Práctica 4: Procesos ARIMA sin estacionalidad

Series Temporales (42229 Modelización estadística). Máster de Bioestadística

Autor: Santiago Agustin Moncalero

Fecha límite entrega: 7 de Abril 2024

# 1 Introducción

Se seguirá trabajando con la serie temporal de visitantes que han tenido los diversos hospedajes en Bélgica, tanto locales como extranjeros. En esta sección se llevará a cabo un análisis ARIMA sin considerar la temporalidad, con el propósito de buscar un modelo predictivo que pueda ser útil para predecir cómo afectó el COVID-19 y como esto afecto a la cantidad de turistas durante los años 2020, 2021 y 2022.

Para realizar esto, se tomará la serie que hemos estado utilizando y se procederá a su análisis anual mediante la función ‘aggregate’.

## Rows: 348 Columns: 1  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (1): Belgium\_TO\_HO  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

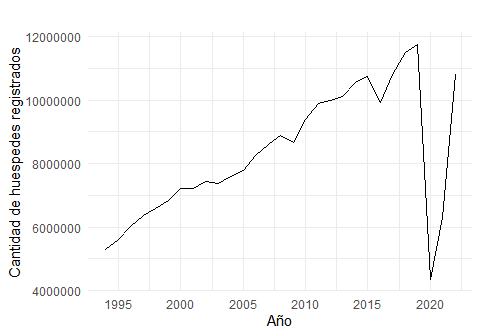


Figure 1.1: Serie temporal anualizada

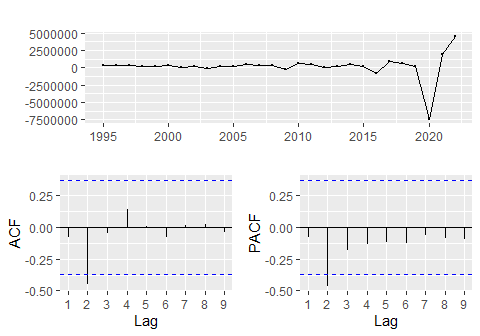
En el gráfico 1.1 se puede ver que es creciente por lo que se deberar diferenciar seguramente, sin embargo se realizara las pruebas de ndiffs.

ndiffs(SerieBelgA)

## [1] 1

# 2 Identificación

ggtsdisplay(diff(SerieBelgA))

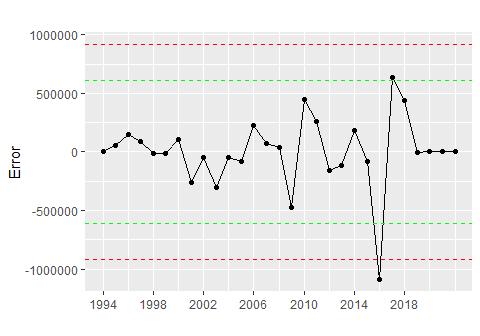


d2020 <- 1\* (time(SerieBelgA) == 2020)  
d2021 <- 1\* (time(SerieBelgA) == 2021)  
d2022 <- 1\* (time(SerieBelgA) == 2022)  
  
summary(auto.arima(SerieBelgA,xreg = cbind(d2020,d2021,d2022)))

## Series: SerieBelgA   
## Regression with ARIMA(0,1,0) errors   
##   
## Coefficients:  
## drift d2020 d2021 d2022  
## 258347.96 -7670244.0 -5962550 -1724047.9  
## s.e. 61185.03 312120.6 449804 560979.2  
##   
## sigma^2 = 109200255508: log likelihood = -393.4  
## AIC=796.8 AICc=799.53 BIC=803.47  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 173.6976 300620.4 185443.9 -0.04340698 2.050258 0.2305851  
## ACF1  
## Training set -0.1777269

arima010 <- Arima(SerieBelgA,   
 order = c(0, 1, 0),  
 include.constant = TRUE,xreg = cbind(d2020,d2021,d2022))  
  
error <- residuals(arima010)  
sderror <- sd(error)  
  
autoplot(error, series="Error",  
 colour = "black",  
 xlab = "",  
 ylab = "Error",  
 main = "") +  
 geom\_hline(yintercept = c(-3, -2, 2, 3)\*sderror,   
 colour = c("red", "green", "green", "red"),   
 lty = 2) +   
 geom\_point() +  
 scale\_x\_continuous(breaks= seq(1990, 2020, 4))

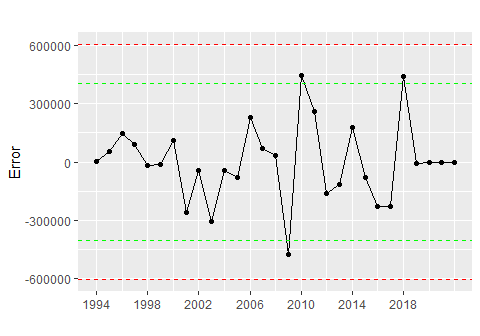
## Scale for x is already present.  
## Adding another scale for x, which will replace the existing scale.



