

Práctica 1 Series Temporales Bioestadística

Irene Extremera Serrano

25/2/2020

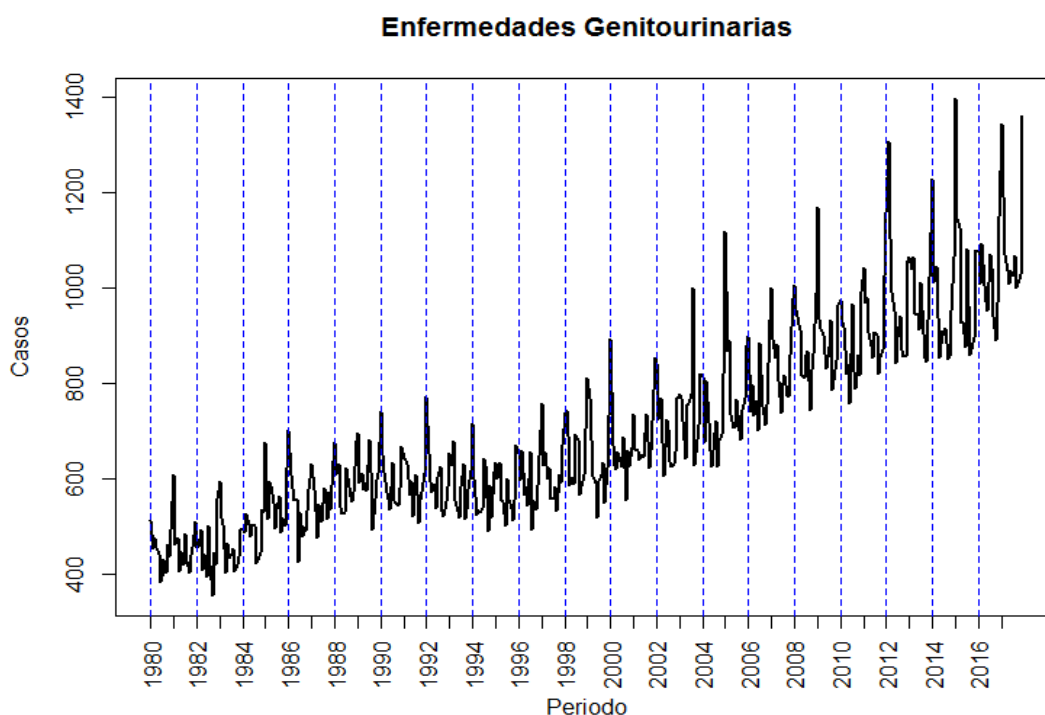
Práctica 1

Ejercicio 1

La serie temporal que voy a analizar contiene información del número de fallecimientos a casusa de enfermedades del sistema genitourinario por meses de enero de 1980 a diciembre de 2017 en España. La serie porcede del INE (Instituto Nacional de Estadística) y tiene un total 456 datos de fallecimientos (uno por mes).

Apartado a: Gráfica de la Serie Temporal.

```
Enf_GU <- read.csv("D:/Desktop/Remember/Estudios/Educación  
Formal/Máster/Máster Valencia/Bioestadística/Curso 1/20 2-6Modelización  
Estadística/Series  
Temporales/Temas/T1/Enfermedades_del_sistema_genitourinario.txt")  
  
Enf_GU <- ts(Enf_GU, start = c(1980,1), freq = 12)  
plot(Enf_GU,xlab = "Periodo", ylab = "Casos", type='l',main =  
"Enfermedades Genitourinarias", lwd = 2, xaxt = "n")  
axis(1, at = 1980:2017, las = 2)  
abline(v = seq(1980,2017,2), lwd = .7, lty = 2, col = "BLUE")
```



```
vector <- c()
for(i in 1:12) vector[i] <- print(Enf_GU[(12*37)+i])
max(vector)-min(vector)

## [1] 360
```

