**Solucionario**

1. Considere la serie pigs: La cantidad de cerdos sacrificados cada mes.

a. Use la función ses() para encontrar los valores óptimos de α y ℓ. Y genere pronósticos

para los próximos 4 meses.

library(fpp2)

library(seasonal)

help(pigs)

pigsdata <- window(pigs, start=1980)

autoplot(pigsdata) +

ggtitle("Total de cerdos sacrificados en Victoria-Australia") +

ylab("# Abs") + xlab("Year")

# Estimate parameters

fc <- ses(pigsdata, h=4)

# Precisión de los errores de entrenamiento de un paso adelante

round(accuracy(fc),2)

summary(fc[["model"]])

autoplot(fc) +

autolayer(fitted(fc), series="Fitted") +

ylab("Total") + xlab("Year")

autoplot(fc) +

autolayer(fitted(fc), series="SES method") +

ggtitle("Forecasts SES de cerdos sacrificados en Victoria-Australia") + xlab("Year") +

ylab("Total") +

guides(colour=guide\_legend(title="Forecast"))

Los valores óptimos de α y ℓ son 0.2971 y 77260.0561, respectivamente.

**Figura 1**

Forecast SES de cerdos sacrificados en Victoria-Australia

|  |
| --- |
|  |
|  |

2.El conjunto de datos *books* contiene las ventas diarias de libros en rústica y de tapa dura en la misma tienda. La tarea es pronosticar las ventas de los próximos cuatro días para libros de bolsillo y de tapa dura.

a. Trace la serie y discuta las características principales de los datos.

#Serie original

help(books)

autoplot(books) +

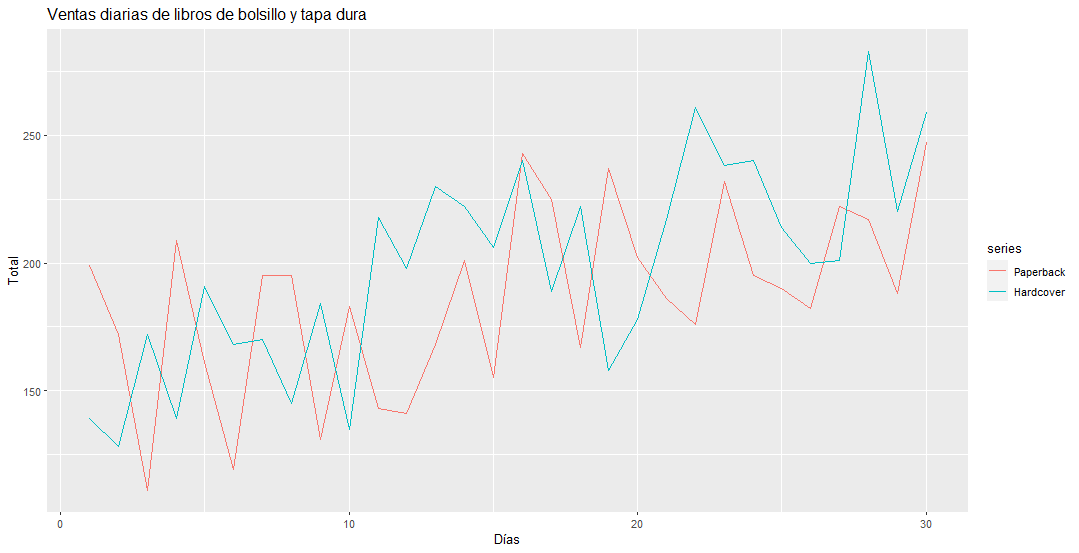
ggtitle("Ventas diarias de libros de bolsillo y tapa dura") +

ylab("Total") + xlab("Year")

Como se observa en la figura 1, ambas series tienen tendencia clara y un componente estacional.

**Figura 1**

*Ventas diarias de libros de bolsillo y tapa dura*



b. Use la función ses() para pronosticar cada serie y trazar los pronósticos.

