Chickenpox in Budapest

Achol Kiir s3569270, Akanksha Sharma s3807428, Anirudhda Pardhi s3807109,

Ghina Khallouf s3491985, Harsha Chandar Thammaredy s3801573, Kushagra Singh s3820789

RMIT

Semester 1, 2021

Table of Contents

Introduction

Chickenpox, also known as varicella, is a disease consisting of itchy red blister that appears all over the body (Frothingham, 2018). This condition is caused by a virus and most commonly children (The Royal Children's Hospital Melbourne, 2021). In Australia in 2016, there were 395 hospital admissions due to the disease, 26% of which were children under the age of 15 (Frothingham, 2018). Chickenpox is known rarely to infect someone more than once. Fortunately, since the introduction of the chickenpox vaccine in mid-1990's cases have decreased.

The chickenpox vaccine is used to treat chickenpox and it is recommended to take two doses (Better Health Channel, 2017; Department of Health, 2019). It is encouraged that the vaccine is administered to children between the ages of 12 to 15 months and a booster for children between 4 and 6 years old (The Royal Children's Hospital Melbourne, 2021). Older children and adults who missed vaccination or have been exposed are able to have catch up vaccination given that the disease is more severe as we get older (Better Health Channel, 2017). Due to its contagious behaviour, it is highly recommended to avoid contact with infected people. Chickenpox outbreaks is known to be highest during winter and early spring (Better Health Channel, 2017).

Problem Statement

The aim of this report is to conduct an analysis on seasonal time series data relating to monthly chickenpox cases to understand the trend and pattern of prevalence. The objective is to narrow possible SARIMA candidate models and use the best model to help forecast or predict monthly chickenpox cases over the course of the next 4 years.

Data

The Hungarian Chickenpox Cases data set was sourced from UCI Machine Learning Repository website (Rozemberczki, Scherer, Kiss, Sarkar, & Ferenci, 2021). The original dataset contains 522 observations of total weekly chickenpox cases in Hungarian cities from 2005 to 2015.

The data was filtered to only show cases in the capital city, Budapest, and was transformed to accumulate total cases by month, giving a new total of 121 observations starting from January 2005 and ending January 2015.

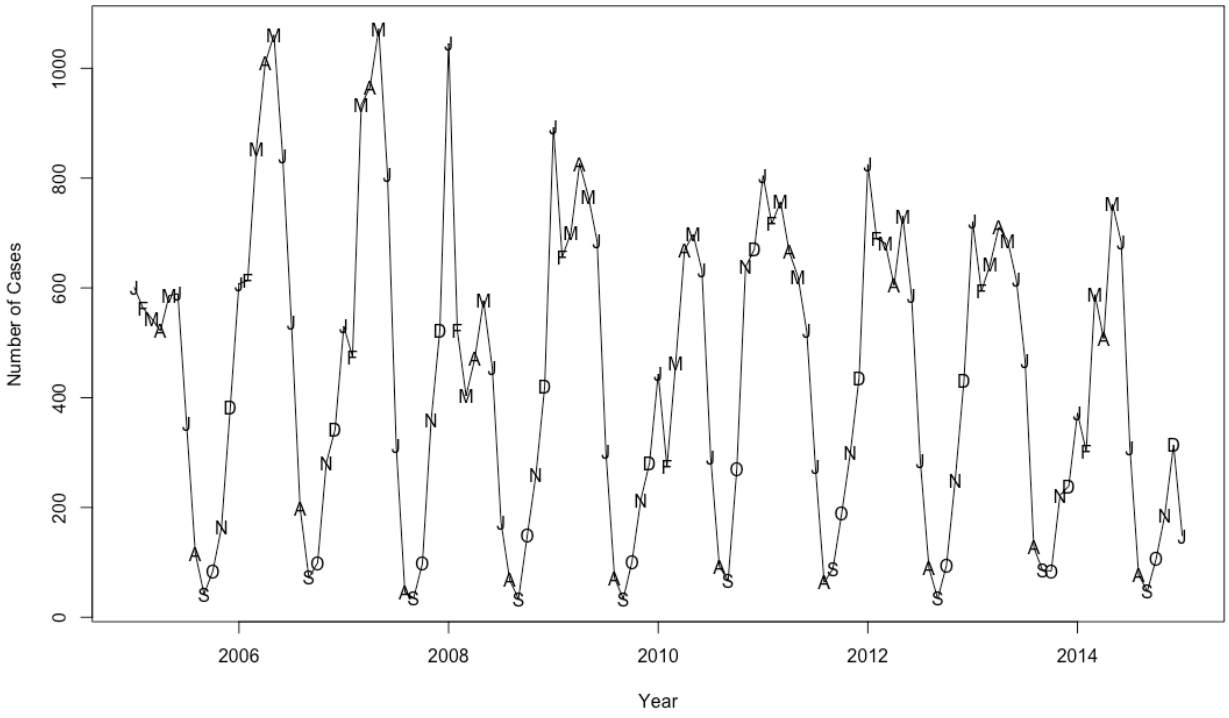
Method

R-Studio Software was used to perform a Time Series Analysis on the data to observe its trend, to fit a model, and to forecast the predicted number of chickenpox cases over the coming years using the selected model.