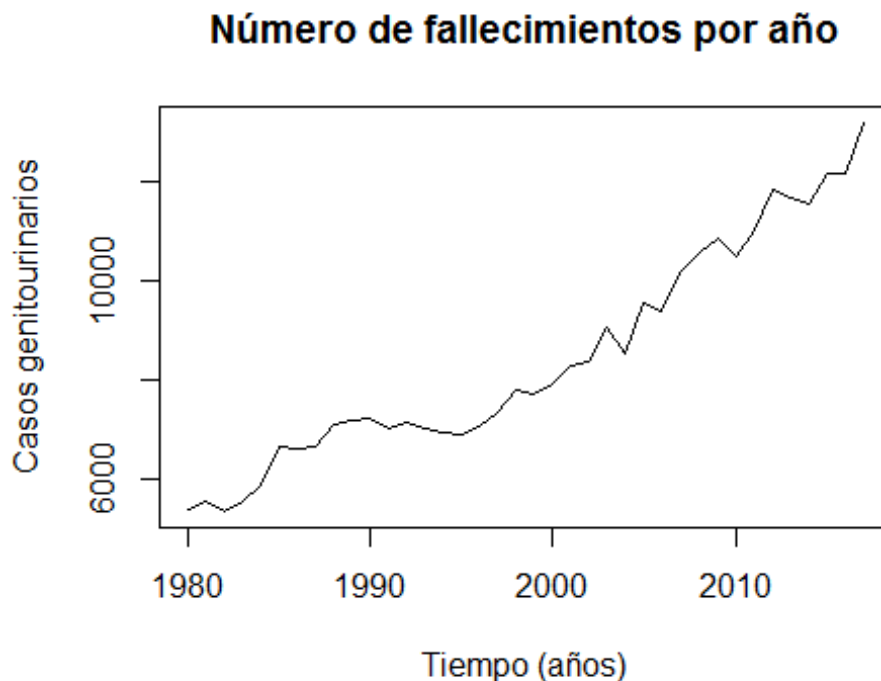
Se puede observar que el número casos ha ido aumentando con el paso del tiempo, por lo que la serie experimenta una tendencia creciente. Sin embargo, parece que aproximadamente entre los años 1989 y 1998 la tendencia se ha mantenido estacionaria.

Por otro lado, las líneas transversales azules discontinuas señalan el mes de enero cada dos años, esto me permite apreciar que hay una ligera componente estacional la cual se va acentuando a medida que lo hace la tendencia ascendente. Además, en el último año la diferencia de incidencia entre el mes con más muertes y el de menos muertes es de 360 casos.

Todo esto me lleva a pensar que el tipo de esquema que sigue la serie temporal es multiplicativo.

Apartado b: Serie Anual.

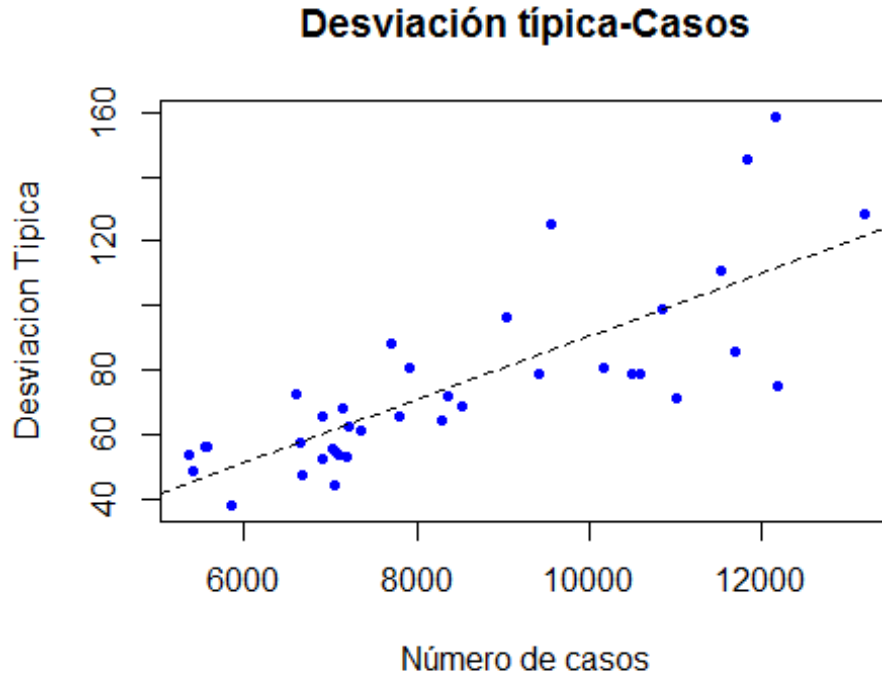
```
enf_gu_sum <- aggregate(Enf_GU,FUN=sum)
plot(enf_gu_sum, main='Número de fallecimientos por año', xlab='Tiempo
(años)',ylab='Casos genitourinarios')
```



