

M11_L4_RomilShah

Romil Shah

August 1, 2016

```
library(RCurl)
## Loading required package: bitops
library(plyr)
## Warning: package 'plyr' was built under R version 3.2.5
library(ggplot2)
## Warning: package 'ggplot2' was built under R version 3.2.5
library(forecast)
## Warning: package 'forecast' was built under R version 3.2.5
## Loading required package: zoo
## Warning: package 'zoo' was built under R version 3.2.5
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
## Loading required package: timeDate
## This is forecast 7.1
library(astsa)
## Warning: package 'astsa' was built under R version 3.2.5
##
## Attaching package: 'astsa'
## The following object is masked from 'package:forecast':
##
##   gas
data <- read.csv("C:/Users/rams1/Desktop/DSCS6030/train_data.txt")

#This dataset consists of the vocal frequencies of the various subjects who
have or dont have Parkinsons. It consists of the voice sampels of sustained
vowels, numbers from 1 to 10 and short sentences or words. On basis of this
```

the parkinsons patient can be identified. There are also labels provided to match with the predicted values. Here using the number '1' for all parkinsons patients, I am trying to find a seasonal model IF it exists for this dataset.

#Link to the dataset:

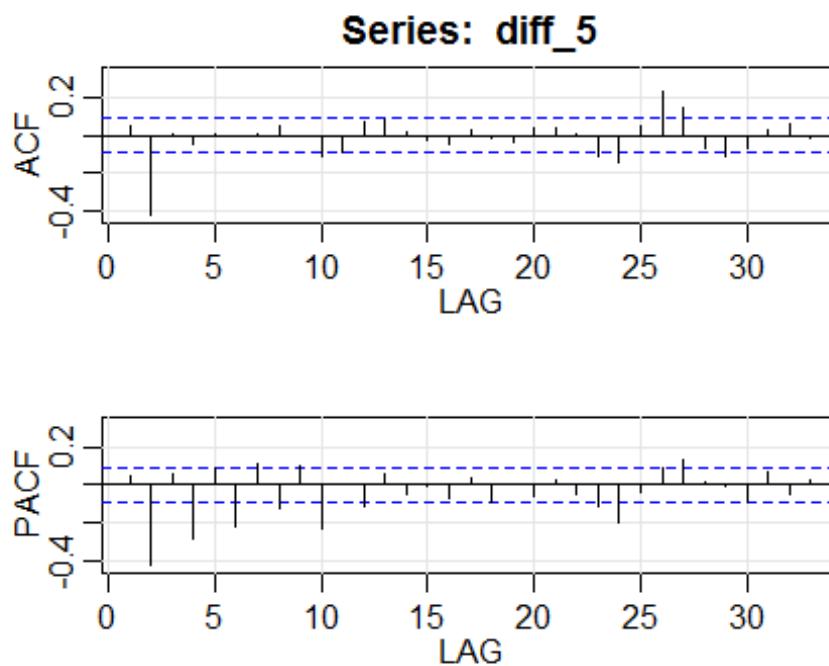
<https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings>

#Obtaining data for subject 1

```
subject1 <- ts(data[1:519,][,5])
```

```
diff_5 <- diff(subject1,2)
```

```
acf2(diff_5)
```



##		ACF	PACF
##	[1,]	0.04	0.04
##	[2,]	-0.42	-0.43
##	[3,]	0.00	0.06
##	[4,]	-0.05	-0.29
##	[5,]	0.01	0.09
##	[6,]	-0.01	-0.22
##	[7,]	0.00	0.11
##	[8,]	0.05	-0.12
##	[9,]	-0.01	0.10
##	[10,]	-0.11	-0.24
##	[11,]	-0.10	0.00
##	[12,]	0.07	-0.11
##	[13,]	0.08	0.06

