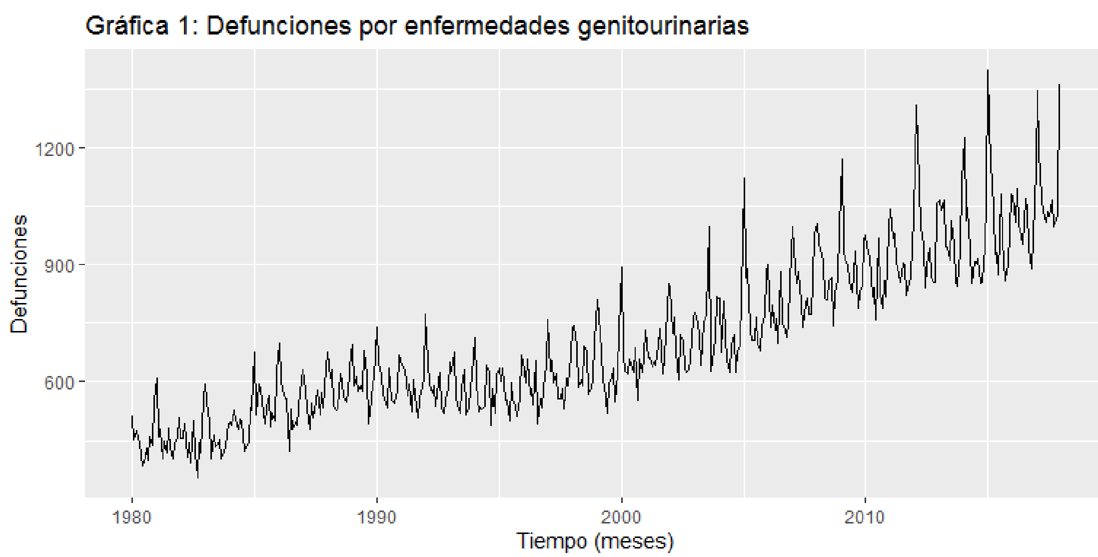
Práctica 5: Series Temporales

Irene Extremera Serrano

29/03/2020

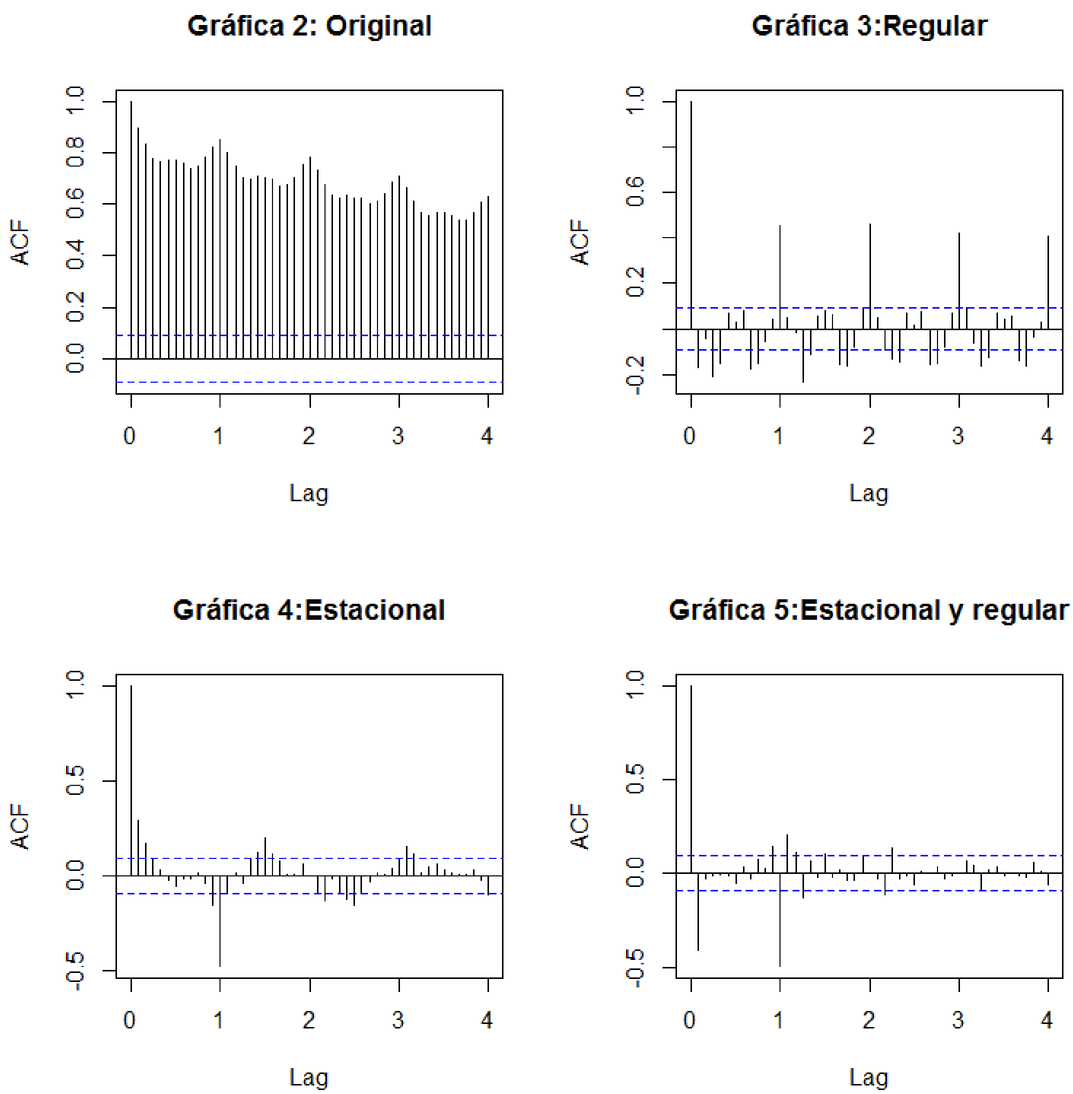
La serie temporal con la que voy a trabajar recopila información sobre fallecimientos causados por enfermedades que afectan al sistema genitourinario, procede del INE y va desde enero de 1980 a diciembre de 2017. El objetivo de esta práctica es analizar la serie temporal con estacionalidad por lo que haré uso de la serie mensual y seguiré la metodología de Box y Jenkins para identificar qué modelo ARIMA genera unas mejores predicciones.

```
# Librerías a usar
library(forecast)
library(tseries)
library(aod)
setwd('D:/Desktop/Remember/Estudios/Educación Formal/Máster/Máster Valenc
ia/Bioestadística/Curso 1/20 2-6Modelización Estadística/Series Temporale
s/Practicas/P5')
# Serie Anual
enf_gu <- read.table('Enfermedades_del_sistema_genitourinario.txt', heade
r = TRUE)
enf_mes <- ts(enf_gu,start = c(1980,1), freq = 12)

#Gráfica de las series
autoplot(enf_mes,main='Gráfica 1: Defunciones por enfermedades genitourin
arias',ylab='Defunciones',xlab='Tiempo (meses)')
```