Results and Analysis

Plotting the Data

Figure 1. Time Series Plot of Number of Chickenpox Cases in Budapest

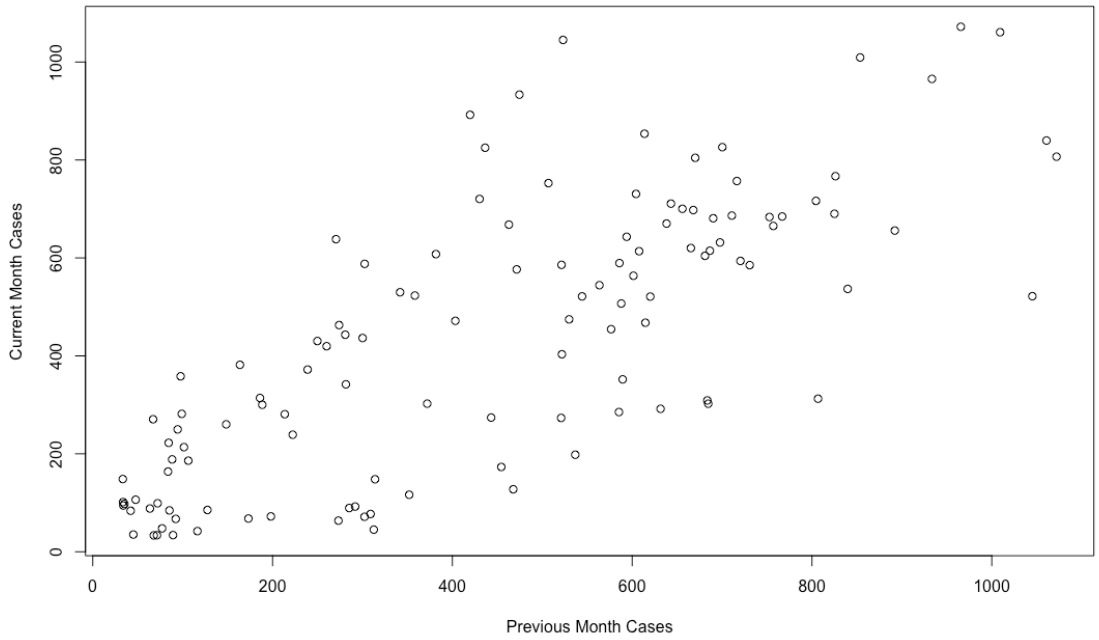


- Trend- Gradual downward trend can be observed clearly from the time series plot in *Figure 1*.
- Seasonality- From the time series plot in *Figure 1*, there is an observable waving pattern, which is a strong indication of seasonality. The grouping among certain monthly values in the peaks and troughs, with case numbers peaking in the Winter/Spring seasons and low during Summer/Autumn, is also a characteristic of seasonality.
- Changing Variance- There is also a noticeable tapering of variance, indicating that the variance in the data slightly narrows as time progresses.

- Behavior- Since there is seasonality present in the plot, it tends to screen out the othe (Frothingham, 2018)r important properties/features of the series. Hence, no comment can be made on AR (Auto Regressive) and MA (Moving average) behavior.
- Change point- No change point can be seen.

The scatterplot in *Figure 2* provides a visual of the relationship between the number of chickenpox cases in a given month and the case numbers of the previous month. There is an apparent moderate to strong positive correlation between neighbouring months. After further analysis, this observation is supported by a correlation coefficient of 0.764 and concludes a moderate-strong positive correlation between these values.

Figure 2. Scatterplot of Chickenpox Numbers of Consecutive Months



Trend Model

Due to the waving pattern in time series plot given in *Figure 1*, both linear and quadratic models were considered unsuitable trend models. To confirm that the data fits a seasonal trend, two seasonal models were tested, one excluding the intercept term and the other including the intercept term. A final harmonic or cosine model was also tested as a possible trend model.

The seasonal model excluding intercept term, shown in *Figure 3*, shows that the coefficients for all months (excluding September) are significant at the $\alpha = 0.05$ level. This model is an overall strong candidate with $R^2 = 0.93$, $F_{12,109} = 119.2$, $p < 0.001$.

Figure 3. Seasonal Trend Model: excluding Intercept Term

```
Call:
lm(formula = chickenpox.ts ~ month. - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-487.39  -57.27  -19.90   43.47  409.75

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
month.January    635.39      43.77  14.518  < 2e-16 ***
month.February   540.64      45.90  11.778  < 2e-16 ***
month.March      656.67      45.90  14.306  < 2e-16 ***
month.April      694.93      45.90  15.139  < 2e-16 ***
month.May        754.99      45.90  16.448  < 2e-16 ***
month.June       641.10      45.90  13.967  < 2e-16 ***
month.July       330.39      45.90   7.198 8.26e-11 ***
month.August      94.89      45.90   2.067  0.04109 *
month.September   53.90      45.90   1.174  0.24286
month.October    127.46      45.90   2.777  0.00647 **
month.November   287.40      45.90   6.261 7.75e-09 ***
month.December   403.71      45.90   8.795 2.37e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 145.2 on 109 degrees of freedom
Multiple R-squared:  0.9292,    Adjusted R-squared:  0.9214
F-statistic: 119.2 on 12 and 109 DF,  p-value: < 2.2e-16
```

The seasonal model including intercept term, given in *Figure 4*, confirms that this model does not fit the data as efficiently as the model that excludes the intercept term. The coefficients from February to June are not significant at the $\alpha = 0.05$ level. When compared to the R^2 of the previous model, 0.93, this model reduces this value to $R^2 = 0.75$. This model is ruled out and it can be concluded that the first seasonal trend model (excluding the intercept) is still the strongest candidate.

Figure 4. Seasonal Trend Model: including Intercept Term

```
Call:
lm(formula = chickenpox.ts ~ month.)

Residuals:
    Min      1Q  Median      3Q     Max
-487.39  -57.27  -19.90   43.47  409.75

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      635.39      43.77  14.518  < 2e-16 ***
month.February   -94.75      63.42  -1.494   0.1381
month.March       21.28      63.42   0.336   0.7379
month.April       59.54      63.42   0.939   0.3499
month.May        119.60      63.42   1.886   0.0620 .
month.June         5.71      63.42   0.090   0.9284
month.July      -305.00      63.42  -4.809 4.89e-06 ***
month.August    -540.50      63.42  -8.522 9.77e-14 ***
month.September -581.49      63.42  -9.168 3.37e-15 ***
month.October   -507.93      63.42  -8.009 1.38e-12 ***
month.November  -347.99      63.42  -5.487 2.69e-07 ***
month.December  -231.68      63.42  -3.653   0.0004 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 145.2 on 109 degrees of freedom
Multiple R-squared:  0.7546,    Adjusted R-squared:  0.7298
F-statistic: 30.46 on 11 and 109 DF,  p-value: < 2.2e-16
```

To verify that the data follows a seasonal trend and not a cosine trend, a harmonic trend model was also tested, and the results are given in *Figure 5*. All coefficients are statistically significant at the $\alpha = 0.05$ level. However, as $R^2 = 0.66$, this trend can also be ruled out. Thus, it can be concluded that the data does indeed follow a seasonal trend, where the seasonal model (excluding the intercept term) best fits the data.

Figure 5. Harmonic/Cosine Trend Model

```
Call:
lm(formula = chickenpox.ts ~ har.)

Residuals:
    Min      1Q  Median      3Q     Max
-396.35  -85.22  -31.96   77.63  515.21

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      436.00      14.97  29.117  < 2e-16 ***
har.cos(2*pi*t)   93.93      21.09   4.454 1.93e-05 ***
har.sin(2*pi*t)  306.00      21.26  14.391  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 164.7 on 118 degrees of freedom
Multiple R-squared:  0.6579,    Adjusted R-squared:  0.6521
F-statistic: 113.5 on 2 and 118 DF,  p-value: < 2.2e-16
```

Transformation

To test the effect of the changing variance, as observed in the time series plot in Figure 1, a log transformation and first differencing was applied to determine whether this would reduce the change in variance, confirm a stationary series and verify whether an ARCH/GARCH model is required.