# Serie original

bookbols <- ts(books[,1], start =1)

bookduro <- ts(books[,2], start =1)

# Estimar parámetros books de bolsillo

fc1 <- ses(bookbols, h=4)

fc2 <- ses(bookduro, h=4)

#Gráfico de libro de bolsillo

autoplot(fc1) +

autolayer(fitted(fc1), series="Serie suavizada", PI=TRUE) +

autolayer(fc1, series="pronóstico") +

ggtitle("Forecasts SES de ventas diarias de libros de bolsillo ") + xlab("Year") +

ylab("") +

guides(colour=guide\_legend(title="Forecast"))

#Gráfico de libro de tapa dura

autoplot(fc2) +

autolayer(fitted(fc2), series="Serie suavizada", PI=TRUE) +

autolayer(fc2, series="pronóstico") +

ggtitle("Forecasts SES de ventas diarias de libros de tapa dura ") + xlab("Year") +

ylab("") +

guides(colour=guide\_legend(title="Forecast"))

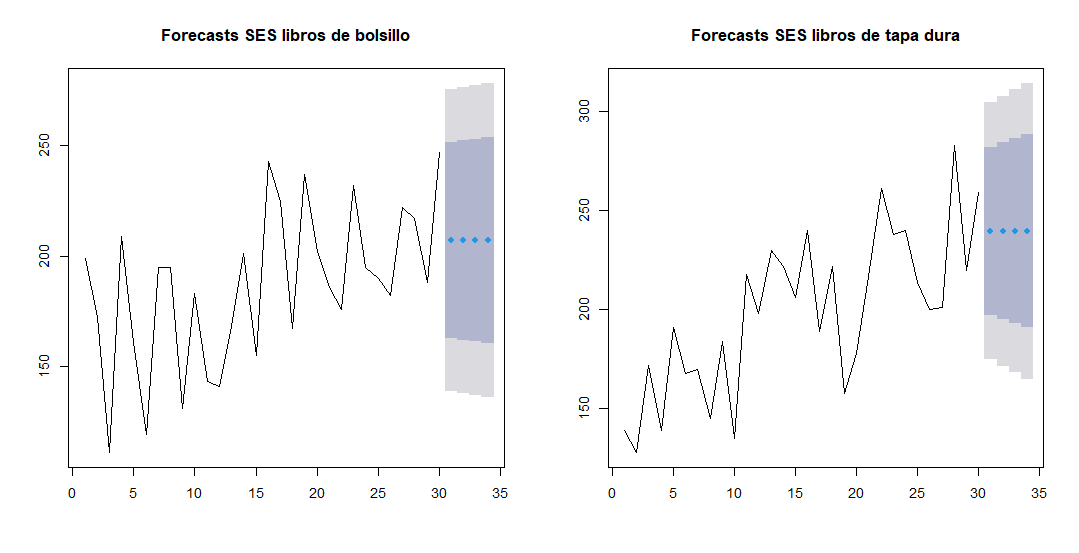
par(mfrow = c(1,2))

plot(fc1, main="Forecasts SES libros de bolsillo")

plot(fc2, main="Forecasts SES libros de tapa dura")