En la gráfica uno se muestra la serie temporal con estacionalidad y con tendencia. Como para realizar el siguiente análisis necesito que la serie sea estacionaria, ergódica y lo más homocedástica posible realizo varias funciones de autocorrelación para comprobar cuántas veces hay que diferenciar estacional y regularmente.

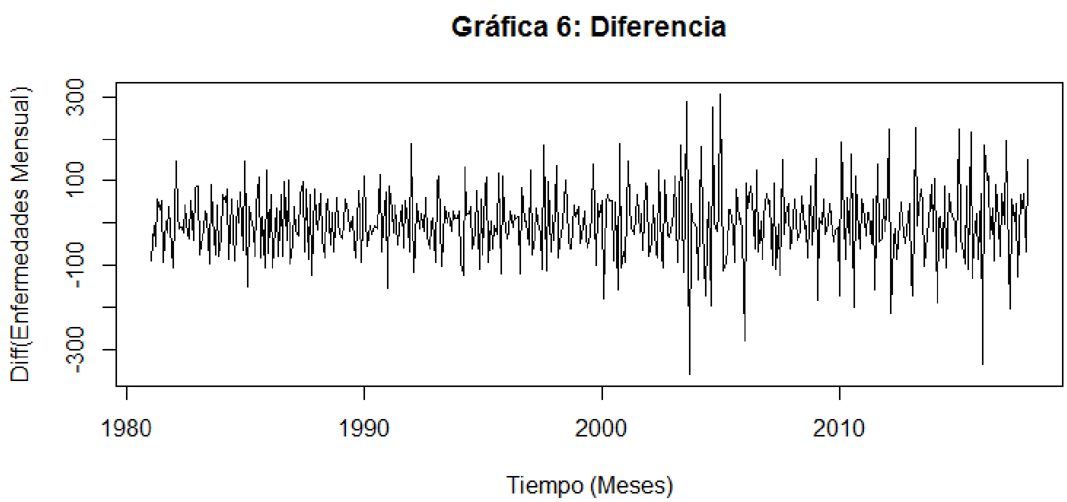
```r
# Funciones de autocorrelación
par(mfrow=c(1,2))
acf(enf_mes, main = "Gráfica 2: Original", lag = 48)
acf(diff(enf_mes), main = "Gráfica 3:Regular", lag =48)
acf(diff(enf_mes,lag=12), main = "Gráfica 4:Estacional", lag = 48)
acf(diff(diff(enf_mes,lag=12)), main = "Gráfica 5:Estacional y regular",
lag = 48)
```



La diferenciación estacional-regular consigue una reducción muy rápida (estacionaria) y además la mayor parte de los valores quedan dentro del intervalo de confianza (ergódica), por lo que me quedo con esta. También podría considerar quedarme con la diferenciación estacional, pero tarda un poco más en decrecer en los primeros valores y prefiero por ello quedarme con la diferenciación estacional-regular.

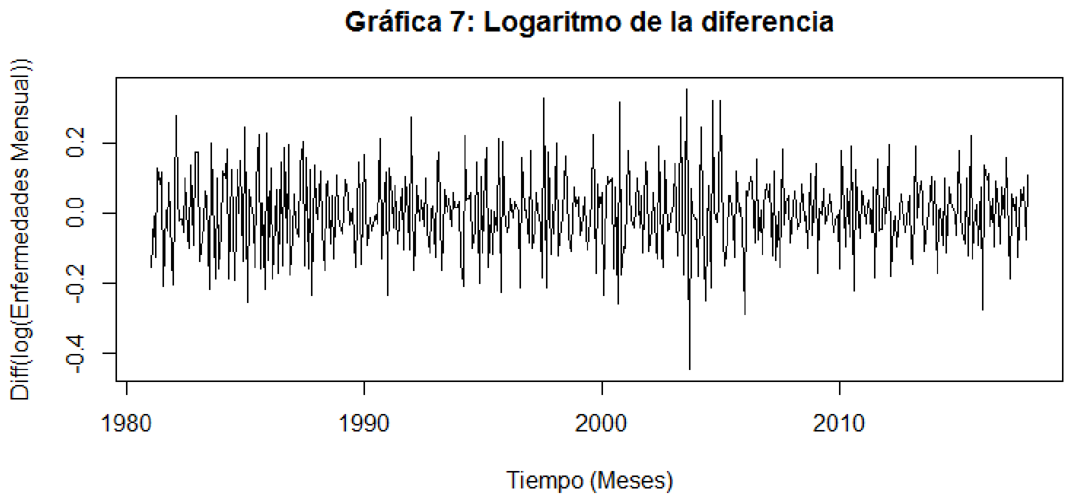
Una vez que me he cerciorado de que la serie es estacionaria y ergódica realizando esa diferenciación dibujo un gráfico para ver cómo es la serie.

```
par(mfrow=c(1,1))
plot(diff(diff(enf_mes,lag=12)),main='Gráfica 6: Diferencia', ylab='Diff(
Enfermedades Mensual)', xlab='Tiempo (Meses)')
```



A la vista de la gráfica 6, he conseguido la estacionaridad en media con esta doble diferenciación (d=D=1). Sin embargo, veo que no hay una estacionaridad en varianza ya que los valores van aumentando en dispersión a medida que el tiempo avanza. Por lo que realizaré el logaritmo de la serie para conseguir la estacionaridad en varianza también.