En este primer gráfico se muestra el número total de fallecimientos por enfermedades genitourinarias (eje de las y) frente a los distintos años (eje de las x). De esta forma elimino la estacionalidad y es más sencillo ver que presenta una tendencia ascendente, lo cual quiere decir que el número de casos en España parece que va a ir aumentando a medida que el tiempo pasa. También se puede observar lo mencionado anteriormente del periodo estacionario que hay entre los años 1989 y 1998.

Apartado c: Esquema de la Serie Temporal.

```
gu_año <- as.numeric(aggregate(Enf_GU, FUN = sum))
gu_añosd <- as.numeric(aggregate(Enf_GU, FUN = sd))
plot(gu_año, gu_añosd, pch = 20, col = "Blue", xlab = "Número de casos",
     ylab = "Desviación Típica", main = "Desviación típica-Casos")
abline(lm(gu_añosd ~ gu_año), lty = 2)
```



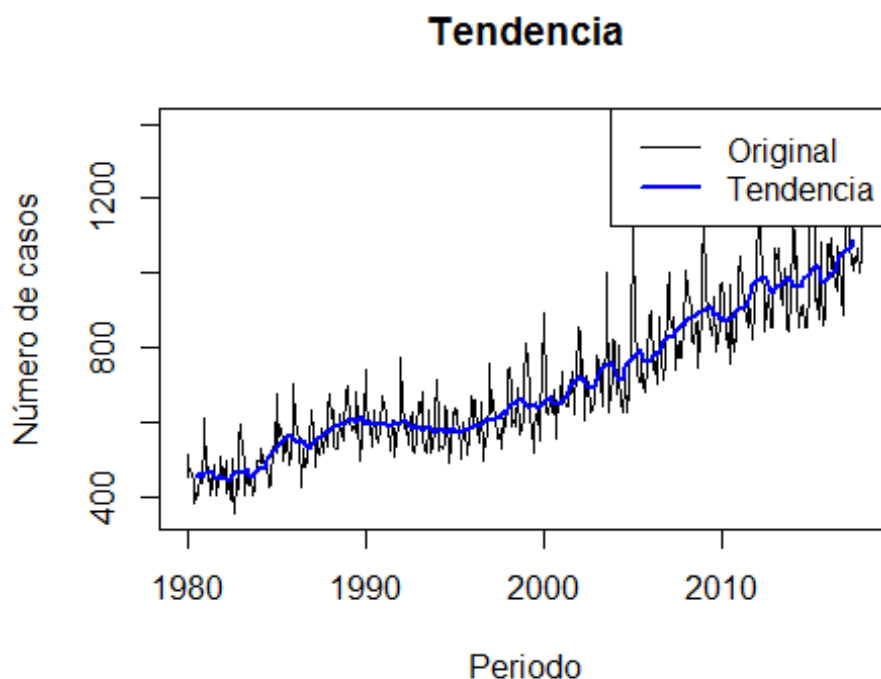
Para poder ver con claridad si la serie presenta un esquema multiplicativo o aditivo, represento la suma del número de fallecimientos en España a causa de enfermedades que afectan al sistema genitourinario (eje de abscisas) con respecto a la desviación típica intra-anual (eje de ordenadas). El resultado es una relación lineal entre ambas variables, esto indica que a medida que aumenta el número de casos la variabilidad aumenta, confirmando así que la serie sigue un esquema multiplicativo.