From the plot in *Figure 6*, applying the transformation eliminates major changes in variance and stabilizes the series, making it stationary. This conclusion is supported by the Dickey-Fuller Test outcome in *Figure 7* that confirms a stationary series, $p = 0.01$.

Figure 6. Stationary Variance Plot

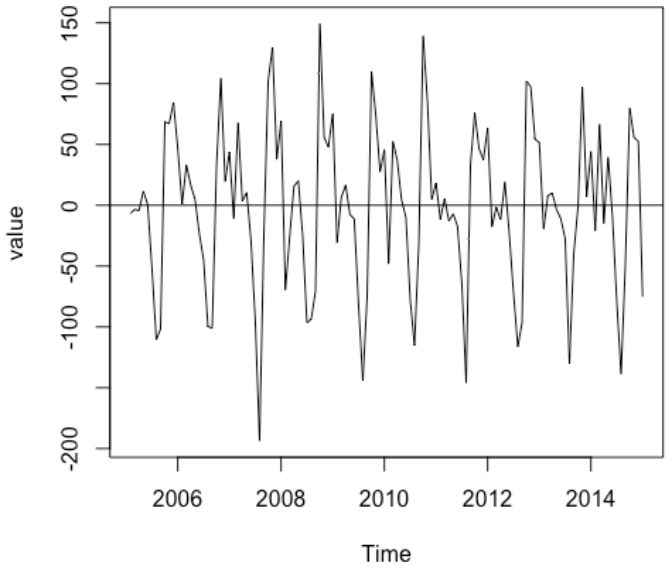
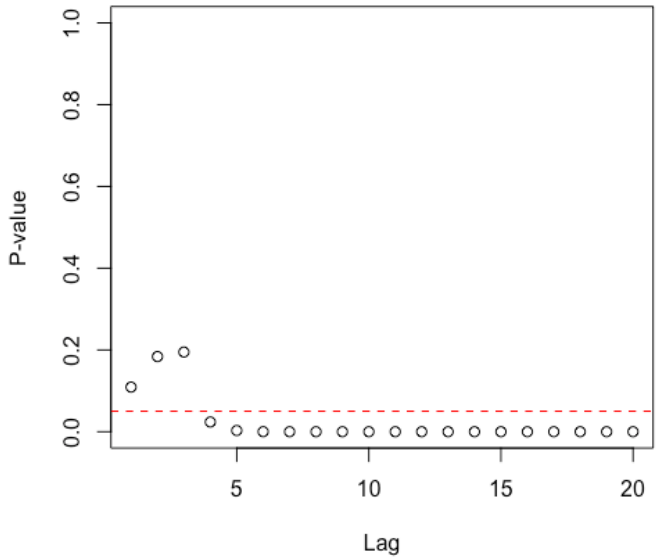