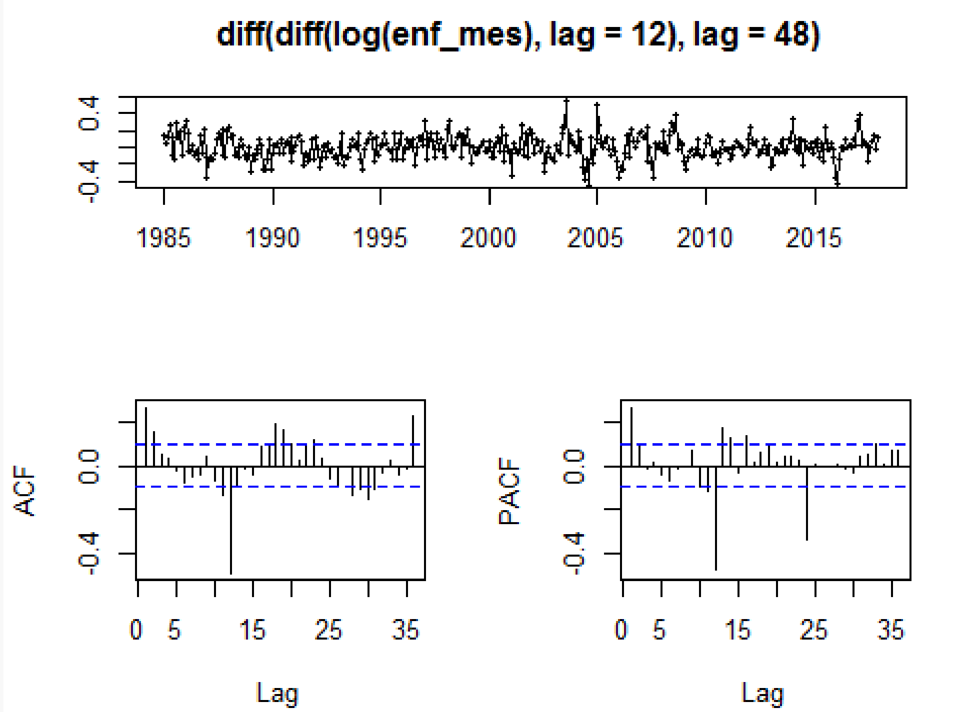
```r
plot(diff(diff(log(enf_mes),lag=12)),main='Gráfica 7: Logaritmo de la dif
erencia', ylab='Diff(log(Enfermedades Mensual))', xlab='Tiempo (Meses)')
```



Como muestra la gráfica 6, efectivamente realizando el logaritmo de la serie consigo la estacionaridad en varianza para la serie temporal mensual.

Una vez conseguido que la serie sea estacionaria, ergódica y lo más homocedástica posible puedo comenzar con su identificación.

```r
tsdisplay(diff(diff(log(enf_mes),lag=12),lag=48))
```



Lo que puedo observar en las distintas gráficas es que en la función de autocorrelación, en la parte regular hay un decrecimiento y un pico en ro12 en la parte estacional, mientras que en la función de autocorrelación parcial observo un decrecimiento en la parte estacional y decrecimiento en la parte regular. Aunque no se ve muy claramente deduzco que tengo un ARIMA(1,1,0)(0,1,1) o un ARIMA(0,1,1) (0,1,1).

Sin embargo, el haber realizado el análisis completo de ARIMA(1,1,0)(0,1,1), el cual no pasó las validaciones, y del modelo propuesto por autoarima (2,1,2)(2,1,0), que pasaba las validaciones pero era muy complejo, probé con un modelo ligeramente más sencillo, un ARIMA(0,1,1)(0,1,1) obteniendo unas predicciones mejores.

La ecuación del modelo que voy a desarrollar es la siguiente:

ARIMA(0,1,1)(0,1,1)

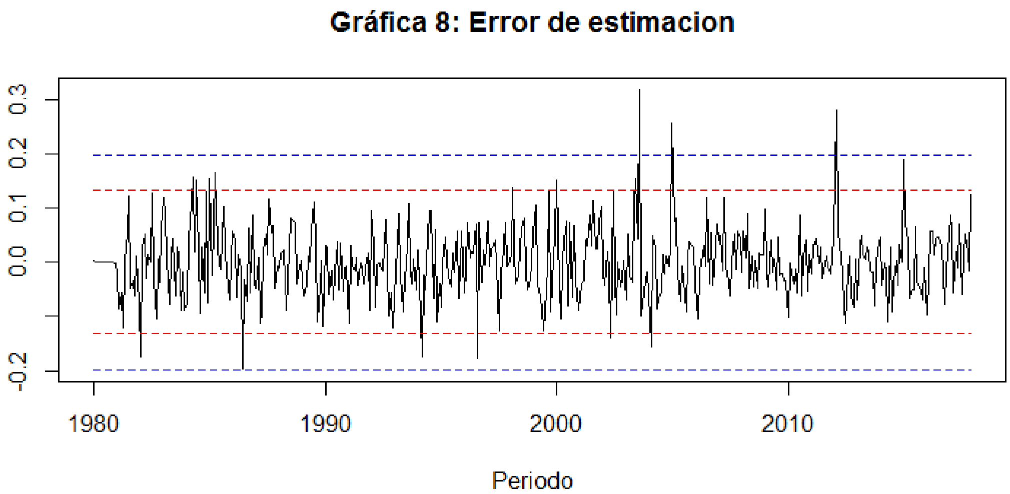
$$(1-L^{12})(1-L)log y_t = c + (1+\theta_1 L)(1+\theta_{12} L^{12})\varepsilon_t$$