Si el esquema de la serie fuese por el contrario aditivo la desviación típica se mantendría constante entre un intervalo de valores a lo largo del tiempo, por lo que se vería una línea horizontal.

Apartado d: Descomposición de la serie.

```
egu_decmul <- decompose(Enf_GU, type = "mult") # Descomposición de la
serie gráfica en cada componente.

tendencia <- window(egu_decmul$trend, start = c(1980, 7), end = c(2017,
6))
ts.plot(Enf_GU, tendencia, plot.type = "single", col = c("Black", "Blue"),
lwd = c(1, 2), xlab = 'Periodo', main = 'Tendencia', ylab = 'Número de casos')
legend("topright", legend = c('Original', 'Tendencia'),
col = c("Black", "Blue"), lty = c(1, 1), lwd = c(1, 2)) # Tendencia
```



En este gráfico se representa la tendencia superpuesta a la serie temporal. Como he dicho anteriormente la tendencia es creciente, lo cual quiere decir que a medida que pasa el tiempo el número de casos va a incrementarse. Además se aprecia bien que de 1980 a 1989 la tendencia es creciente, mientras que a partir de 1989 se vuelve ligeramente estacionaria volviendo a cambiar en 1998 a una tendencia creciente.

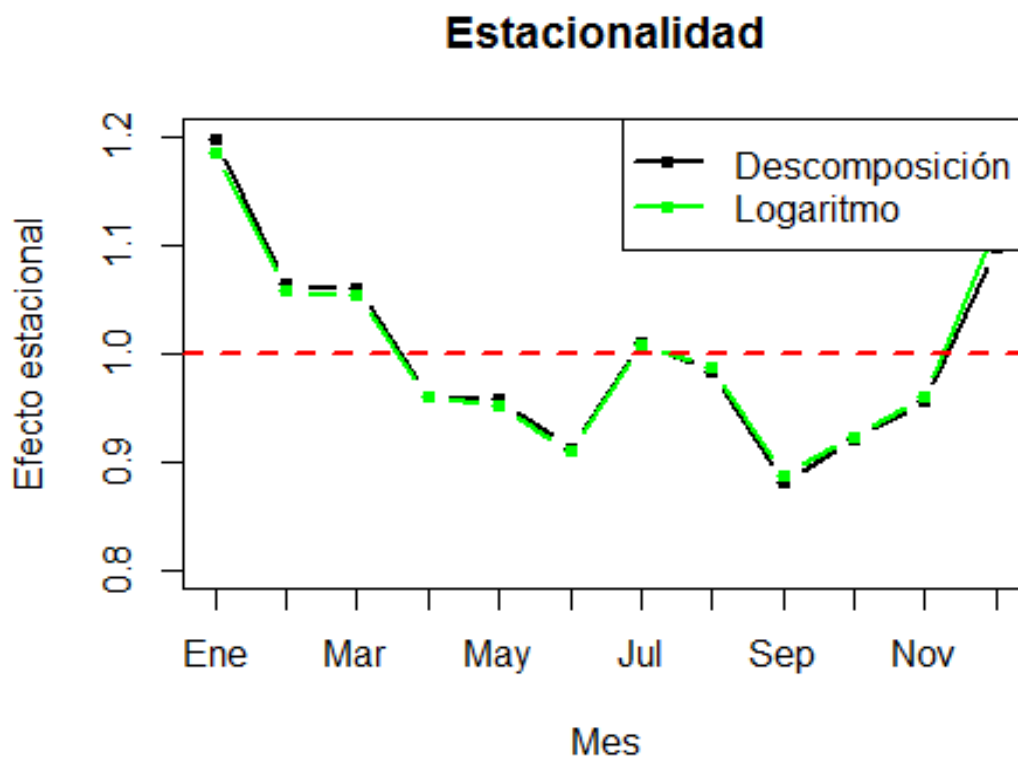