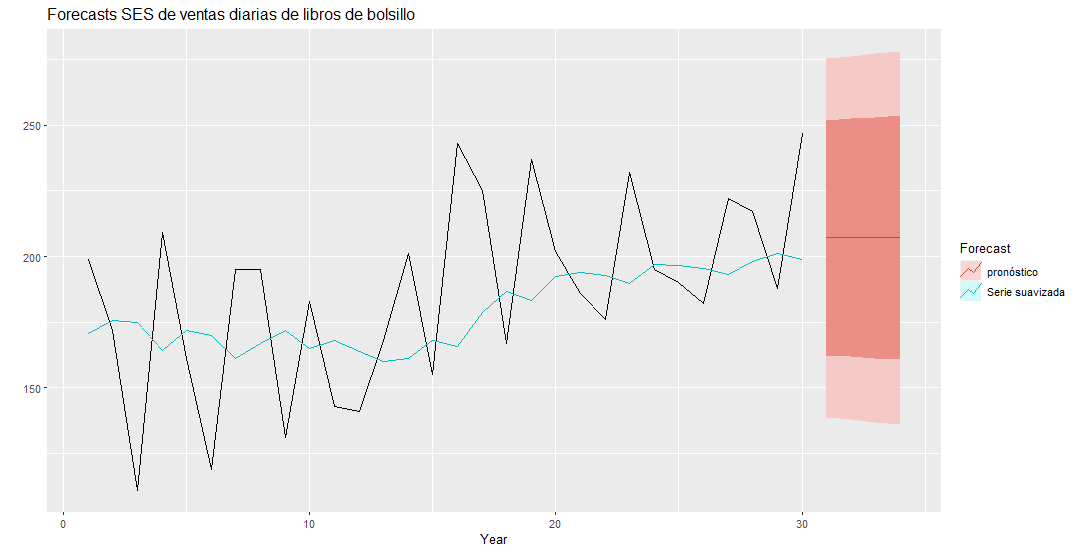
**Figura 2**

*Pronósticos SES de libros de bolsillo “en rústica” y de tapa dura*



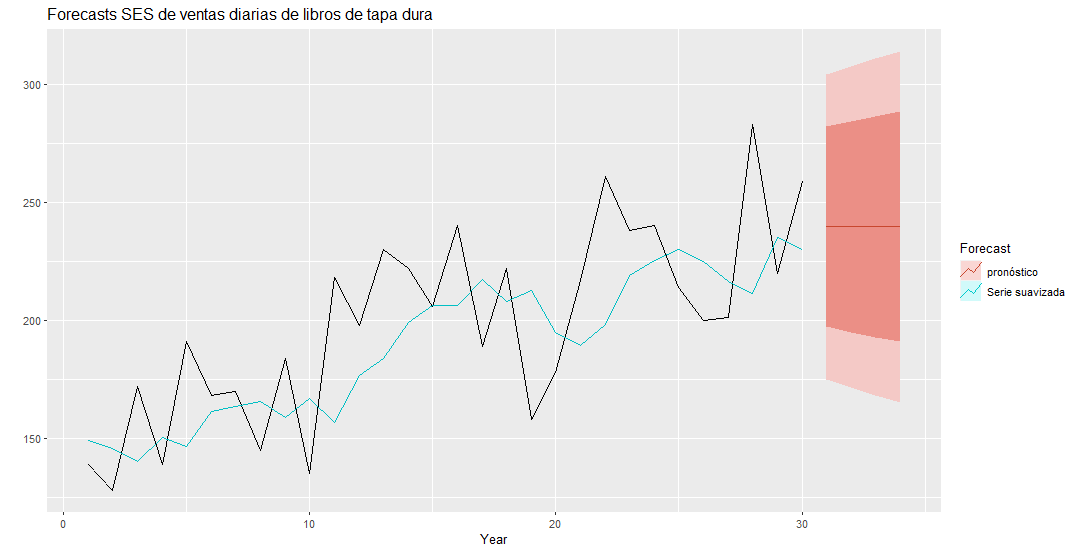
**Figura 3**

*Pronóstico SES de ventas diarias de libros de bolsillo*



**Figura 4**

*Pronóstico SES de ventas diarias de libros de tapa dura*

****

c. Calcule los valores RMSE para los datos de entrenamiento en cada caso.

original

# Estimar parámetros books de bolsillo

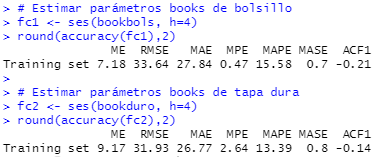
fc1 <- ses(bookbols, h=4)

round(accuracy(fc1),2)

# Estimar parámetros books de tapa dura

fc2 <- ses(bookduro, h=4)

round(accuracy(fc2),2)



Para la serie *en rústica* el RSME fue de 33,64 libros y para tapa dura fue de 31,93.