A continuación incluiré el desarrollo de la estimación y predicción del modelo ARIMA(0,1,1)(0,1,1) que es con el que me he quedado.

```
#Realizo el ARIMA(0,1,1)(0,1,1) con lambda = 0 porque uso el logaritmo en la serie
arima0 <-  Arima(enf_mes,order=c(0,1,1),
                 seasonal=list(order=c(0,1,1),period=12),
                 lambda=0)
arima0

## Series: enf_mes
## ARIMA(0,1,1)(0,1,1)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ma1     sma1
##       -0.8223  -0.9993
## s.e.   0.0377   0.1143
##
## sigma^2 estimated as 0.004352:  log likelihood=554.46
## AIC=-1102.93   AICc=-1102.87   BIC=-1090.64
```

Ahora voy a mirar si hay algún valor atípico que merezca la pena incluir en el modelo.

```
#Gráfica del residuo
es <- sqrt(arima0$sig)
ts.plot(arima0$residuals, 2*es, -2*es, 3*es, -3*es,
        xlab = "Periodo",
        plot.type = "response",
        ylab = "",
        main = "Gráfica 8: Error de estimacion",
        lty = c(1, 2, 2, 2, 2),
        col = c("black", "red", "red", "blue", "blue"))
```




```r
# Intervenciones
arima0$residuals > 3*es
arima0$residuals < -3*es
```

Se puede observar en la gráfica 8 que hay cinco valores que superan las tres desviaciones típicas por encima y por debajo por lo que sería necesario incluir esos valores en el modelo. Estos valores atípicos corresponden a: agosto de 2003, enero de 2005 y febrero de 2012.

Incluyo las intervenciones en su respectivo modelo y veo si significativas.

```r
#arima0rima
d2003 <- rep(0, length(arima0$residuals)) #agosto
d2005 <- rep(0, length(arima0$residuals)) #enero
d2012 <- rep(0, length(arima0$residuals)) #febrero
ii <- order(abs(arima0$residuals), decreasing = TRUE)
d2003[ii[3]] <- 1
d2005[ii[2]] <- 1
d2012[ii[1]] <- 1

arima01 <- Arima(enf_mes,order= c(0,1,1),seasonal=list(order=c(0,1,1),per
iod=12),lambda=0,xreg=cbind(d2003,d2005,d2012))
arima01

## Series: enf_mes
## Regression with ARIMA(0,1,1)(0,1,1)[12] errors
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ma1     sma1   d2003   d2005   d2012
##       -0.8270  -1.0000  0.2247  0.2634  0.3457
## s.e.   0.0323   0.0622  0.0593  0.0594  0.0594
##
```

```
## sigma^2 estimated as 0.003795:  log likelihood=586.16
## AIC=-1160.32   AICc=-1160.12   BIC=-1135.75
```

```r
wald.test(b = coef(arima01), Sigma = vcov(arima01), Terms = 1)
```

```
## Wald test:
## ----------
## 
## Chi-squared test:
## X2 = 653.7, df = 1, P(> X2) = 0.0
```

```r
wald.test(b = coef(arima01), Sigma = vcov(arima01), Terms = 2)
```

```
## Wald test:
## ----------
## 
## Chi-squared test:
## X2 = 258.2, df = 1, P(> X2) = 0.0
```

```r
wald.test(b = coef(arima01), Sigma = vcov(arima01), Terms = 3)
```

```
## Wald test:
## ----------
## 
## Chi-squared test:
## X2 = 14.3, df = 1, P(> X2) = 0.00015
```

```r
wald.test(b = coef(arima01), Sigma = vcov(arima01), Terms = 4)
```

```
## Wald test:
## ----------
## 
## Chi-squared test:
## X2 = 19.6, df = 1, P(> X2) = 9.3e-06
```

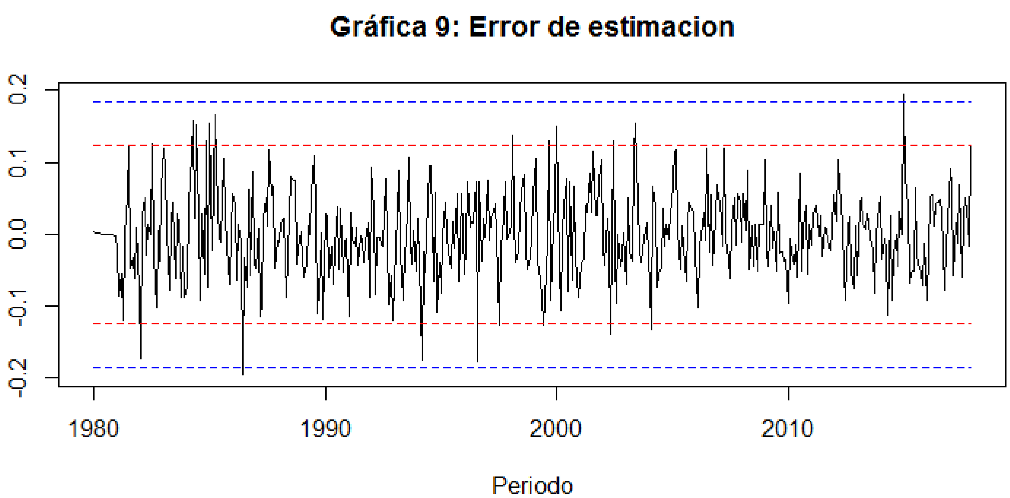
```r
wald.test(b = coef(arima01), Sigma = vcov(arima01), Terms = 5)
```

```
## Wald test:
## ----------
## 
## Chi-squared test:
## X2 = 33.9, df = 1, P(> X2) = 5.9e-09
```

Todos los coeficientes asociados a los valores atípicos son significativos y los coeficientes ma1 y sma1 son significativos, por lo que todo ha de incluirse en el modelo ARIMA(0,1,1) (0,1,1) mejorado.