Este comportamiento puede ser debido a que las condiciones de vida 1980 a 1989 mejoraron un poco y eso aumentó la esperanza de vida, de modo que las personas al vivir más, desarrollaban más enfermedades y entre ellas las que afectan al sistema genitourinario. El que se mantenga estacionaria puede indicarme que esas condiciones de calidad de vida se mantuvieron a lo largo de ese periodo y a partir de 1998 volvieron a mejorar de tal forma que la esperanza de vida ha ido aumentando gradualmente y con ella el número de fallecimientos debidos a estas enfermedades.

```

estacionalidad <- tapply(Enf_GU / mean(Enf_GU), cycle(Enf_GU), mean)

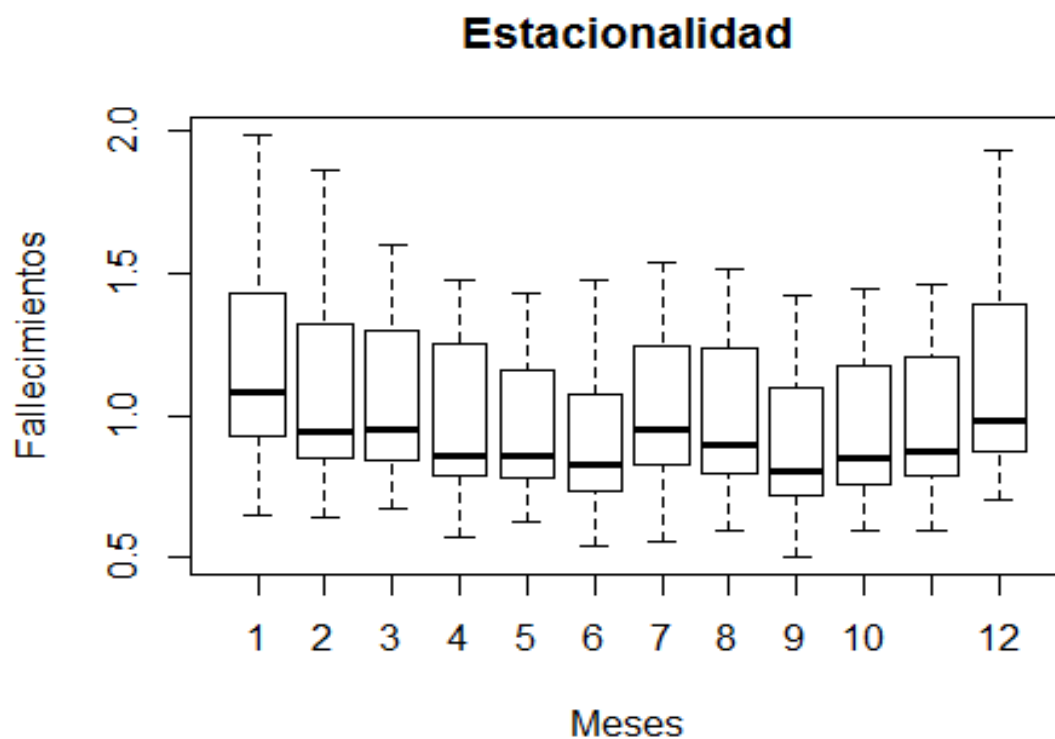
plot(egu_decmul$figure, type = 'b', xlab = 'Mes', ylab = 'Efecto
estacional', col = "Black", pch = 20, lwd = 2, xaxt = "n", ylim = c(0.8,
1.2), main = 'Estacionalidad')
lines(1:12, estacionalidad, col = "Green", pch = 20, lwd = 2, type = "b")
axis(1, at = 1:12, labels = c("Ene", "Feb", "Mar", "Abr", "May", "Jun",
"Jul", "Ago", "Sep", "Oct", "Nov", "Dic"))
abline(h = 1, lty = 2, lwd = 2, col = "Red")
legend("topright", legend = c('Descomposición', 'Logaritmo'), col =
c("Black", "Green"), pch = 20, lwd = 2)