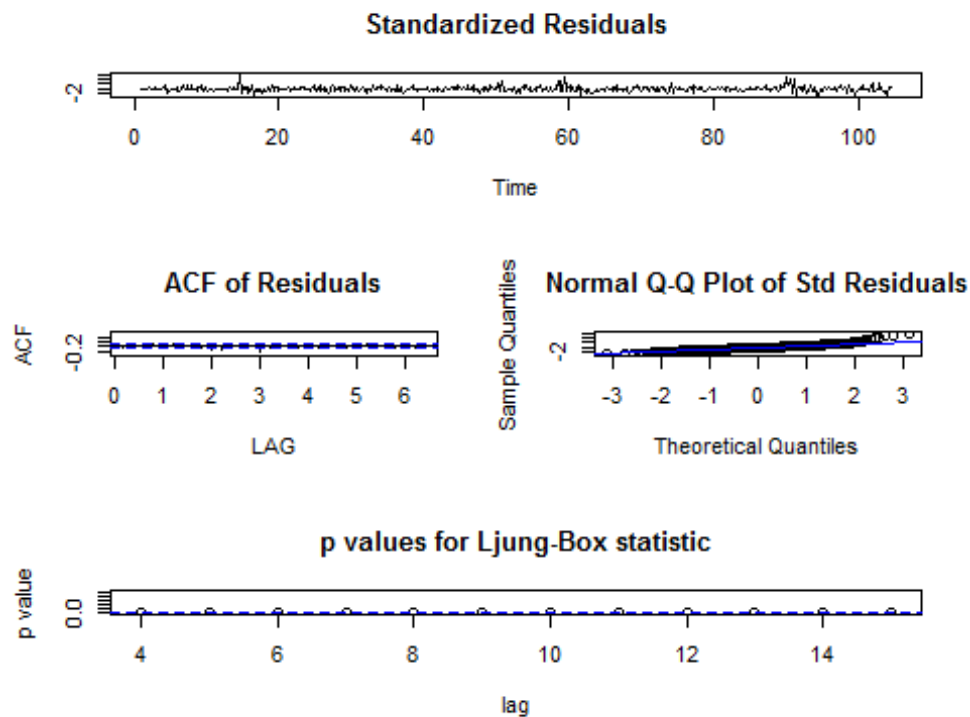
```

## [14,]  0.01 -0.06
## [15,] -0.03  0.00
## [16,] -0.05 -0.07
## [17,]  0.03  0.03
## [18,] -0.01 -0.10
## [19,] -0.04  0.01
## [20,]  0.04 -0.07
## [21,]  0.03  0.03
## [22,]  0.00 -0.05
## [23,] -0.11 -0.11
## [24,] -0.15 -0.20
## [25,]  0.05 -0.04
## [26,]  0.23  0.09
## [27,]  0.14  0.13
## [28,] -0.07  0.02
## [29,] -0.11 -0.01
## [30,] -0.07 -0.10
## [31,]  0.02  0.07
## [32,]  0.06 -0.05
## [33,] -0.02  0.02

subject1<-ts(subject1, freq=5)
mod1<-sarima(subject1, 1,0,0,2,1,0,5)

## initial  value 0.236458
## iter    2 value 0.101278
## iter    3 value 0.075097
## iter    4 value 0.071503
## iter    5 value 0.071453
## iter    6 value 0.071437
## iter    7 value 0.071437
## iter    8 value 0.071436
## iter    8 value 0.071436
## iter    8 value 0.071436
## final   value 0.071436
## converged
## initial  value 0.064001
## iter    2 value 0.063972
## iter    3 value 0.063969
## iter    4 value 0.063969
## iter    5 value 0.063968
## iter    5 value 0.063968
## iter    5 value 0.063968
## final   value 0.063968
## converged

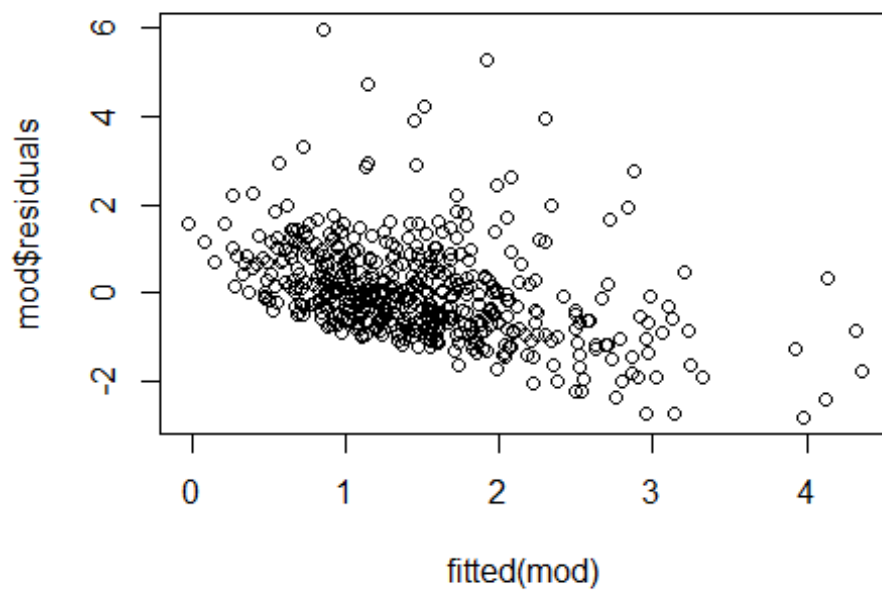
```