Figure 7. Dickey-Fuller Test for Stationarity

```
        Augmented Dickey-Fuller Test

data:  r.pox
Dickey-Fuller = -6.2111, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Figure 8. McLeod Li Test Plot



As seen in *Figure 8*, McLeod Li test concludes that there is no threat of volatility clustering in the transformed data due to the non-significant values at lags 1 to 3.

From these results, it can be concluded that an ARCH/GARCH model is not required.

Residual Approach

The residual approach is chosen over classical approach because its more consistent with the results and it makes it easier to find appropriate models. From *Figure 9*, a slowly decaying pattern just at the periods can be observed.

Figure 9. Autocorrelation Function (ACF) Plot without Differencing

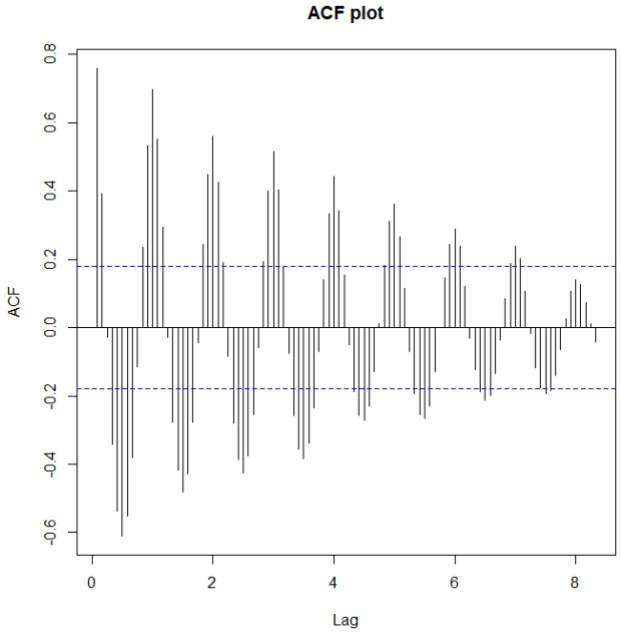
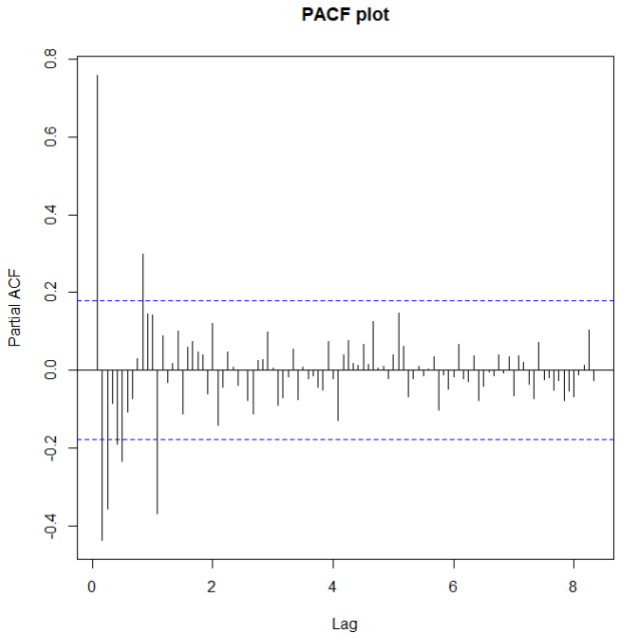


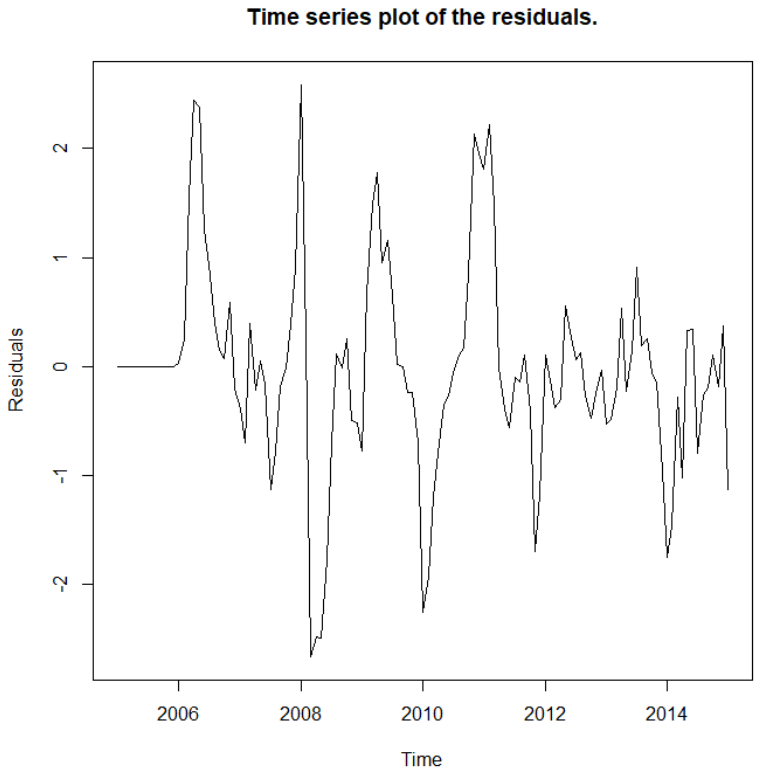
Figure 10. Partial Autocorrelation Function (PACF) Plot without Differencing



Model Identification and Specification

Given that the need for ARCH/GARCH was found unfitting, and the confirmation that the data follows a seasonal trend, it was determined that a SARIMA model would best fit the data. Hence, SARIMA (0,0,0) * (0,1,0) model was fitted where value for the first seasonal difference i.e., D = 1 and will proceed with the inspection of residuals.

Figure 11. Time Series Plot of the residuals for SARIMA (0,0,0) * (0,1,0)



From *Figure 11*, the effect of seasonal trend has been filtered out after applying first seasonal differencing and the behaviour of seasonal component is clear.

Figure 12. Autocorrelation Function (ACF) Plot SARIMA (0,0,0) * (0,1,0)

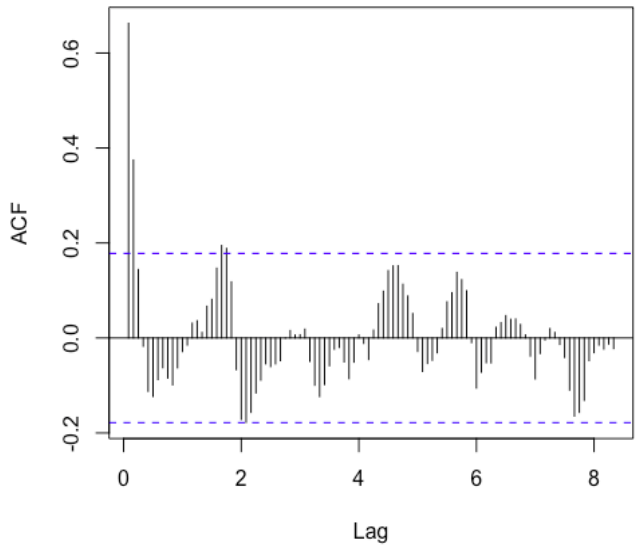
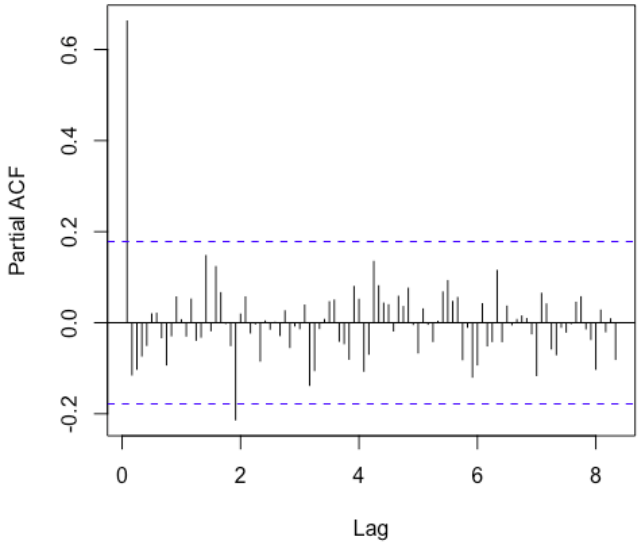


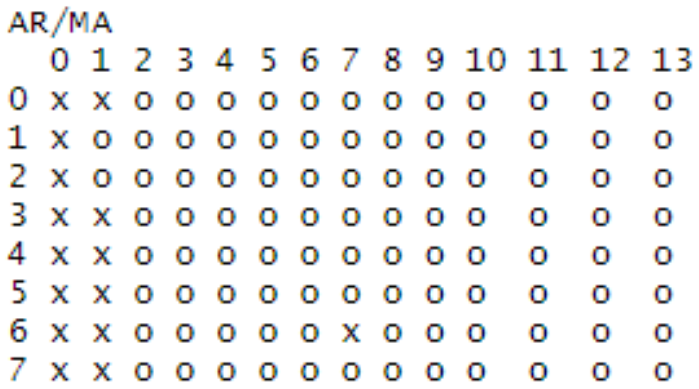
Figure 13. Partial Autocorrelation Function (PACF) Plot SARIMA (0,0,0) * (0,1,0)



In *Figure 12*, there is 1 significant value that is crossing the significant boundary between 0 to 1, thus, giving the value of q =1. There is also one seasonal lag at 2, therefore, giving the value of Q = 1.

Similarly in *Figure 13*, we observe no significant partial autocorrelation value touching or crossing the significant boundaries at lag 1 to 4, therefore, there is no seasonal lag is left, giving P = 0.

Figure 14. Extended Autocorrelation Function Results SARIMA (0,0,0) * (0,1,0)



From *Figure 14*, after employing EACF, ACF, and PACF test to figure out the best SARIMA models, we select two best candidate model sets:

1. SARIMA (1,0,1) * (0,1,1)
2. SARIMA (1,0,2) * (0,1,1)

Model Fitting

SARIMA Model (1,0,1)x(0,1,1)12

Figure 15. SARIMA (1,0,1) * (0,1,1)12 Model

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ar1   0.60314    0.10649  5.6637 1.481e-08 ***
ma1   0.27831    0.12790  2.1759 0.0295628 *
sma1 -0.75000    0.20789 -3.6077 0.0003089 ***
sma2 -0.24999    0.12423 -2.0123 0.0441865 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results of fitting the SARIMA(1,0,1)x(0,1,1)12 model are given in *Figure 15*. We can observe that the *p*-value of all the coefficients are less than 0.05, which means that the coefficient estimates are significant at the $\alpha = 0.05$ level.

SARIMA Model (1,0,2)x(0,1,1)12

Figure 16. SARIMA (1,0,2) * (0,1,1)12 Model

```
z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1   0.557904   0.159476  3.4983 0.0004681 ***
ma1   0.324373   0.176512  1.8377 0.0661091 .
ma2   0.060863   0.139984  0.4348 0.6637194
sma1 -0.756797   0.210031 -3.6033 0.0003142 ***
sma2 -0.243184   0.124735 -1.9496 0.0512236 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results of fitting the SARIMA (1,0,2) * (0,1,1)12 model is given in *Figure 16.* We can observe that the *p*-values of ma1, ma2 and sma2 all exceed the α = 0.05 level of significance, which means estimates of these coefficients are not significant.

