Problem 3 (6 credits)

HW2

Danny Moncada (monca016)

February 22, 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.

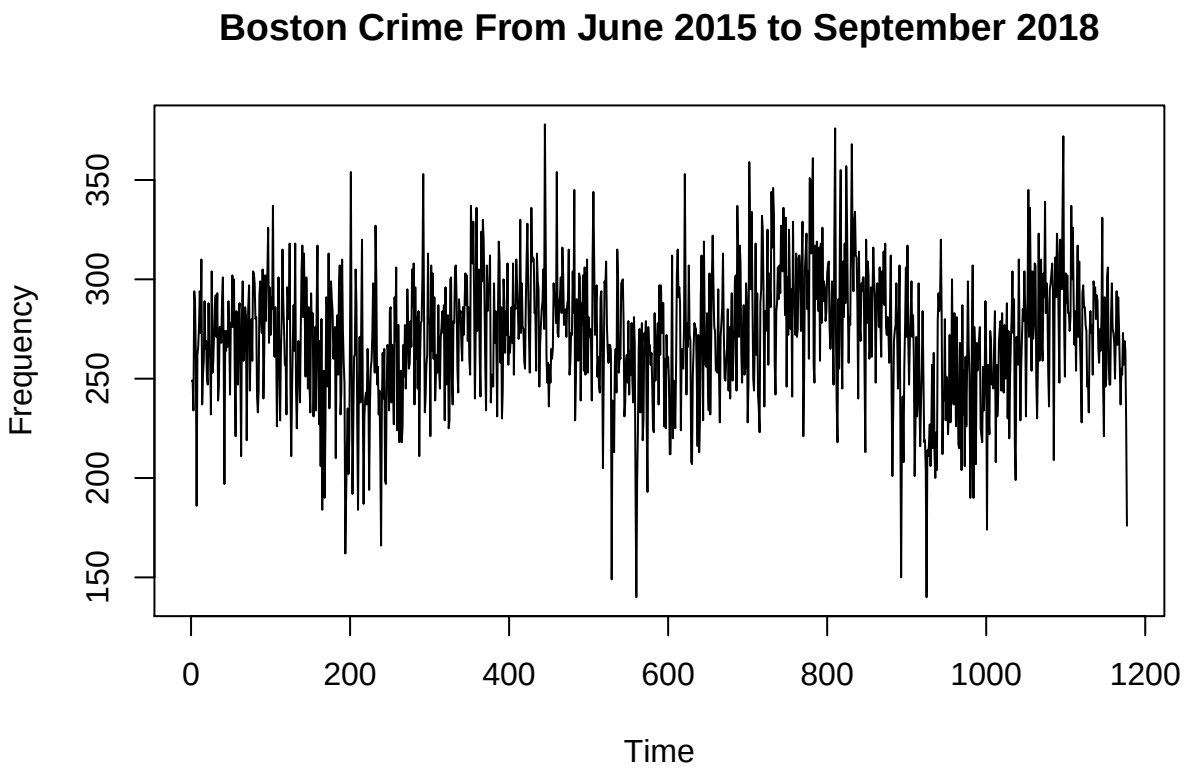
```
setwd("~/MSBA 2020 All Files/Spring 2020/MSBA 6430 - Advanced Issues in Business Analytics/HW2")

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_aggr[,4]

ts.plot(Y, ylab = "Frequency", main = "Boston Crime From June 2015 to September 2018")
```



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

```
arima_fit <- auto.arima(Y)
arima_fit
```

```
## Series: Y
## 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:

$$Y_t = 0.99 \cdot (Y_{t-1} - 270.64) + e_t - 0.71 \cdot e_{t-1} - 0.25 \cdot e_{t-2} + 0.04 \cdot e_{t-3} + 270.64$$

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.