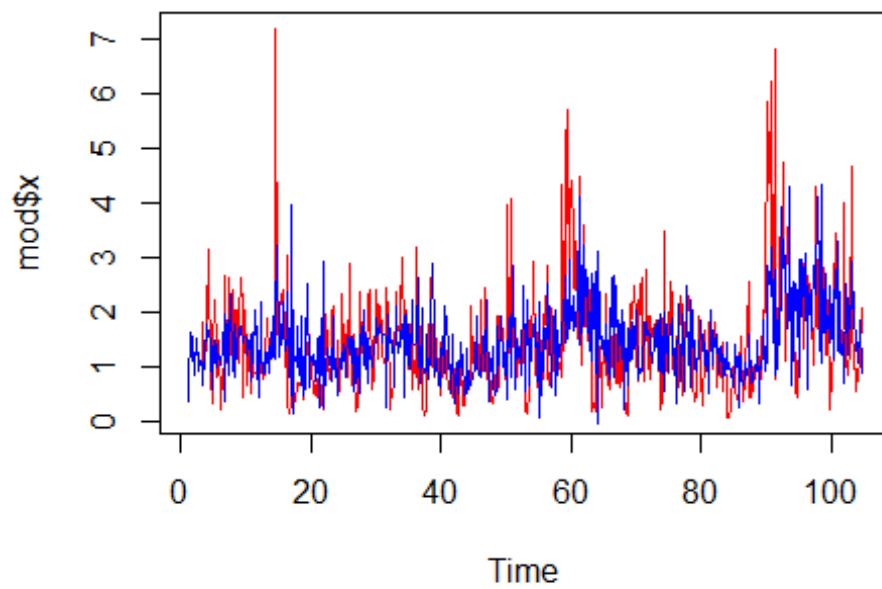
```
mod<-Arima(subject1,order=c(1, 0, 0),
            seasonal=list(order=c(2, 1, 0), period=12))
mod

## Series: subject1
## ARIMA(1,0,0)(2,1,0)[12]
##
## Coefficients:
##          ar1      sar1      sar2
##      0.3918 -0.5320 -0.3193
## s.e.  0.0412  0.0423  0.0431
##
## sigma^2 estimated as 1.189:  log likelihood=-764.19
## AIC=1536.39  AICc=1536.47  BIC=1553.3

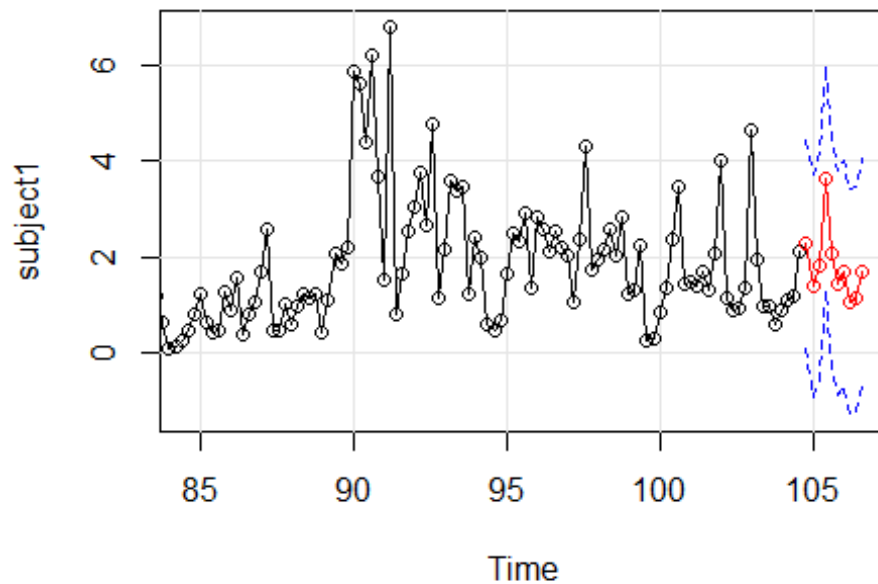
#plot model
plot(fitted(mod), mod$residuals)
```



```
plot(mod$x, col='red')  
lines(fitted(mod), col='blue')
```



```
#forecasting
sarima.for(subject1, 10, 1,0,0,2,1,0,12)
```



```
## $pred
## Time Series:
## Start = c(104, 5)
## End = c(106, 4)
## Frequency = 5
## [1] 2.267552 1.395113 1.812963 3.648007 2.056248 1.445198 1.665387
## [8] 1.042814 1.116152 1.690323
##
## $se
## Time Series:
## Start = c(104, 5)
## End = c(106, 4)
## Frequency = 5
## [1] 1.087054 1.167454 1.179302 1.181109 1.181386 1.181429 1.181435
## [8] 1.181436 1.181436 1.181436

predict(mod, n.ahead=10)

## $pred
## Time Series:
## Start = c(104, 5)
## End = c(106, 4)
## Frequency = 5
## [1] 2.254919 1.377717 1.793613 3.628133 2.035843 1.424627 1.644715
```

```
## [8] 1.022193 1.095550 1.669661
##
## $se
## Time Series:
## Start = c(104, 5)
## End = c(106, 4)
## Frequency = 5
## [1] 1.090356 1.171046 1.182943 1.184759 1.185037 1.185080 1.185087
## [8] 1.185088 1.185088 1.185088
```

#Taking a look at the plot, it is seen that the seasonal model is able to find the patterns which would repeat for the patients with parkinsons disease in voice data for number '1'.

#Thus using this, I was able to set a prediction for the future patient voice data. Seasonal model can also be applied for gait data as there are patterns which might repeat in case of either healthy or parkinsons patients.