3. Continuaremos con las ventas diarias de libros de bolsillo y tapa dura del conjunto de datos books.

a. Aplicar el método lineal de Holt a las series paperback y las hardback. Calcular los pronósticos de cuatro días en cada caso.

# books de bolsillo y tapa dura

fc1 <- holt(bookbols, h=4)

fc2 <- holt(bookduro, h=4)

autoplot(fc1) +

autolayer(fitted(fc1), series="Serie suavizada", PI=TRUE) +

autolayer(fc1, series="pronóstico") +

ggtitle("Forecast lineal Holt de ventas diarias de libros de bolsillo ") + xlab("Year") +

ylab("") +

guides(colour=guide\_legend(title="Forecast"))

autoplot(fc2) +

autolayer(fitted(fc2), series="Serie suavizada", PI=TRUE) +

autolayer(fc2, series="pronóstico") +

ggtitle("Forecast lineal de Holt de ventas diarias de libros de tapa dura ") + xlab("Year") +

ylab("") +

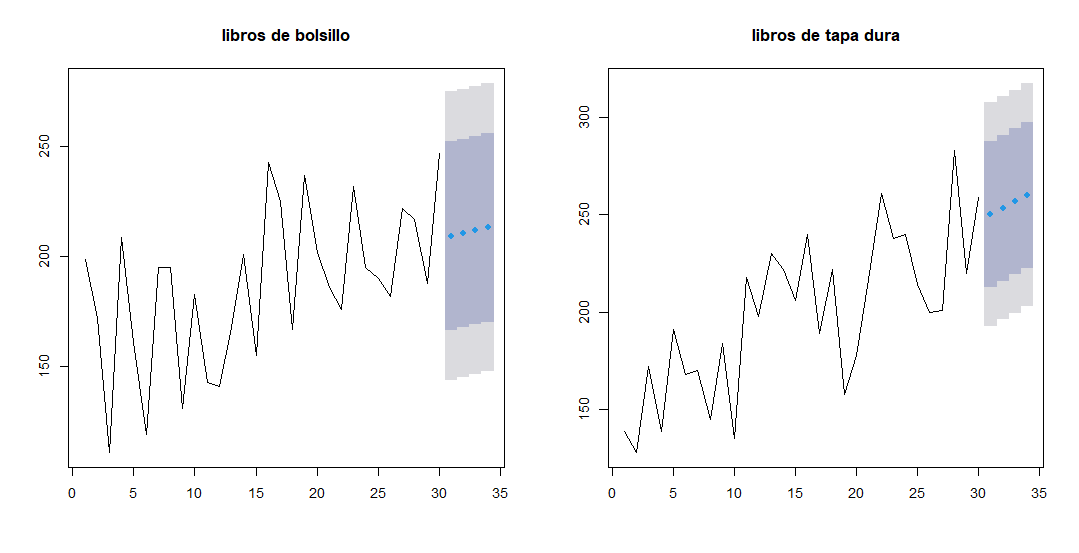
guides(colour=guide\_legend(title="Forecast"))

par(mfrow = c(1,2))

plot(fc1, main="libros de bolsillo")