Una vez mejorado el modelo vuelvo a comprobar si tiene valores atípicos.

```r
#Gráfica del residuo
arima02 <- Arima(enf_mes,order=c(0,1,1),
                 seasonal=list(order=c(0,1,1),period=12),
                 lambda=0,
                 xreg=cbind(d2003,d2005,d2012))
es <- sqrt(arima02$sig)
ts.plot(arima02$residuals, 2*es, -2*es, 3*es, -3*es,
        xlab = "Periodo",
        plot.type = "response",
        ylab = "",
        main = "Gráfica 9: Error de estimacion",
        lty = c(1, 2, 2, 2, 2),
        col = c("black", "red", "red", "blue", "blue"))
```




```r
# Intervenciones
arima02$residuals > 3*es
arima02$residuals < -3*es
```

En la gráfica nueve se observa que en el modelo mejorado aun hay dos valores atípicos, uno corresponde con junio de 1986 y otro con febrero de 2012. El valor atípico de 2012 ya está incluido por lo que el valor que voy a meter en el modelo va a ser el de junio de 1986.

```r
#arima02rima
d1986 <- rep(0, length(arima0$residuals)) #junio

ii <- order(abs(arima0$residuals), decreasing = TRUE)
d1986[ii[456]] <- 1

arima03 <- Arima(enf_mes,order= c(0,1,1),seasonal=list(order=c(0,1,1),period=12),lambda=0,xreg=cbind(d2003,d2005,d2012,d1986))
```

```r
arima03

## Series: enf_mes
## Regression with ARIMA(0,1,1)(0,1,1)[12] errors
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ma1     sma1   d2003   d2005   d2012    d1986
##       -0.8278   -1.000  0.2248  0.2636  0.3459  -0.0504
## s.e.   0.0323    0.058  0.0593  0.0594  0.0594   0.0593
##
## sigma^2 estimated as 0.003797:  log likelihood=586.52
## AIC=-1159.04   AICc=-1158.78   BIC=-1130.38

wald.test(b = coef(arima03), Sigma = vcov(arima03), Terms = 1)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 656.5, df = 1, P(> X2) = 0.0

wald.test(b = coef(arima03), Sigma = vcov(arima03), Terms = 2)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 297.6, df = 1, P(> X2) = 0.0

wald.test(b = coef(arima03), Sigma = vcov(arima03), Terms = 3)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 14.4, df = 1, P(> X2) = 0.00015

wald.test(b = coef(arima03), Sigma = vcov(arima03), Terms = 4)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 19.7, df = 1, P(> X2) = 9.1e-06
```

```r
wald.test(b = coef(arima03), Sigma = vcov(arima03), Terms = 5)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 33.9, df = 1, P(> X2) = 5.7e-09
```

```r
wald.test(b = coef(arima03), Sigma = vcov(arima03), Terms = 6)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 0.72, df = 1, P(> X2) = 0.4
```

Se observa que todos los coeficientes salen significativos a excepción de junio de 1986 con un p valor de 0.4 por lo que no lo incluiré en el modelo definitivo.

La ecuación definitiva por lo tanto será la siguiente:

$$\text{ARIMA}(0,1,1)(0,1,1)$$

$$(1-L^{12})(1-L)log\mathrm{y_t} = (1+\theta_1 L)(1+\theta_{12}L^{12})\varepsilon_\mathrm{t} + AI_1\mathrm{d2003} + AI_2\mathrm{d2005} + AI_3\mathrm{d2012}$$

$$TVAnual\mathrm{y_t} = TVAnual\mathrm{y_{t_{t\text{-}1}}} + \varepsilon_\mathrm{t} + \theta_1\varepsilon_\mathrm{t\text{-}1} + \theta_{12}\varepsilon_\mathrm{t\text{-}12} + \theta_1\theta_{12}\varepsilon_\mathrm{t\text{-}13} + AI_1\mathrm{d2003} + AI_2\mathrm{d2005} + AI_3\mathrm{d2012}$$

$$TVAnualy_t = TVAnual\mathrm{y_{t_{t\text{-}1}}} + \varepsilon_\mathrm{t} + (-0.83)\varepsilon_\mathrm{t\text{-}1} + (-1)\varepsilon_\mathrm{t\text{-}12} + (-0.83)(-1)\varepsilon_\mathrm{t\text{-}13} + 0.22\mathrm{d2003} + 0.26\mathrm{d2005} + 0.34\mathrm{d2012}$$

$$\widehat{TV}Anual\mathrm{y_t} = TVAnual\mathrm{y_{t_{t\text{-}1}}} + (-0.83)\varepsilon_\mathrm{t\text{-}1} + (-1)\varepsilon_\mathrm{t\text{-}12} + (-0.83)(-1)\varepsilon_\mathrm{t\text{-}13} + 0.22\mathrm{d2003} + 0.26\mathrm{d2005} + 0.34\mathrm{d2012}$$

La tasa de variación anual del número de fallecimientos a causa de enfermedades que afectan al sistema genitourinario se mantiene constante con respecto al periodo anterior ($TVAnual = TVAnual_\mathrm{t\text{-}1}$)

Si la tasa de variación anual del mes previo fue anómala ($\varepsilon_\mathrm{t\text{-}1}$diferente de cero) o la tasa de variación anual de hace un año fue anómala ($\varepsilon_\mathrm{t\text{-}12}$diferente de cero), corrijo para compensar (términos en medias móviles).

Aparte, por alguna causa que desconozco en 2003, 2005 y 2012 hubo respectivamente un 22%, 26% y 34% más de fallecimientos de lo que se esperaba para ese año.

```r
#Calidad del modelo
accuracy(arima02)
```

```
##                    ME     RMSE      MAE        MPE     MAPE      MASE      ACF1
## Training set 1.645685 43.20709 32.55196 -0.1647271 4.661491 0.6041749 0.1235425
```

No se puede contrastar i el residuo tiene media cero pero el error medio es 1.646.

```
#Homocedasticidad
Box.test(arima02$residuals^2,lag=2, type='Ljung-Box')
##  Box-Ljung test
##
## data:  arima02$residuals^2
## X-squared = 0.31781, df = 2, p-value = 0.8531

Box.test(arima02$residuals^2,lag=24, type='Ljung-Box')
##  Box-Ljung test
##
## data:  arima02$residuals^2
## X-squared = 35.409, df = 24, p-value = 0.06257
```

Se acepta que el residuo es homocedástico (para lag 2 p valor de 0.85 y para lag 24, 0.0625).

```
#Incorrelación
Box.test(arima02$residuals,lag=2, type='Ljung-Box')
##  Box-Ljung test
##
## data:  arima02$residuals
## X-squared = 5.4255, df = 2, p-value = 0.06635

Box.test(arima02$residuals,lag=24, type='Ljung-Box')
##  Box-Ljung test
##
## data:  arima02$residuals
## X-squared = 31.724, df = 24, p-value = 0.134
```

El residuo es incorrelado (para lag 2 p valor de 0.066 y para lag 24, 0.134)

```
#Normalidad
jarque.bera.test(arima02$residuals)

##
##  Jarque Bera Test
##
## data:  arima02$residuals
## X-squared = 3.4852, df = 2, p-value = 0.1751
```

Para un p valor de 0.175 se acepta la hipótesis de normalidad.

Una vez que el modelo ha sido validado puedo ver cómo de bueno es para realizar predicciones.

```
accuracy(arima02)

##                    ME     RMSE      MAE        MPE     MAPE      MASE      ACF1
## Training set 1.645685 43.20709 32.55196 -0.1647271 4.661491 0.6041749 0.1235425
```

Para empezar el MAPE es de 4.66% lo cual quiere decir que el modelo es bueno a la hora de realizar predicciones. Además, el ACIF1 es bastante bajo, de 0.12, lo cual indica que la capacidad de mejora de las predicciones por intervalo es bastante baja. Aparte, en media el modelo se equivoca 43.20 casos (RMSE).

A continuación realizaré una predicción del número de fallecimientos a tres años vista y su intervalo al 80 y 95.

```
#Predicción
parima02 <- forecast(arima02, h=36,level =c(80,95), xreg = cbind(rep(0,36
),rep(0,36),rep(0,36)))

parima02

##          Point Forecast      Lo 80    Hi 80     Lo 95    Hi 95
## Jan 2018       1328.561 1226.4293 1439.198 1175.5816 1501.448
## Feb 2018       1182.515 1090.3145 1282.512 1044.4533 1338.826
## Mar 2018       1190.882 1096.7449 1293.099 1049.9625 1350.715
## Apr 2018       1083.325  996.5398 1177.669  953.4491 1230.893
## May 2018       1082.096  994.2770 1177.672  950.7111 1231.639
## Jun 2018       1028.974  944.4045 1121.116  902.4871 1173.188
## Jul 2018       1142.005 1046.9839 1245.649  999.9266 1304.270
## Aug 2018       1103.178 1010.2812 1204.616  964.3146 1262.038
## Sep 2018       1003.977  918.4406 1097.480  876.1510 1150.453
## Oct 2018       1050.016  959.5314 1149.033  914.8322 1205.176
## Nov 2018       1090.396  995.3799 1194.481  948.4806 1253.544
## Dec 2018       1260.252 1149.2353 1381.993 1094.4819 1451.130
## Jan 2019       1361.541 1239.9138 1495.098 1179.9904 1571.024
## Feb 2019       1211.869 1102.4330 1332.168 1048.5603 1400.612
## Mar 2019       1220.444 1109.0606 1343.013 1054.2740 1412.804
## Apr 2019       1110.217 1007.8397 1222.994  957.5231 1287.261
## May 2019       1108.958 1005.6557 1222.871  954.9252 1287.836
## Jun 2019       1054.517  955.3077 1164.028  906.6256 1226.532
## Jul 2019       1170.353 1059.1727 1293.204 1004.6589 1363.374
## Aug 2019       1130.562 1022.1366 1250.490  969.0144 1319.043
## Sep 2019       1028.899  929.3002 1139.173  880.5400 1202.255
## Oct 2019       1076.081  970.9593 1192.584  919.5344 1259.279
## Nov 2019       1117.463 1007.3170 1239.653  953.4748 1309.655
## Dec 2019       1291.536 1163.1090 1434.143 1100.3772 1515.902
## Jan 2020       1395.339 1255.0090 1551.359 1186.5295 1640.895
## Feb 2020       1241.951 1115.9417 1382.190 1054.4973 1462.728
## Mar 2020       1250.739 1122.7351 1393.337 1060.3654 1475.292
## Apr 2020       1137.776 1020.3402 1268.729  963.1626 1344.046
## May 2020       1136.486 1018.2005 1268.512  960.6523 1344.503
## Jun 2020       1080.693  967.2899 1207.392  912.1574 1280.369
## Jul 2020       1199.405 1072.5277 1341.292 1010.8898 1423.076
## Aug 2020       1158.627 1035.0899 1296.908  975.1183 1376.670
## Sep 2020       1054.440  941.1345 1181.387  886.1692 1254.663
## Oct 2020       1102.793  983.3820 1236.704  925.4964 1314.054
## Nov 2020       1145.202 1020.2632 1285.441  959.7406 1366.502
## Dec 2020       1323.596 1178.1225 1487.032 1107.7020 1581.568
```

```r
#Diferencia de un año a otro por meses
parima02$mean[1:12]-parima02$mean[13:24]
```

```
##  [1] -32.97937 -29.35400 -29.56170 -26.89178 -26.86128 -25.54260 -28.34840
##  [8] -27.38459 -24.92210 -26.06493 -27.06729 -31.28370
```

```r
parima02$mean[13:24]-parima02$mean[25:36]
```