Of the two SARIMA models fitted, the first SARIMA model (1,0,1) * (0,1,1)12 is the best choice by looking at the z test of coefficients.

Diagnostic Checking

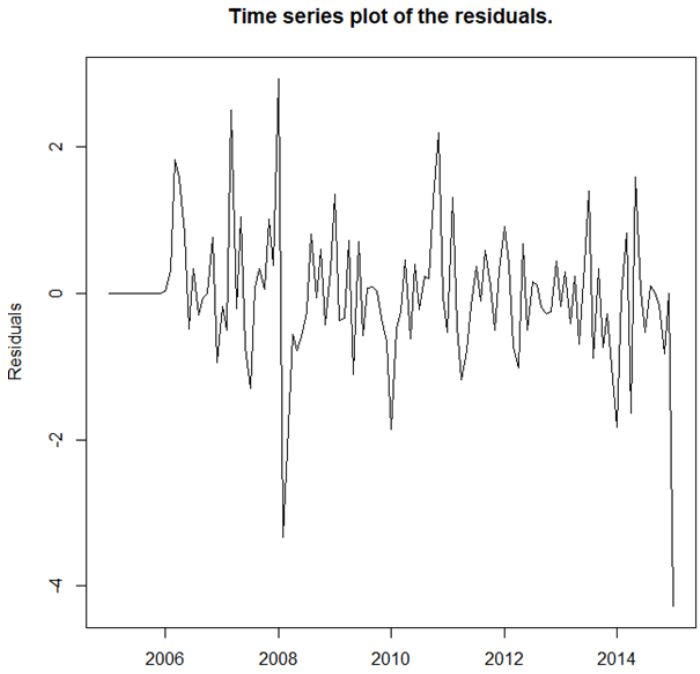
We use the goodness of fit parameters to check the efficiency of model selected SARIMA model.

Figure 17. Summary of Selected SARIMA (1,0,1) * (0,1,1)12 Model

```
Series: chickenpox.ts
ARIMA(1,0,1)(0,1,1)[12]

Coefficients:
         ar1     ma1     sma1
      0.6102  0.2581  -0.8852
s.e.  0.1088  0.1280   0.2243

sigma^2 estimated as 12014:  log likelihood=-674.16
AIC=1356.33   AICc=1356.71   BIC=1367.09

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set -5.021897 102.5901 67.13901 -3.373768 21.60144 0.4799373 0.0132186
```

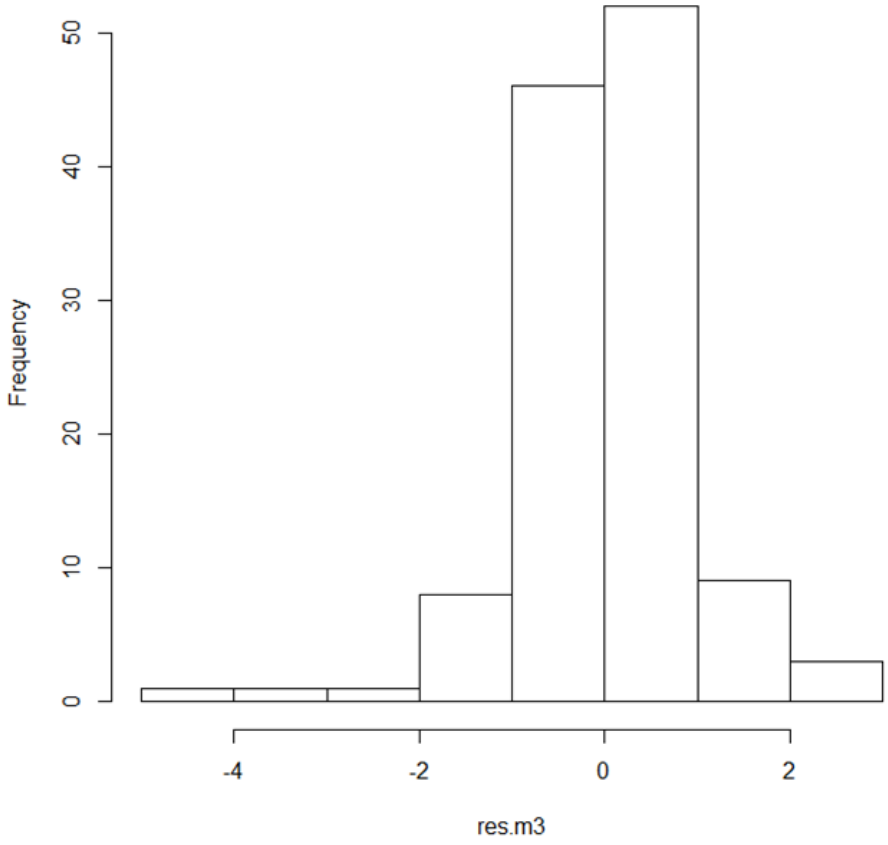
As observed in *Figure 17*, our model parameter values, ME, RMSE, MAE, MPE, MAPE and MASE, all have quite low values which suggests that our selected model fits well.

Figure 18. Time Series Plot of Residuals of Selected SARIMA (1,0,1) * (0,1,1)12 Model