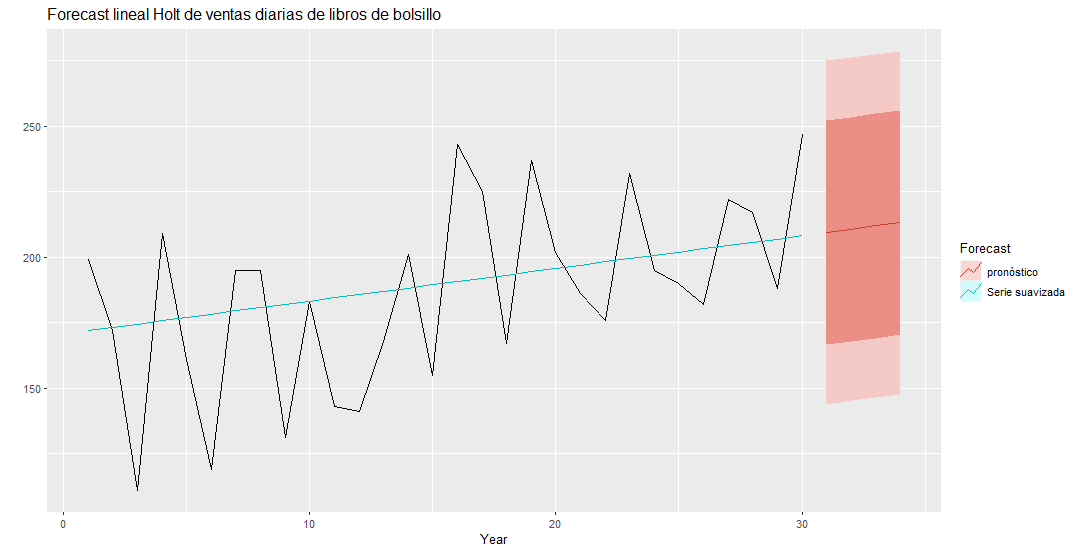
plot(fc2, main="libros de tapa dura")

**Figura 5**

*Pronóstico lineal de Holt para libros de bolsillo “en rústica” y de tapa dura* ****

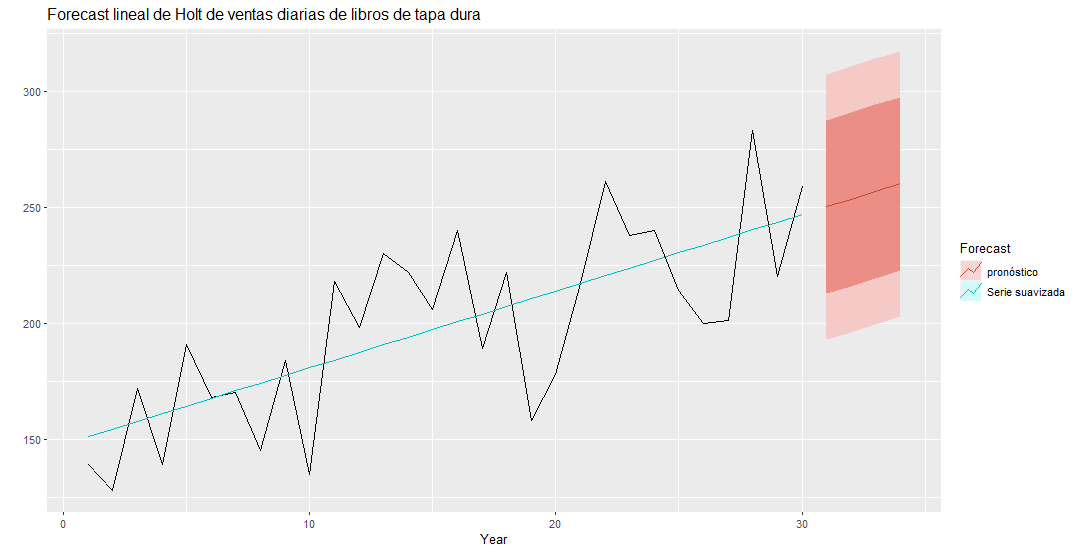
**Figura 6**

*Pronóstico lineal de Holt para ventas diarias de libros de bolsillo*



**Figura 7**

*Pronóstico lineal de Holt para ventas diarias de libros de tapa dura*

****

b. Compare las medidas RMSE del método de Holt para las dos series con las del suavizado exponencial simple en la pregunta anterior. (Recuerde que el método de Holt está usando un parámetro más que SES). Discuta los méritos de los dos métodos de pronóstico para estos conjuntos de datos.