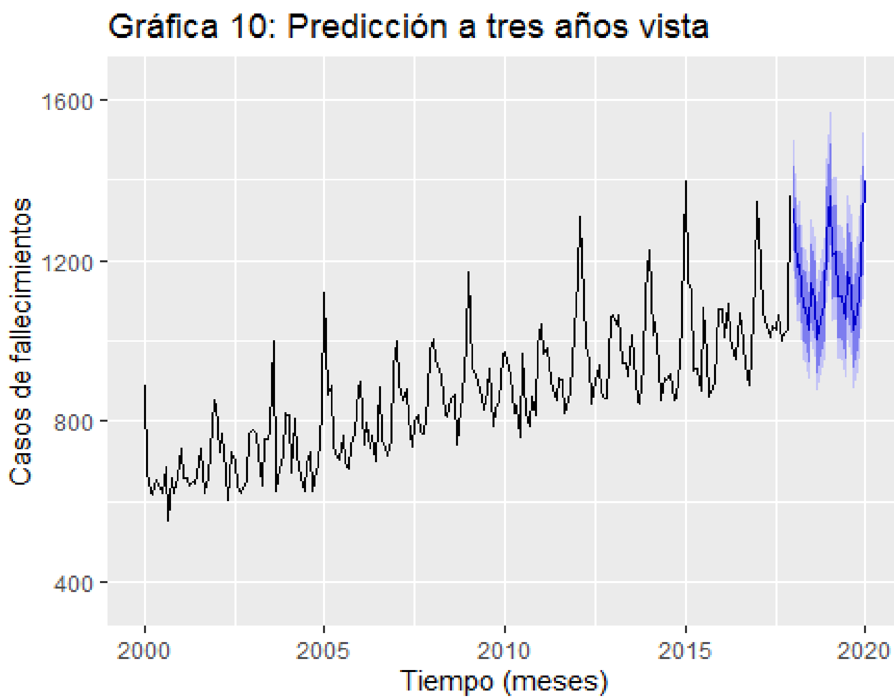
```
##  [1] -33.79802 -30.08266 -30.29552 -27.55933 -27.52807 -26.17665 -29.05210
##  [8] -28.06437 -25.54075 -26.71195 -27.73919 -32.06027
```

```r
# Meses en los que el número de fallecimientos es mayor y menor por año
which.max(parima02$mean[1:12])
## [1] 1
which.max(parima02$mean[13:24])
## [1] 1
which.max(parima02$mean[25:36])
## [1] 1
which.min(parima02$mean[1:12])
## [1] 9
which.min(parima02$mean[13:24])
## [1] 9
which.min(parima02$mean[25:36])
## [1] 9
```

En los distintos años predichos se observa que tanto el número de fallecimientos como su intervalo es mayor en los meses más fríos en especial en enero y más bajos en meses con temperaturas más suaves como septiembre. Otra cosa que queda por mencionar es que hay un incremento progresivo en el número de fallecimientos de un año a otro, por lo que la tendencia es positiva.

Para verlo de forma gráfica:

```r
autoplot(parima02, ylab='Casos de fallecimientos',xlab='Tiempo (meses)',
xlim=c(2000,2020),main='Gráfica 10: Predicción a tres años vista')
```



En la gráfica 10 se puede observar lo mencionado anteriormente, el número de fallecimientos aumenta de un año para otro lo cual hace que la tendencia sea positiva y en los tres años coinciden los meses con más número de muertes (enero) y con menos (septiembre).

Para finalizar la práctica haré una comparación con el modelo de alisado obtenido en la práctica 2.

```
enf_ets <- ets(enf_mes)
summary(enf_ets)

## ETS(M,A,M)
##
## Call:
##  ets(y = enf_mes)
##
##   Smoothing parameters:
##     alpha = 0.2026
##     beta  = 1e-04
##     gamma = 1e-04
##
##   Initial states:
##     l = 439.5583
##     b = 1.3845
##     s = 1.0986 0.9523 0.9181 0.8804 0.9851 1.0087
##            0.9091 0.9574 0.9637 1.0594 1.0631 1.204
##
```

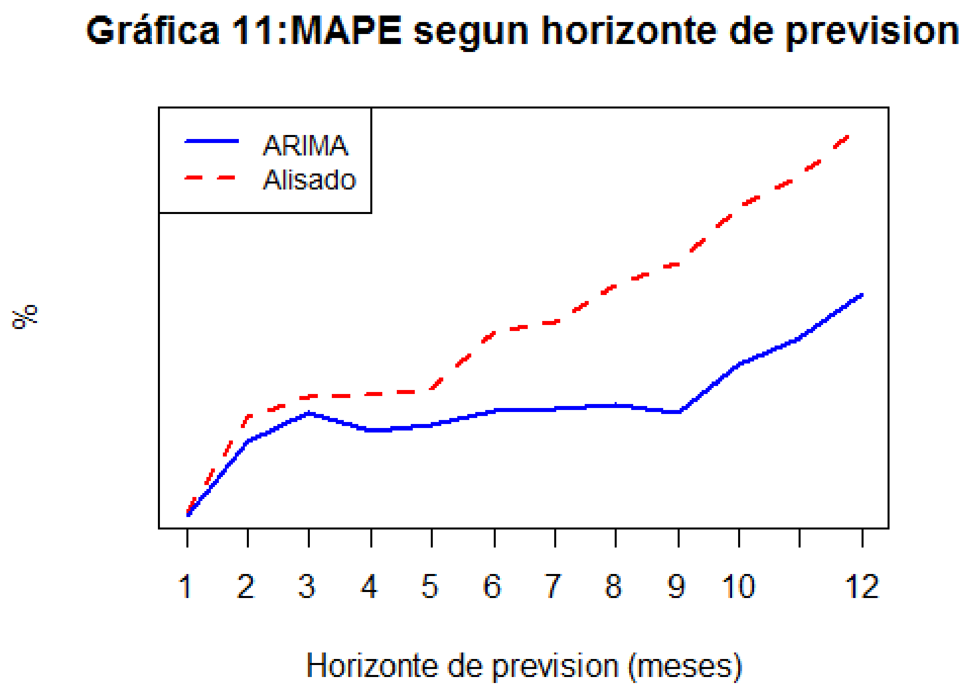
```
##   sigma:  0.0672
##
##      AIC     AICc      BIC
## 6293.591 6294.988 6363.673
##
## Training set error measures:
##                     ME     RMSE     MAE        MPE     MAPE      MASE      ACF1
## Training set 0.6504983 48.41639 34.6059 -0.3850739 4.945122 0.6422966 0.1147674
```