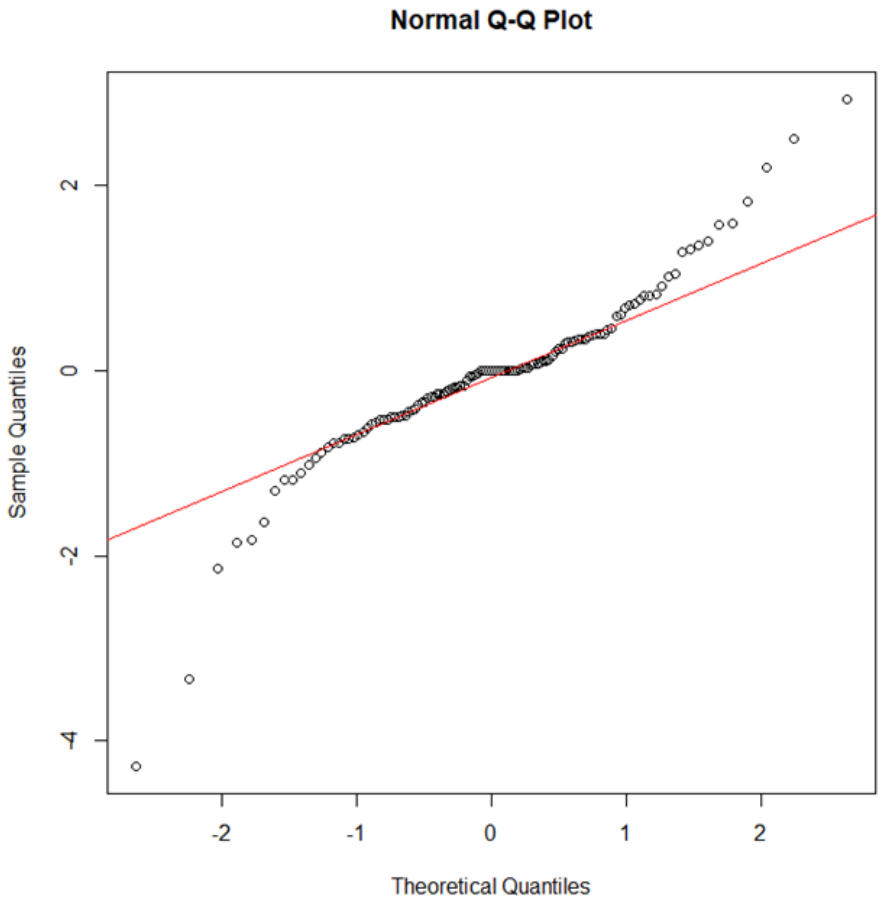
In the residual time series plot of the selected SARIMA model given in *Figure 18*, there are no observable trends or irregularities and it can be seen with respect to baseline x-axis(x=0), that the data points are located above and below randomly.

Figure 19. Histogram of Residuals of Selected SARIMA (1,0,1) * (0,1,1)12 Model



The histogram seen in *Figure 19* suggests that residual may seem symmetric, however it is slightly skewed to the left making it difficult to assume normality. Further analysis is conducted to conclude whether the residuals fit a normal distribution.

Figure 20. Normal Q-Q Plot for Residuals of Selected SARIMA (1,0,1) * (0,1,1)12 Model



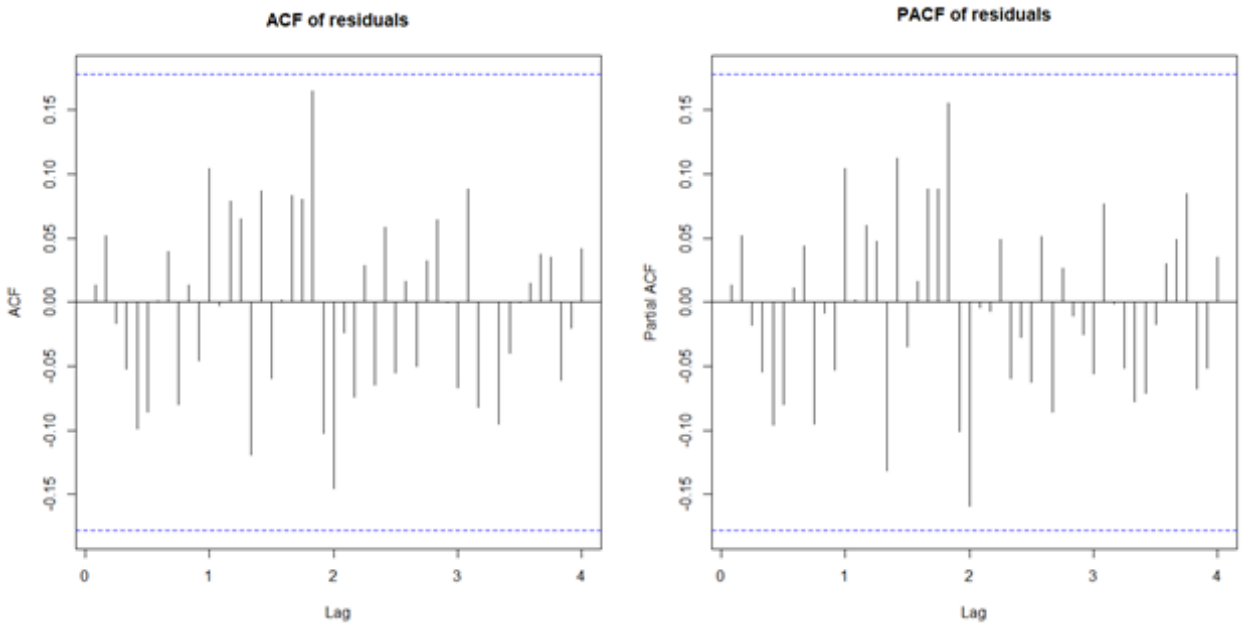
In QQ plot, *Figure 20*, we can observe that data points are deviating away from both extreme ends of line which indicates that the residuals stray away from normality.

Figure 21. Shapiro-Wilk Test for Normality of Residuals of Selected SARIMA (1,0,1) * (0,1,1)12 Model

```
> shapiro.test(res.m3)

        Shapiro-Wilk normality test

data:  res.m3
W = 0.91402, p-value = 1.007e-06
```

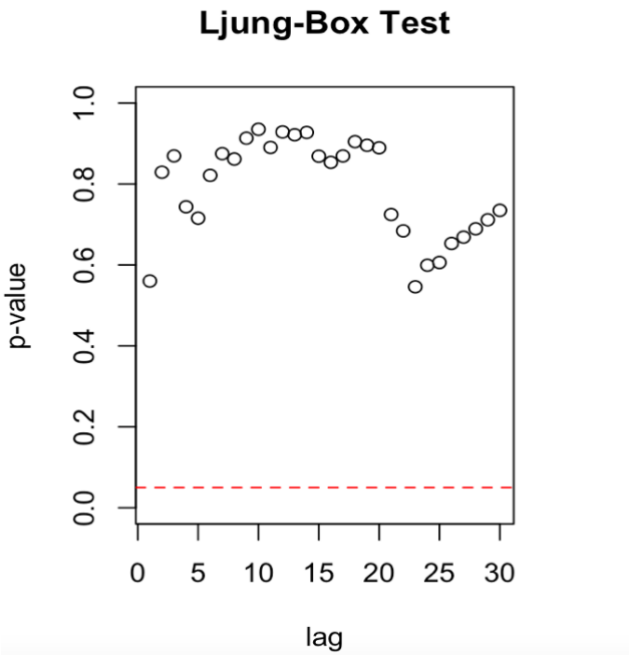
Shapiro-Wilk normality test results, given in *Figure 21*, deem that the distribution of residuals are significantly different from a normal, $p < 0.001$, therefore we cannot assume that they fit a normal distribution. However, according to the Central Limit Theorem, since the number of observations exceed 30, we can assume normality. Although we have white noise residuals, the large valued residuals make it impossible to conclude the normality of residuals by either the QQ plot or Shapiro test at a 5% level of significance.