```
##          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

"The first, second, and third coefficients are significant."
```

```
## [1] "The first, second, and third coefficients are significant."
```

```
"The standard error*2 for ar1 is 0.0108, while the coefficient is 0.9888."
```

```
## [1] "The standard error*2 for ar1 is 0.0108, while the coefficient is 0.9888."
```

```
"The standard error*2 for ma1 is 0.0596, while the coefficient is 0.7142."
```

```
## [1] "The standard error*2 for ma1 is 0.0596, while the coefficient is 0.7142."
```

```
"The standard error*2 for ma2 is 0.0694, while the coefficient is 0.2542."
```

```
## [1] "The standard error*2 for ma2 is 0.0694, while the coefficient is 0.2542."
```

d) (2 credits)

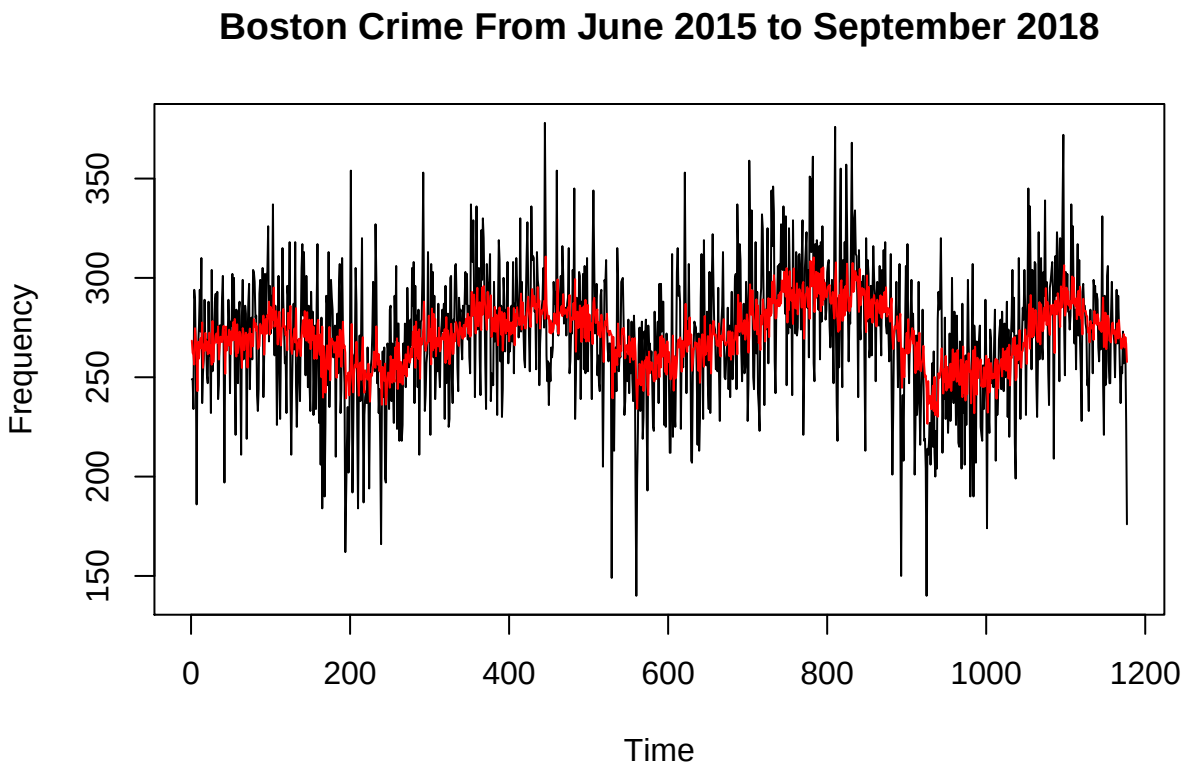
Please superimpose the fitted values on the original crime frequency time series. Does the model sufficiently explain the data?

Hints:

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

```
fitted_vals = Y - arima_fit$residuals

par(mfrow=c(1,1))
ts.plot(Y, ylab = "Frequency", main = "Boston Crime From June 2015 to September 2018")
lines(fitted_vals,col="red")
```




```
"The auto arima model DOES sufficiently explain the data - it generally follows the process laid out by
```

```
## [1] "The auto arima model DOES sufficiently explain the data - it generally follows the process laid
```

```
"It doesn't capture the fringe points but not every model can be perfect."
```

```
## [1] "It doesn't capture the fringe points but not every model can be perfect."
```