El valor de MAPE (error porcentual absoluto medio) es de 4.945% lo cual indica que la calidad de predicción del modelo es muy alta, sin embargo, la del modelo ARIMA es ligeramente mejor 4.66%. Aparte, el valor de MASE es del 0.642% por lo que el modelo de alisado no es mucho mejor con respecto al modelo más sencillo que podría aplicarse. Para terminar el ACF1 toma un valor de 0.115% ligeramente inferior al anterior, 0.12 y el RMSE en alisado es de 48.41 y de 43.20 en ARIMA.

Con el objetivo de analizar finamente la calidad del ajuste, voy a estimar el error según el horizonte temporal mediante la aplicación del procedimiento de origen de predicción móvil. Por ello voy a usar cinco años para la estimación del error y predecir en un horizonte temporal de 12 meses.

```
k <- 60                   #Mínimo número de datos para estimar
h <- 12                    #Horizonte de las prediciciones
T <- length(enf_mes)     #Longitud serie
s<-T - k - h               #Total de estimaciones

mapeArima <- matrix(NA, s, h)
mapeAlisado <- matrix(NA, s, h)

X <- cbind(d2003,d2005,d2012)  #Valores atípicos

for (i in 0:s) {
  train.set <- subset(enf_mes, start = i + 1, end = i + k)
  test.set <-  subset(enf_mes, start = i + k + 1, end = i + k + h)

  X.train <- X[(i + 1):(i + k),]
  hay <- colSums(X.train)
  X.train <- X.train[, hay>0]

  X.test <- X[(i + k + 1):(i + k + h),]
  X.test <- X.test[, hay>0]

  if (length(X.train) > 0) {
    fit <- try(Arima(train.set,
                 order = c(0, 1, 1),
                 seasonal = list(order = c(0, 1, 1), period = 12),
                 lambda = 0,   #
                 xreg=X.train), silent = TRUE)} else {
                   fit <- try(Arima(train.set,
                                order = c(0, 1, 1),
```

```r
                                seasonal=list(order=c(0,1,1),period= 12),
                                lambda = 0),
                             silent = TRUE)
                 }

  if (!is.element("try-error", class(fit))) {
    if (length(X.train) > 0) fcast <- forecast(fit, h = h, xreg = X.test)
else
      fcast <- forecast(fit, h = h)
    mapeArima[i,] <- 100*abs(test.set - fcast$mean)/test.set}

  fit <- ets(train.set, model = "MAM", damped = FALSE)
  fcast<-forecast(fit, h = h)
  mapeAlisado[i,] <- 100*abs(test.set - fcast$mean)/test.set}

errorArima <- colMeans(mapeArima, na.rm = TRUE)
errorArima

##  [1] 5.512621 5.806408 5.912495 5.842077 5.869323 5.923227 5.925331 5.942259
##  [9] 5.912545 6.099003 6.204672 6.373606

errorAlisado <- colMeans(mapeAlisado)
errorAlisado

##  [1] 5.532467 5.899591 5.975203 5.983157 6.006476 6.223820 6.262532 6.403772
##  [9] 6.492914 6.703717 6.825096 7.027012

plot(errorAlisado,
     type = 'l',col = 'red', lwd = 2, lty = 2,
     xlab = 'Horizonte de prevision (meses)', ylab = '%',
     main = 'Gráfica 11:MAPE segun horizonte de prevision',
     xaxp = c(1,12,11), yaxp = c(2,4,4))
lines(errorArima,
     col = "blue", lwd = 2, lty = 1)
     legend("topleft", legend = c("ARIMA","Alisado"),
     col = c("Blue","Red"), lwd = 2, lty = c(1, 2), cex = .9)
```



En la gráfica 11 se refleja cómo en ambos métodos el MAPE aumenta de forma progresiva tanto en alisado como en ARIMA. En predicciones a un mes o dos vista el modelo de alisado (5.53% y 5.89%) obtiene predicciones muy similares a las que hace ARIMA (5.51% y 5.80%). Sin embargo, a largo plazo parece que el error del modelo ARIMA es menor que el error que comete alisado, 6.37% frente a 7.03%. Se puede observar como en el modelo ARIMA el error se mantiene mas o menos constante prediciendo de 3 a nueve meses vista y a partir de ahí se incrementa progresivamente.

Para concluir, en caso de tener que decantarme por algún modelo me decantaría sin duda por el de alisado debido a que es muchísimo más sencillo de obtener y las predicciones que da son similares a las del ARIMA a corto plazo. Sin embargo, en caso de querer hacer predicciones a largo plazo, elegiría el ARIMA sin dudarlo aunque sea más costoso de obtener.