```



Aquí se ilustra la componente estacional estimada a partir del método de descomposición (línea negra) y la componente estacional estimada a partir la media intra-anual y el ratio de la serie. Como se puede comprobar, ambas estimaciones son muy similares, en donde la componente estacional estimada a partir del método descriptivo parece presenta un efecto estacional más suave que la estimada por el método de descomposición.

```
boxplot((Enf_GU / mean(Enf_GU)) ~ cycle(Enf_GU), outline =
FALSE,main='Estacionalidad', xlab='Meses', ylab = 'Fallecimientos')
```



```
tapply(Enf_GU / mean(Enf_GU), cycle(Enf_GU), mean) # Estacionalidad
```

	1	2	3	4	5	6	7	8
##	1.1859558	1.0581184	1.0530318	0.9597532	0.9520859	0.9105331	1.0078511	0.9869065
##	9	10	11	12				
##	0.8874566	0.9232495	0.9597532	1.1153049				

De manera numérica y gráfica se observa cómo va variando el número de fallecimientos a lo largo del año usando los valores de los distintos años. Aquellos meses que son mas fríos la incidencia es mayor (1.115 diciembre, 1.186 enero, 1.058 febrero y 1.053 marzo) y a medida que el clima se suaviza también lo hace esta incidencia. Cabe mencionar que en el mes de julio, uno de los más cálidos, hay un ligero aumento en el número de fallecimientos (1.008). Por lo tanto, las enfermedades que afectan al sistema genitourinario se manifiestan de forma más acusada tanto en meses muy fríos como en meses muy calurosos como lo es el mes de julio.

Apartado e: Descripción detallada.

Como recopilación de lo visto con anterioridad, la serie temporal de enfermedades del sistema genitourinario presenta una tendencia ascendente siendo estacionaria entre los años de 1990 y 1998 aproximadamente, y en 1998 vuelve a retomar esa ascendencia. Esto puede explicarse en relación al aumento de la calidad de vida que ha prolongado la esperanza de vivir más años y con ello el aumento de las diversas enfermedades.

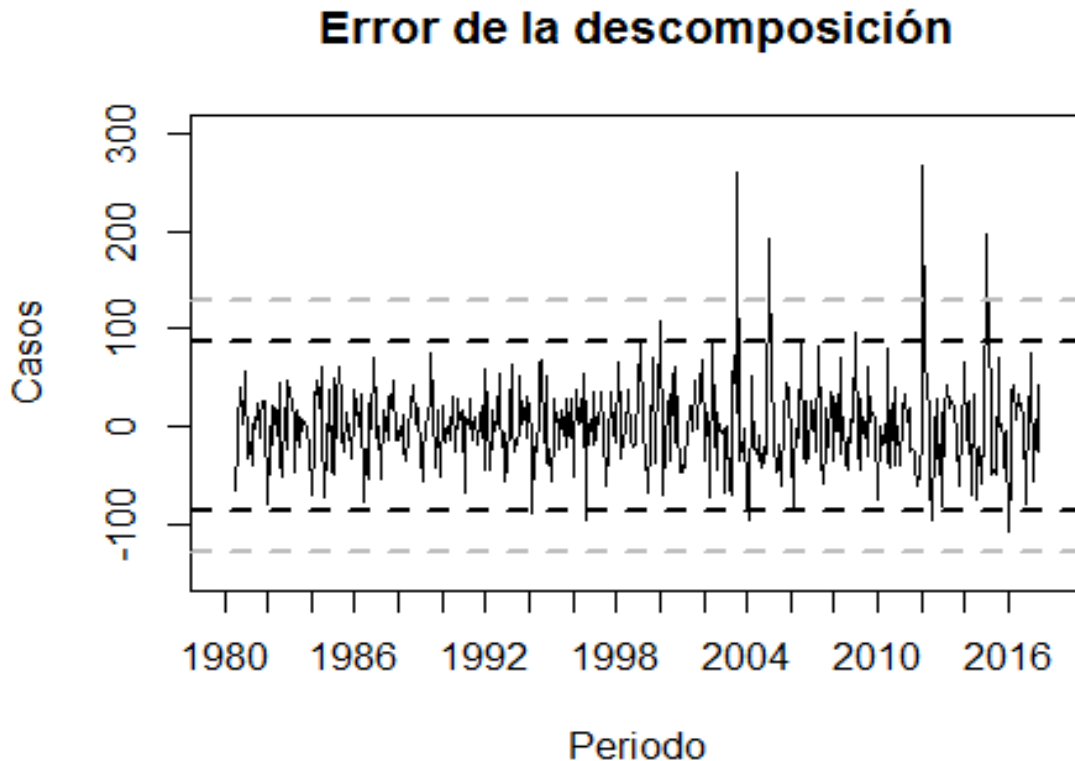
Con respecto a la estacionalidad se ha podido ver que en los meses de diciembre, enero, febrero y marzo la incidencia es mayor, lo cual correspondería a los meses más frío y hay además hay un pequeño aumento también en el mes de julio, un mes cálido y esto puede estar relacionado con el cambio de estación de primavera a verano.

