks

# Executive Summary

# Bonus question: