Problem 3 (6 credits)

HW2

Michael Brucek (mbrucek), Eduardo Chavez (echavezh), Kevin Grady (grady133), Danny Moncada (monca016), Tian Zhang (zhan7003)

March 02, 2020

```
suppressWarnings(suppressPackageStartupMessages({
  library(TSA)
  library(forecast)
  library(ggplot2)
  library(dplyr)
}))
```

Boston Crime Data Analysis

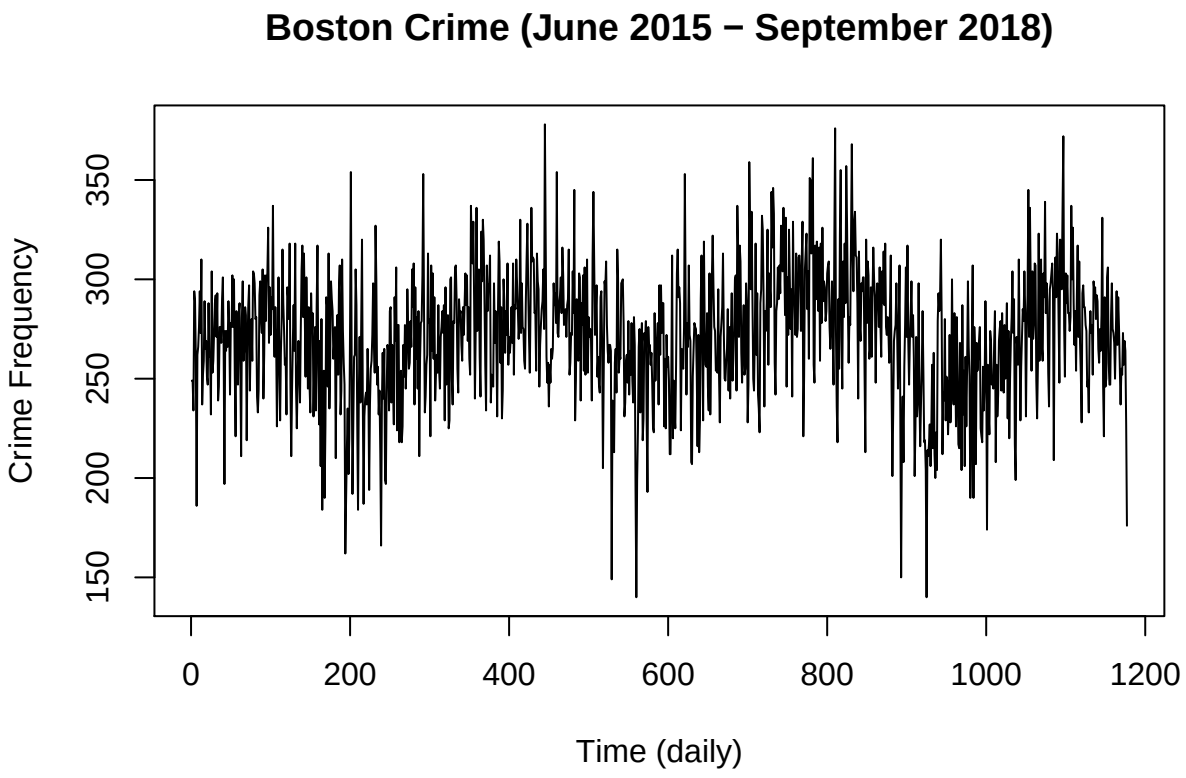
Question 1

Please pull out the crime frequency data we got from Homework 1 - Problem 3 - Question 4. You may re-plot the time series to refresh yourself about the pattern.

```
crime=read.table("crime.txt",header=T)

N=dim(crime)[1]
crime_aggr=aggregate(rep(1,N),list(year=crime[,1],month=crime[,2],day=crime[,3]),sum)
crime_aggr=crime_aggr[order(crime_aggr[,1],crime_aggr[,2],crime_aggr[,3]),]

Y_crime = crime_aggr[,4]
ts.plot(Y_crime, main = "Boston Crime (June 2015 - September 2018)",
        xlab = "Time (daily)", ylab = "Crime Frequency")
```



a) (1 credit)

First, let's fit an `auto.arima()` to find out a good ARIMA model for the data. Again, notice that, `auto.arima()` provides a "good" model but not necessarily the optimal. We will learn more concrete model selection techniques in Lecture 6.

Hints:

- use `auto.arima()` function

```
auto_crime <- auto.arima(Y_crime)
auto_crime
```

```
## Series: Y_crime
## ARIMA(1,0,3) with non-zero mean
##
## Coefficients:
##          ar1      ma1      ma2     ma3      mean
##       0.9888  -0.7142  -0.2542  0.0446  270.6409
## s.e.  0.0054   0.0298   0.0347  0.0292    5.5130
##
## sigma^2 estimated as 880.5:  log likelihood=-5658.27
## AIC=11328.53   AICc=11328.61   BIC=11358.96
```

b) (2 credits)

What's the model? For example

$$(Y_t - 10) = 0.4 \cdot (Y_{t-1} - 10) + e_t - 0.8 \cdot e_{t-1}$$

Hints:

- The mean value comes with every Y_t. In the example above, the mean value is 10.
- R assumes positive sign for MA models. In the example above, R would show -0.8, rather than +0.8 for the MA(1) coefficient

Please write down the model below:

$$(Y_t - 270.6409) = 0.9888 \cdot (Y_{t-1} - 270.6409) + e_t - 0.7142 \cdot e_{t-1} - 0.2542 \cdot e_{t-2} + 0.0446 \cdot e_{t-3}$$

c) (1 credit)

Are any of the coefficients significant?

