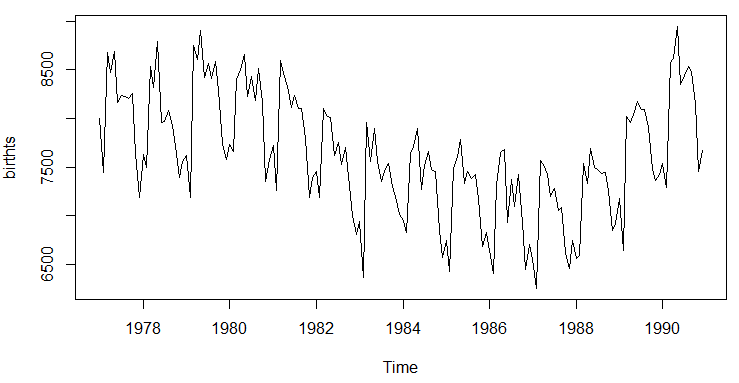
Homework 1 – Due in class on Tuesday, March 20th

**Oliver Yuan, Ruyi Shi**

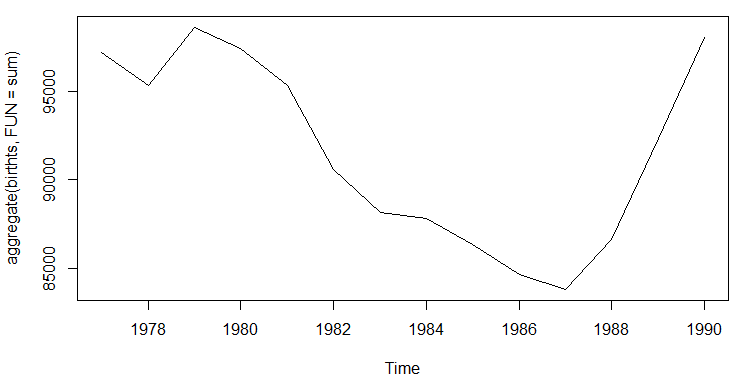
1. Choose a time index and aggregate the data (if needed). Submit a plot of the aggregated data (and be sure to mention the time index and aggregation method chosen).



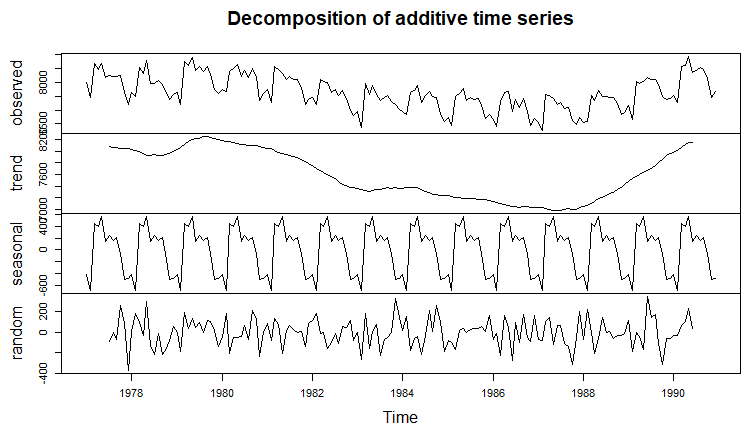
**We aggregated the data by month and sum all values in each month.**

1. Do the data exhibit trend? Seasonality? Cyclical behavior? Include (and refer to) the appropriate plots in your answer.