Figure 22. ACF and PACF Plots of Residuals of Selected SARIMA (1,0,1) * (0,1,1)12 Model



From *Figure 22*, according to ACF and PACF plots, there is no significant lag or autocorrelation components left between 0 and 1 as there are no values touching or crossing the significant boundaries. Similarly, if we observe lags after 1, there is no correlation value touching or crossing the significant boundaries at lag 1 to 4, so we can say that no seasonal lag is left.

Figure 23. Box-Ljung Test of Selected SARIMA (1,0,1) * (0,1,1)12 Model




```
        Box-Ljung test

data:  window(rstandard(m3.cpox), start = c(2005, 2))
X-squared = 16.809, df = 22, p-value = 0.7738
```

The Box-Ljung test uses the following hypotheses:

H_0: The residuals are independently distributed.

H_A: The residuals are not independently distributed; they exhibit serial correlation

Ideally, we would like to fail to reject the null hypothesis. That is, we would like to see the *p*-value of the test be greater than 0.05 because this means the residuals for our time series model are independent, which is often an assumption we make when creating a model.

As per Ljung box test on Figure 23, the given *p*-value = 0.77 which is greater than significant level 0.05 there seems to be no independence of error. Hence, we can say residuals are white noise.

Predictions

Based on selected SARIMA(1,0,1)x(0,1,1)12 model, forecasting is executed to predict the trend over the next 4 years, 2015 to 2019. The forecast model in *Figure 24* shows point values with lower and higher bound values and goodness of prediction can be seen in internal spikes within the forecasting slopes. The predicted case numbers for the next 10 years are given in *Figure 25*. As per forecasting pattern, we can say that there will be decrease in number of chickenpox patients of Budapest city.

Figure 24. Plot of Time Series with 4 Year Forecast based on SARIMA(1,0,1)x(0,1,1)12

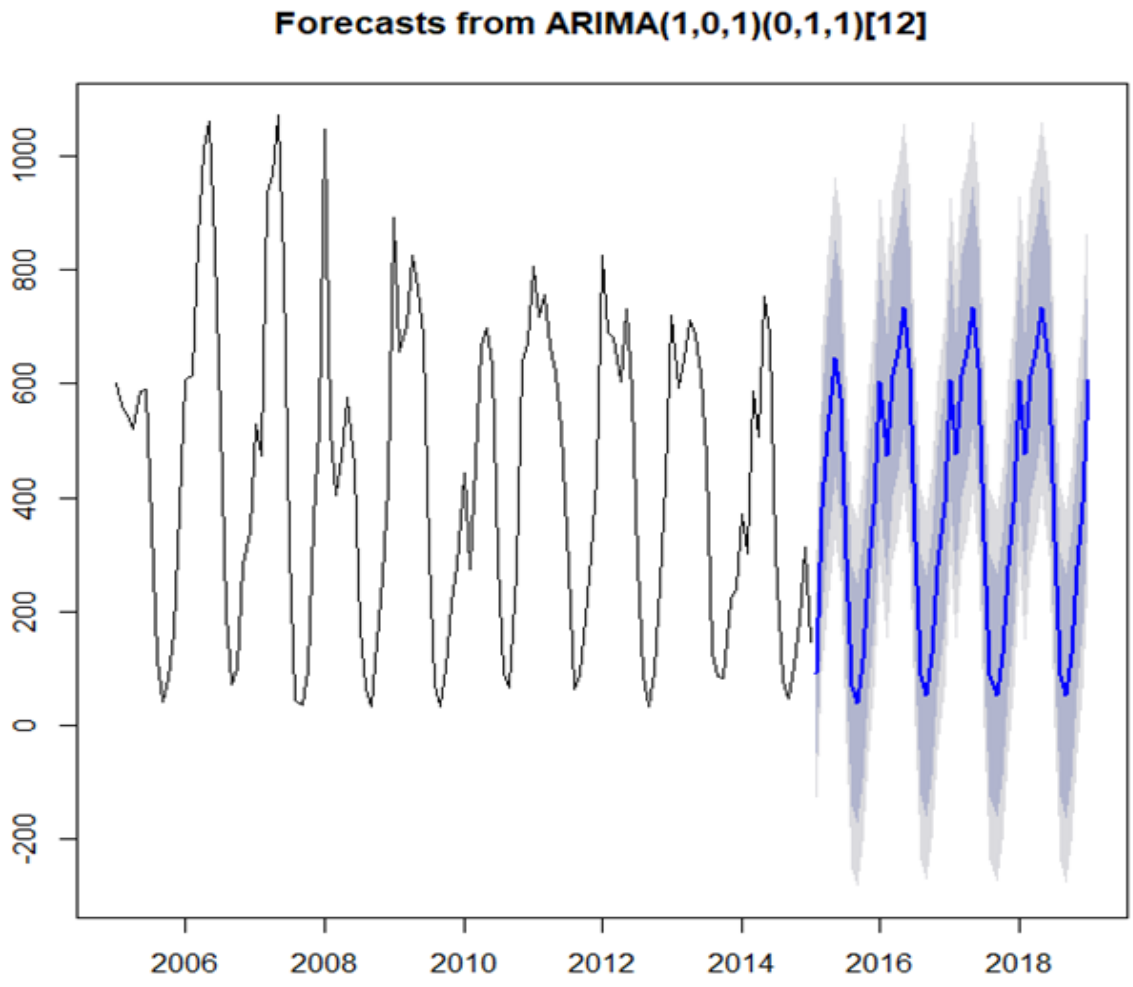


Figure 25. 10 Month Prediction of Chickenpox Cases of Selected SARIMA (1,0,1) * (0,1,1)12 Model

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Feb 2015	91.28569	-50.64554	233.2169	-125.7793979	308.3508
Mar 2015	380.99795	193.04827	568.9476	93.5537180	668.4422
Apr 2015	516.11331	313.68608	718.5405	206.5275835	825.6990
May 2015	645.68286	438.12362	853.2421	328.2483875	963.1173
Jun 2015	575.40451	365.96713	784.8419	255.0976817	895.7113
Jul 2015	292.51576	82.38449	502.6470	-28.8522790	613.8838
Aug 2015	70.26964	-140.11655	280.6558	-251.4882665	392.0275
Sep 2015	40.04714	-170.42629	250.5206	-281.8441924	361.9385
Oct 2015	116.85126	-93.63427	327.3368	-205.0585845	438.7611
Nov 2015	275.20456	64.76924	485.6399	-46.6284884	597.0376

Conclusion

The Chickenpox Cases data is confirmed to follow a seasonal trend, with case numbers peaking in the Winter/Spring seasons and low during Summer/Autumn. The variance in the data was resolved using transformation which concluded no requirement of an ARCH/GARCH model. Multiple SARIMA models were tested to verify which one best fit the data. SARIMA(1,0,1)*(0,1,1)12 model was chosen and was used to forecast a prediction of cases over the next 4 years, from 2015 to 2019. It can be concluded that there will be a gradual decrease of chickenpox cases in Budapest over the coming years.