Hints:

- A coefficient is significant if its magnitude is (roughly) at least twice as large as its standard error.

```
#Please write down your answer below
auto_crime
```

```
## Series: Y_crime
## ARIMA(1,0,3) with non-zero mean
##
## Coefficients:
##          ar1      ma1      ma2     ma3      mean
##       0.9888  -0.7142  -0.2542  0.0446  270.6409
## s.e.  0.0054   0.0298   0.0347  0.0292    5.5130
##
## sigma^2 estimated as 880.5:  log likelihood=-5658.27
## AIC=11328.53   AICc=11328.61   BIC=11358.96
```

```
#All of the coefficients of the model except for ma3 (0.0446) are significant according
#to the table above.  All of these coefficients have a magnitude twice as large as its
#standard error.
```

d) (2 credits)

Please superimpose the fitted values on the original crime frequency time series. Does the model sufficiently explain the data?

Hints:

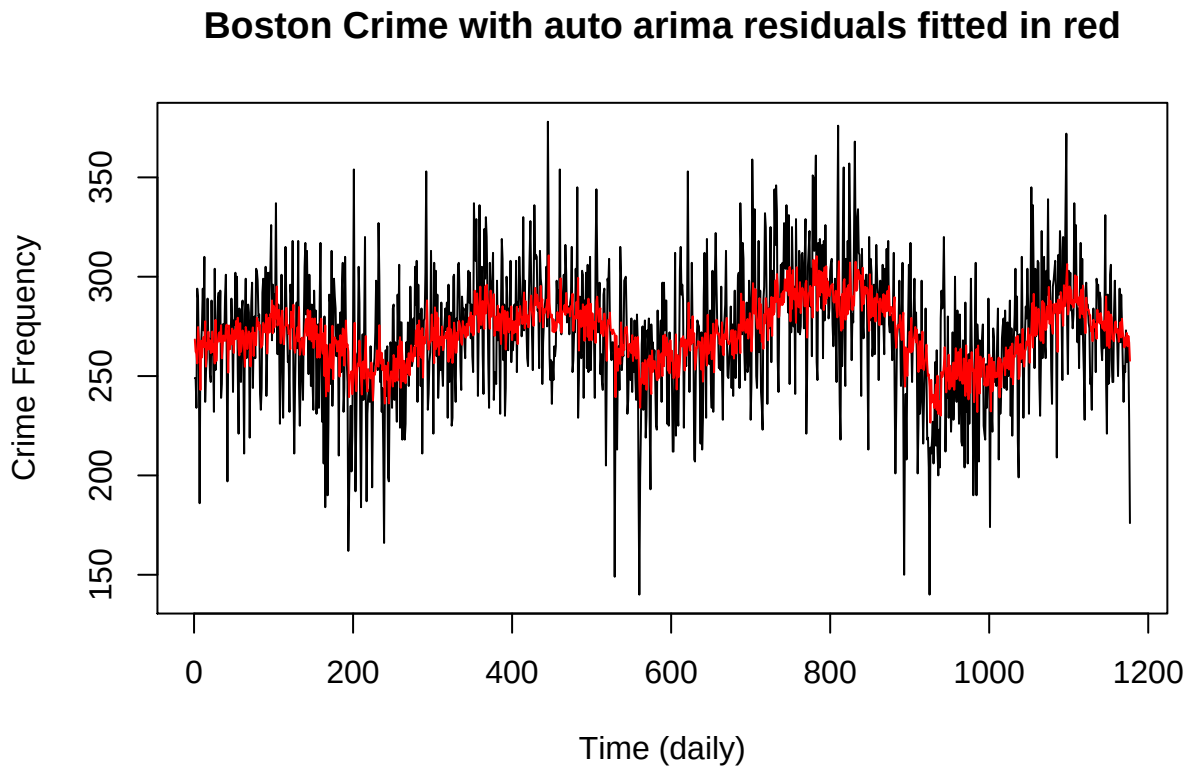
- The fitted values can be calculated by `the original time series - arima_fit$residuals`

```
crime_aggr$fitted <- Y_crime - auto_crime$residuals
t.test(Y_crime, crime_aggr$fitted)
```

```
##
##  Welch Two Sample t-test
##
## data:  Y_crime and crime_aggr$fitted
## t = 0.0088586, df = 1622.7, p-value = 0.9929
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.060779  2.079479
## sample estimates:
## mean of x mean of y
##  271.0901  271.0807
```

```r
#According to the two sample t-test results above, we cannot reject the null hypothesis
#that the true difference in means is equal to 0.

ts.plot(Y_crime, main = "Boston Crime with auto arima residuals fitted in red",
                  xlab = "Time (daily)", ylab = "Crime Frequency")
lines(crime_aggr$fitted, col = "red")
```




```r
#Combining with the time series plot of the original data and the fitted series,
#we can safely say that the model generated via the auto arima process sufficiently
#explains the Boston crime data.
```