Aparte, con respecto al residuo, se identifican varios valores atípicos (intervenciones) a lo largo de la serie en donde la cantidad de fallecimientos se dispara en un mes concreto como ocurre en octubre de 2003, enero de 2005, febrero de 2012 y enero de 2015. Cabe mencionar que también hay valores ligeramente altos y bajos a lo largo de la serie que superan las dos desviaciones típicas, sin embargo, la mayor parte del residuo queda dentro de esas dos desviaciones por lo que puede considerarse ruido blanco.

Para finalizar, en cuanto al esquema que sigue la serie temporal se ha visto que es multiplicativo ya que la desviación típica intra-anual aumenta en relación al número de casos de fallecimientos.

Ejercicio 2. Análisis del residuo.

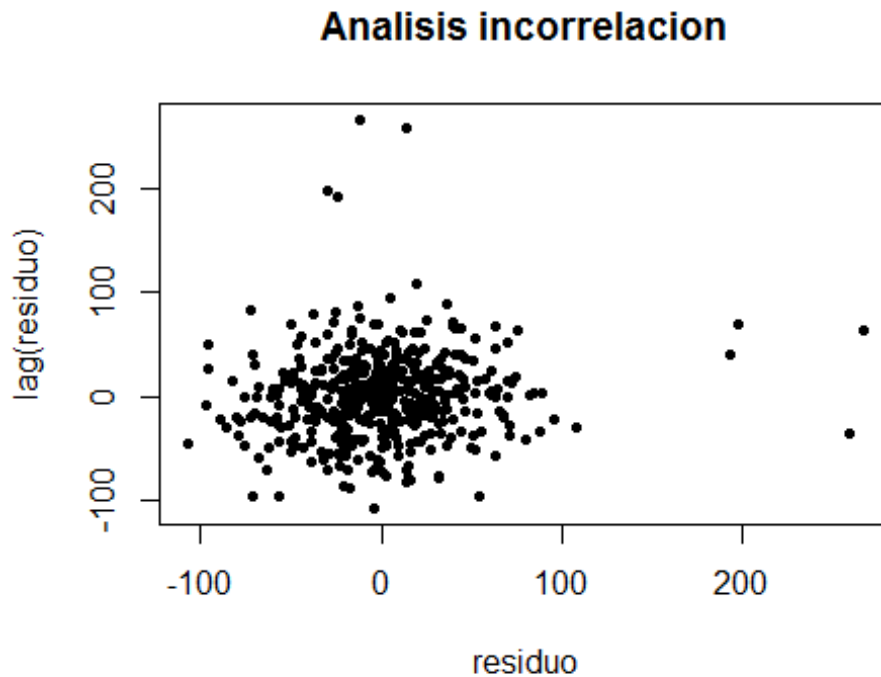
```
residuo <- egu_decmul$x - egu_decmul$trend*egu_decmul$seasonal
se <- sd(residuo, na.rm = TRUE) #Calculo sus desviaciones típicas.
plot(residuo,
     xlab = 'Periodo', ylab = 'Casos', xaxt = "n",
     main = "Error de la descomposición",
     col = "Black", lwd = 1, ylim = c(-150, 300))
axis(1, at = seq(1980, 2017, 2))
abline(h = c(-3 * se, -2 * se, 2 * se, 3 * se),
      lty = 2, lwd = 2, col = c("Grey", "Black", "Black", "Grey"))
#residuo > 3*se #Intervenciones.
```