References

Better Health Channel. (2017). *Chickenpox - immunisation*. Retrieved from Better Health Channel: https://www.betterhealth.vic.gov.au/health/healthyliving/chickenpox-immunisation

Department of Health. (2019). *Varicella(Chickenpox)*. Retrieved from Australian Government: https://immunisationhandbook.health.gov.au/vaccine-preventable-diseases/varicella-chickenpox

Frothingham, S. (2018). *Chickenpox in Adults*. Retrieved from Healthline: https://www.healthline.com/health/chickenpox-in-adults

Rozemberczki, B., Scherer, P., Kiss, O., Sarkar, R., & Ferenci, T. (2021). *Hungarian Chickenpox Cases Data Set*. Retrieved from UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/Hungarian+Chickenpox+Cases

The Royal Children's Hospital Melbourne. (2021). *Chickenpox*. Retrieved from The Royal Children's Hospital Melbourne: https://www.rch.org.au/kidsinfo/fact_sheets/Chickenpox_Varicella/

Appendix: R Code

```
library(tidyr)
library(dplyr)
library(TSA)
library(tseries)
library(forecast)
library(wktmo)
library(lmtest)
library(LBQPlot)
library(FitAR)

test <- hungary_chickenpox$BUDAPEST
monthlyData <- weekToMonth(test, year = 2005, wkIndex = 1, wkMethod = "ISO")
monthlyData <- weekToMonth(test, datStart = "03-01-2005", wkMethod = "startDat")

monthlyData <- monthlyData %>% select(value)

class(monthlyData)

chickenpox.ts<- ts(monthlyData, start = c(2005, 1), end = c(2015, 1), frequency = 12)

class(chickenpox.ts)

plot(chickenpox.ts, type = 'o', ylab = 'Number of Cases', xlab = 'Year',
    main = 'Time Series Plot of Number of Chickenpox Cases in Budapest')

plot(window(chickenpox.ts,start=c(2005,1)),ylab='Number of Cases', xlab = 'Year', main ="'Time
Series Plot of Number of Chickenpox Cases in Budapest")
Month=c('J','F','M','A','M','J','J','A','S','O','N','D')
points(window(chickenpox.ts,start=c(2005,1)),pch=Month)

plot(y=chickenpox.ts,x=zlag(chickenpox.ts),ylab='Current Month Cases', xlab='Previous Month
Cases',
    main = "Scatter plot of case numbers of consecutive months")

y = chickenpox.ts
x = zlag(chickenpox.ts)
index = 2:length(x)
cor(y[index],x[index])

## Seasonal

month.= season(chickenpox.ts) # period added to improve table display and this line sets up
indicators
model1=lm(chickenpox.ts~month.-1) # -1 removes the intercept term
```

```
summary(model1)

model2=lm(chickenpox.ts~month.)
summary(model2)

## Harmonic/Cosine

har.=harmonic(chickenpox.ts,1)
model3=lm(chickenpox.ts~har.)
summary(model3)

## Stationarity

returnSeries=diff(log(chickenpox.ts))*100

head(returnSeries)

McLeod.Li.test(y=returnSeries)

r.pox=diff(log(chickenpox.ts))*100
plot(r.pox,main="Chickenpox Returns")
abline(h=0)

McLeod.Li.test(y=r.pox)

adf.test(r.pox)
acf(r.pox)
pacf(r.pox)
eacf(r.pox)

qqnorm(r.pox)
qqline(r.pox)
shapiro.test(r.pox)

## Model Specification

acf(chickenpox.ts, lag.max = 100, main="ACF plot")
pacf(chickenpox.ts, lag.max = 100, main="PACF plot")


m1.cpox = arima(chickenpox.ts,order=c(0,0,0),
            seasonal=list(order=c(0,1,0), period=12))

res.m1 = rstandard(m1.cpox)
plot(res.m1,xlab='Time',ylab='Residuals',
    main="Time series plot of the residuals.")
```

```
acf(res.m1, lag.max = 100, main="ACF of the residuals after fitting the first seasonal
difference")
pacf(res.m1, lag.max = 100, main="PACF of the residuals after fitting the first seasonal
difference")

m2.cpox = arima(chickenpox.ts,order=c(0,0,0),
            seasonal=list(order=c(0,1,2), period=12))
res.m2 = rstandard(m2.cpox)
plot(res.m2,xlab='Time',ylab='Residuals',
    main="Time series plot of the residuals.")
acf(res.m2, lag.max = 100, main="ACF of the residuals after fitting the first seasonal
difference")
pacf(res.m2, lag.max = 100, main="PACF of the residuals after fitting the first seasonal
difference")

adf.test(res.m2)# shows we don't need to look into normal residual models
eacf(res.m2)# For identification of SARIMA models


# Model Fitting we have got SARIMA(1,0,1)x(0,1,1) and SARIMA(1,0,2)x(0,1,1) and see if we
can get white noise on residuals.

m3.cpox = arima(chickenpox.ts,order=c(1,0,1),
            seasonal=list(order=c(0,1,2), period=12))
coeftest(m3.cpox)

m3.cpoxS = arima(chickenpox.ts,order=c(1,0,2),
             seasonal=list(order=c(0,1,2),
                        period=12))
coeftest(m3.cpoxS)


# Forecasting
frc = forecast ( m3.cpox , h=48)

frc

plot(frc)


# Residual Analysis :

res.m3 = rstandard(m3.cpox)

plot(res.m3,xlab='Time',ylab='Residuals',

    main="Time series plot of the residuals.")
```

```
par(mfrow=c(1,2));hist(res.m3)
qqnorm(res.m3);qqline(res.m3,col=2)
shapiro.test(res.m3)
par(mfrow=c(1,2))
acf(res.m3,lag.max=48,main="ACF of residuals")
pacf(res.m3,lag.max=48,main="PACF of residuals")
LBQPlot(res.m3,lag.max=30)
Box.test(window(rstandard(m3.cpox),start=c(2005,2)),lag=22,type='Ljung-Box',fitdf=0)
```