Parece ser que el año 2016 fue un año inusual, al estar relacionado con el atentado que tuvo belgica en el 2016 que afecto al aeropuerto y lo mantovo cerrado sumado a que debio generar menos ganas de hacer turismo.

d2016 <- 1\* (time(SerieBelgA) == 2016)  
  
  
arima010 <- Arima(SerieBelgA,   
 order = c(0, 1, 0),  
 include.constant = TRUE,xreg = cbind(d2016,d2020,d2021,d2022))  
  
error <- residuals(arima010)  
sderror <- sd(error)  
  
autoplot(error, series="Error",  
 colour = "black",  
 xlab = "",  
 ylab = "Error",  
 main = "") +  
 geom\_hline(yintercept = c(-3, -2, 2, 3)\*sderror,   
 colour = c("red", "green", "green", "red"),   
 lty = 2) +   
 geom\_point() +  
 scale\_x\_continuous(breaks= seq(1990, 2020, 4))

## Scale for x is already present.  
## Adding another scale for x, which will replace the existing scale.



# 3 Validación

coeftest(arima010)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## drift 258348 40330 6.4059 0.0000000001495 \*\*\*  
## d2016 -860911 142585 -6.0379 0.0000000015617 \*\*\*  
## d2020 -7670244 205660 -37.2958 < 0.00000000000000022 \*\*\*  
## d2021 -5962550 296436 -20.1141 < 0.00000000000000022 \*\*\*  
## d2022 -1724048 369737 -4.6629 0.0000031178259 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Todas las variables son significativas y explican que:

* Para el 2016, se tuvo 860.991 visitantes menos de lo esperado, esto por el atentado que se tuvo ese año: [Diario El Mundo - Atentados en Brusselas - 2016](https://www.elmundo.es/internacional/2016/03/22/56f0f2cf22601d20498b4648.html)
* Para el 2021, se tuvo 7.670.244 visitantes menos de lo esperado.
* Para el 2022, se tuvo 5.962.550 visitantes menos de lo esperado.
* Para el 2023, se tuvo 1.724.048 visitantes menos de lo esperado.

summary(arima010)

## Series: SerieBelgA   
## Regression with ARIMA(0,1,0) errors   
##   
## Coefficients:  
## drift d2016 d2020 d2021 d2022  
## 258347.96 -860910.5 -7670244.0 -5962549.9 -1724047.9  
## s.e. 40329.49 142585.4 205659.5 296435.9 369737.2  
##   
## sigma^2 = 49498798008: log likelihood = -381.73  
## AIC=775.46 AICc=779.46 BIC=783.45  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 173.6976 198135.5 141811.7 -0.01871124 1.621725 0.1763318  
## ACF1  
## Training set -0.1496057

Box.test(error, lag = 2, type = "Ljung-Box")

##   
## Box-Ljung test  
##   
## data: error  
## X-squared = 2.8843, df = 2, p-value = 0.2364

Box.test(error^2, lag = 2, type = "Ljung-Box")

##   
## Box-Ljung test  
##   
## data: error^2  
## X-squared = 2.4509, df = 2, p-value = 0.2936

shapiro.test(error)

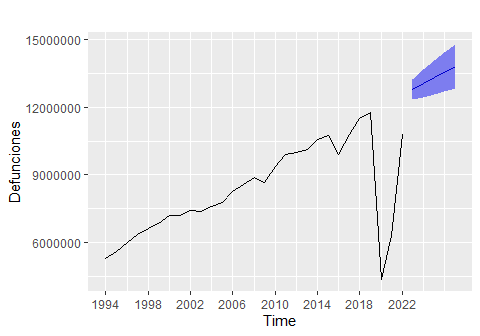
##   
## Shapiro-Wilk normality test  
##   
## data: error  
## W = 0.96834, p-value = 0.5155

Las hipótesis de normalidad, incorrelación y homocedasticidad se aceptan.