**Based on the plot below, we don’t see any cyclical behavior in the dataset.**

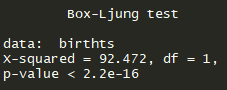


**Based on the plot below, we don’t see a persistent trend but do see seasonality in our dataset.**

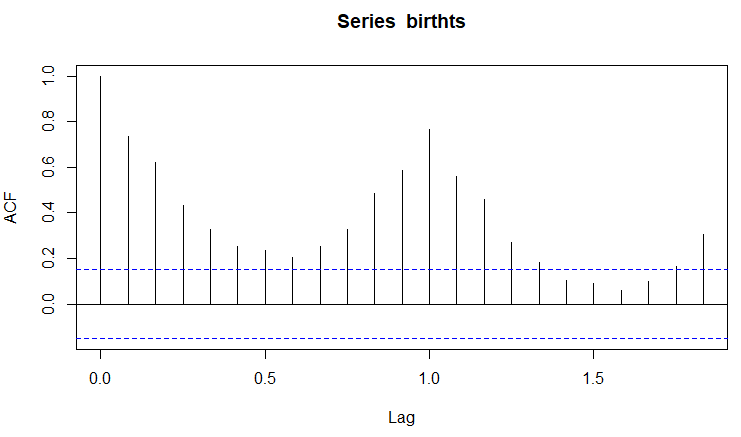


1. Is this a white noise series? Be sure to state the hypotheses and report the p-value when answering this question.

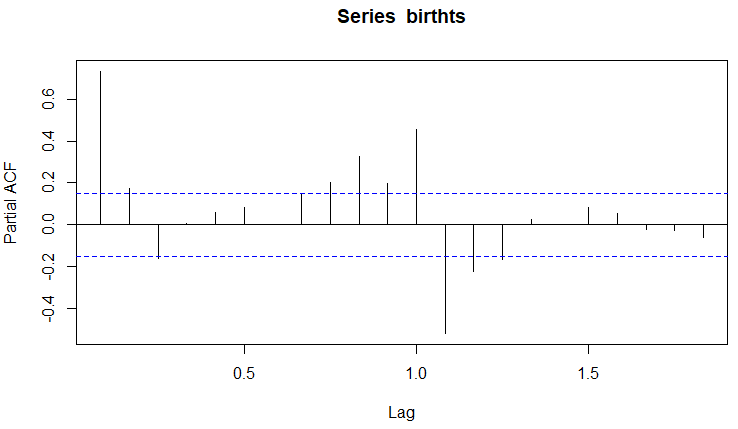
**The null hypothesis is that this is a white noise series while the alternative hypothesis is that this is not a white noise series. We got a p-value less than 2.2e-16. Therefore, we can reject the null hypothesis and conclude that this is not a white noise series.**

****

1. Are there any “significant” autocorrelations? Partial autocorrelations? Include the appropriate plots and circle or highlight the “significant” values.



**Based on the ACF plot above, we can see that most of autocorrelations are significant (lines that exceed the blue line).**

****

**Based on the PACF plot above, we can see that 1st, 7th, 8th, 9th, 10th, 11th, 12th order partial autocorrelations are significant (lines that exceed the blue line).**

1. How do you interpret the first order autocorrelation for this data set (e.g., explain in words what it tells you)? How do you interpret the 2nd order partial autocorrelation?

**The first order autocorrelation is greater than 0 and is above the 95% confidence interval, which suggests that the time series value is strongly affected by previous time series value.**

**The second partial autocorrelation is within the 95% confidence interval, which suggests that the time series value is not strongly affected by the second previous time series value.**

**Appendix:**

###### load packages ######

library(tidyverse)

library(lubridate)

library(forecast)

library(sarima)

###### data import and format ######

rawdata <- read.csv("Daily\_Births.csv")

rawdata$Date <- ymd(rawdata$Date)

rawdata$monthly <- substring(rawdata$Date,1,7)

###### aggregate data by month and create time series object ######

data\_m <- rawdata %>%

select(2,3) %>%

group\_by(monthly) %>%

dplyr::summarise(count = sum(Num\_Births)) %>%

arrange(monthly)

birthts <- ts(data\_m$count, frequency = 12, start = c(1977,1))

###### plot ######

plot.ts(birthts)

###### decomposing ######

### cyclical

cycle(birthts)

plot(aggregate(birthts, FUN = sum))

### trend seasonality and random

decomposedRes <- decompose(birthts)

plot(decomposedRes)

### white noise test

ltest <- Box.test(birthts,type = "Ljung-Box")

### autocorrelations

acf(birthts)

### partial autocorrelations

pacf(birthts)