En el gráfico anterior se observa cómo varía el residuo a lo largo de los años y parece heterocedástico e incorrelado. Además, se aprecian varios valores atípicos (intervenciones) que superan las tres desviaciones típicas (99.7%) por la superior: octubre de 2003, enero de 2005, febrero de 2012 y enero de 2015, que superaron los valores reales los casos estimados. Sin embargo, por la parte inferior no hay ningún mes que sobrepase las tres desviaciones típicas.

Además, se puede observar que a lo largo de los años el error se mantiene entre dos desviaciones típicas, lo cual puede considerarse ruido blanco

```
plot(residuo, lag(residuo), pch = 20, main = "Análisis incorrelacion")  
#Incorrelación
```

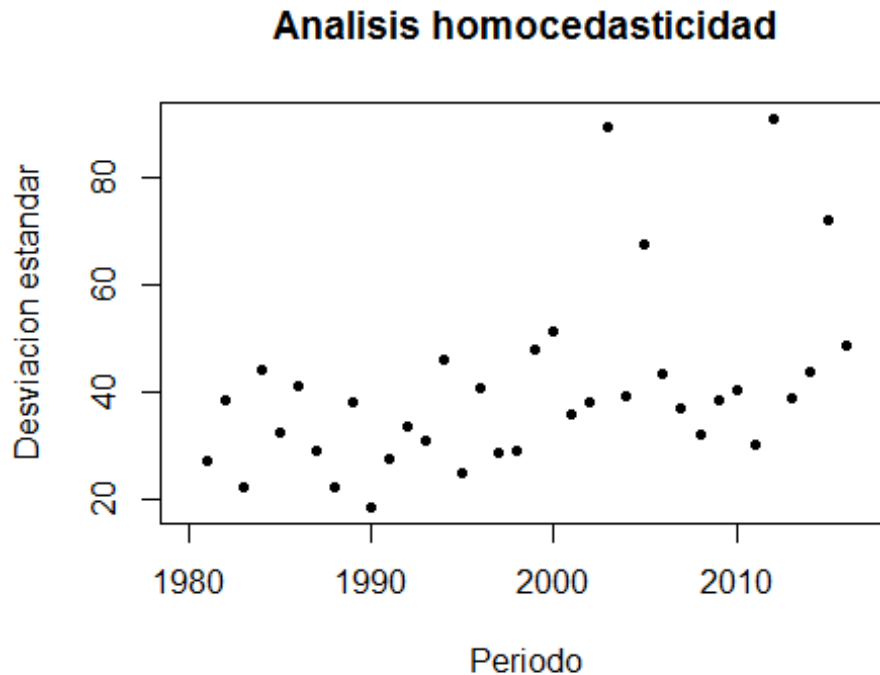


Por otro lado, cuando realizo el análisis de incorrelación, observo que el residuo se distribuye como una nube de puntos sin mostrar patrón alguno, lo cual me informa de que el residuo de la serie temporal es incorrelado.

```

xd <- aggregate(residuo,FUN=sd)
plot(xd,type = "p", pch = 20,
xlab = 'Periodo', ylab = "Desviacion estandar",
main = "Análisis homocedasticidad") #Heterocedasticidad

```



Para finalizar me fijo en el gráfico que enfrenta la desviación estándar con los distintos años y veo que el residuo es heterocedástico, pues no se mantiene constante a medida que pasa el tiempo sino que sus valores varían en una franja de valores. Se aprecia que a medida que pasa el tiempo la desviación típica va aumentando ligeramente y en ocasiones aparecen valores atípicos con una desviación bastante grande a diferencia del resto.

Haciendo uso de la función `locate()` identifiqué que estos valores no se corresponden con los valores atípicos que superan las 3 desviaciones típicas, que era lo que esperaba, sino que corresponden a los valores que sobrepasan las dos desviaciones típicas.