# 4 Predicción

parima010 <- forecast(arima010,   
 h = 5,   
 level = 95,  
 xreg= cbind(d2016=rep(0, 5), d2020=rep(0, 5), d2021=rep(0, 5),d2022=rep(0, 5)))  
  
autoplot(parima010,   
 ylab = "Defunciones",  
 main = "") +  
 scale\_x\_continuous(breaks= seq(1990, 2025, 4))

## Scale for x is already present.  
## Adding another scale for x, which will replace the existing scale.



k <- 10   
h <- 5   
T <- length(SerieBelgA)   
s <- T - k - h   
  
mapeArima <- matrix(NA, s + 1, h)  
  
X <- data.frame(cbind(d2016, d2020, d2021, d2022))  
  
for (i in 0:s) {  
 train.set <- subset(SerieBelgA, start = i + 1, end = i + k)  
 test.set <- subset(SerieBelgA, start = i + k + 1, end = i + k + h)   
   
 X.train <- data.frame(X[(i + 1):(i + k),])  
 X.test <- data.frame(X[(i + k + 1):(i + k + h),])  
   
 hay <- colSums(X.train)   
   
 if(sum(hay) == 0) {  
 X.train <- NULL  
 X.test <- NULL  
 } else {  
 X.train <- as.matrix(X.train[, hay>0])  
 X.test <- as.matrix(X.test[, hay>0])  
 }  
   
 fit <- Arima(train.set,   
 include.constant = FALSE,  
 order = c(0, 1, 0),  
 xreg = X.train)  
   
 fcast <- forecast(fit, h = h, xreg = X.test)  
   
 mapeArima[i + 1,] <- 100\*abs(test.set - fcast$mean)/test.set  
   
}  
  
mapeArima <- colMeans(mapeArima)  
mapeArima

## [1] 3.770551 5.974864 17.362115 23.403658 24.497099

# 5 Comparación de modelo de Alisado

Se quiere comparar la p

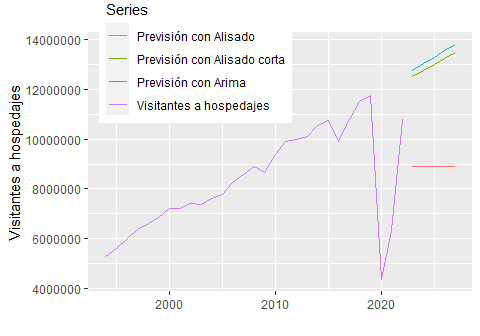
corto<-window(SerieBelgA,end=2019)  
summary(alisado <- ets(SerieBelgA))

## ETS(M,N,N)   
##   
## Call:  
## ets(y = SerieBelgA)   
##   
## Smoothing parameters:  
## alpha = 0.3413   
##   
## Initial states:  
## l = 6594449.256   
##   
## sigma: 0.172  
##   
## AIC AICc BIC   
## 921.0494 922.0094 925.1513   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 234855.6 1596700 1020753 -1.157505 14.85812 1.269227 0.1653713

summary(alisado2 <- ets(corto))

## ETS(M,A,N)   
##   
## Call:  
## ets(y = corto)   
##   
## Smoothing parameters:  
## alpha = 0.0361   
## beta = 0.0001   
##   
## Initial states:  
## l = 5251345.71   
## b = 244201.4577   
##   
## sigma: 0.0328  
##   
## AIC AICc BIC   
## 741.2139 744.2139 747.5044   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -50550.53 276093.5 200450.9 -0.6868952 2.380907 0.5797593  
## ACF1  
## Training set 0.2659279

palisado <- forecast(alisado, h = 5)   
palisado2 <- forecast(alisado2, h = 8) # 8 porque son los 3 años del covid y 5 adicionales  
  
autoplot(SerieBelgA,  
 series = "Visitantes a hospedajes",  
 main="",  
 xlab="",   
 ylab="Visitantes a hospedajes") +  
 autolayer(parima010$mean, series = "Previsión con Arima") +  
 autolayer(window(palisado$mean,start=2023), series = "Previsión con Alisado") +  
 autolayer(window(palisado2$mean,start=2023), series = "Previsión con Alisado corta") +  
 labs(colour = "Series") +   
 theme(legend.position=c(0.28,0.9))



## 5.1 Comparación de la calidad de los modelos.

| Modelo | AICc | RMSE | MPE | MAPE | MASE | ACF1 |
| --- | --- | --- | --- | --- | --- | --- |
| Modelo Alisado | 922.00 | 1596700 | -1.16 | 14.85 | 1.26 | 0.16 |
| Modelo Alisado Corto | 744.21 | 276093 | -0.69 | 2.38 | 0.58 | 0.27 |
| ARIMA (0,1,0) | 779.46 | 198135 | -0.02 | 1.62 | 0.18 | -0.15 |

## 5.2 Comparación de las previsiones