#3.b

# libro de bolsillo

fcses <- ses(bookbols, h=4)

fcholt <- holt(bookbols, h=4)

round(accuracy(fcses),2)

# libros de tapa dura

round(accuracy(fcholt),2)

fcses2 <- ses(bookduro, h=4)

fcholt2 <- holt(bookduro, h=4)

round(accuracy(fcses2),2)

round(accuracy(fcholt2),2)

Para ambas series el método de pronóstico lineal Holt fue el mejor por obtener el menor error cuadrático medio. La medida de RMSE para los libros de bolsillo usando el método SES fue de 33,64 y para el de Holt fue de 31,14. Asimismo para el de tapa dura fue de 31,93 y 27,19, respectivamente.

Cabe recordar que el método de Holt se usa cuando la serie tiene una clara tendencia, lo contrario ocurre con el SES. Al observar las series estas tienen una tendencia evidente al alza, por lo que el primer método es el más adecuado.

**Tabla 1**

*Evaluación de la precisión del pronóstico*

|  |  |  |
| --- | --- | --- |
| Tipo de libros | RMSE usando SES | RMSE usando Holt |
| De bolsillo | 33,64 | 31,14 |
| De tapa dura | 31,93 | 27,19 |

c. Compare los pronósticos para las dos series usando ambos métodos. ¿Cuál crees que

es mejor?

#3.c

autoplot(bookbols) +

autolayer(fitted(fcses), series="suavizado SES") +

autolayer(fitted(fcholt), series="suavizado Holt") +

autolayer(fcses, series="pronóstico SES", PI=FALSE) +

autolayer(fcholt, series="pronóstico lineal Holt", PI=FALSE) +

ggtitle("Forecast SES y lineal Holt de ventas diarias de libros de bolsillo ") + xlab("Year") +

ylab("") +

guides(colour=guide\_legend(title="Forecast"))

autoplot(bookduro) +

autolayer(fitted(fcses2), series="suavizado SES") +

autolayer(fitted(fcholt2), series="suavizado Holt") +

autolayer(fcses2, series="pronóstico SES", PI=FALSE) +

autolayer(fcholt2, series="pronóstico lineal Holt", PI=FALSE) +

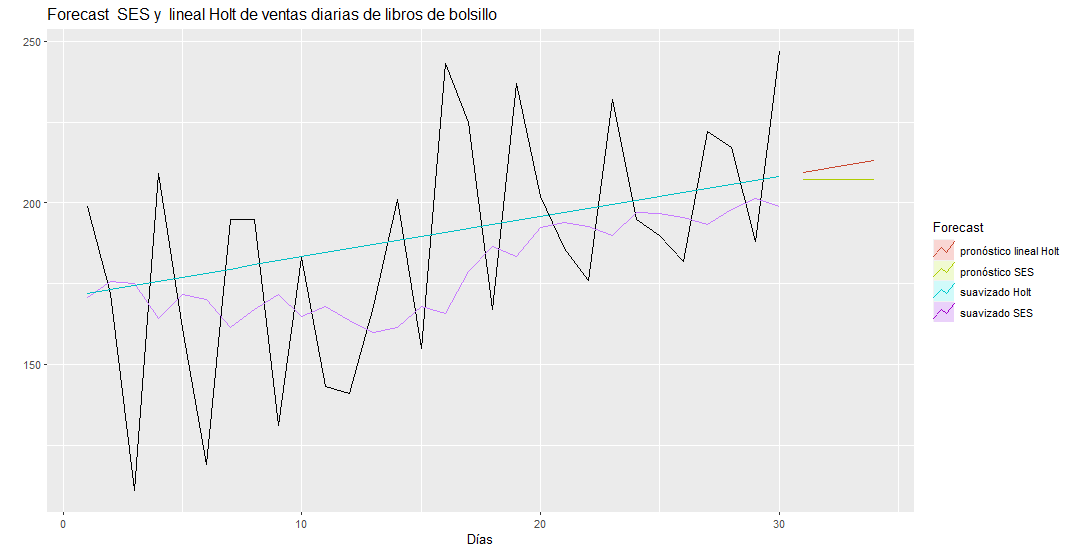
ggtitle("Forecast SES y lineal Holt de ventas diarias de libros de tapa dura ") + xlab("Year") +

ylab("") +

guides(colour=guide\_legend(title="Forecast"))

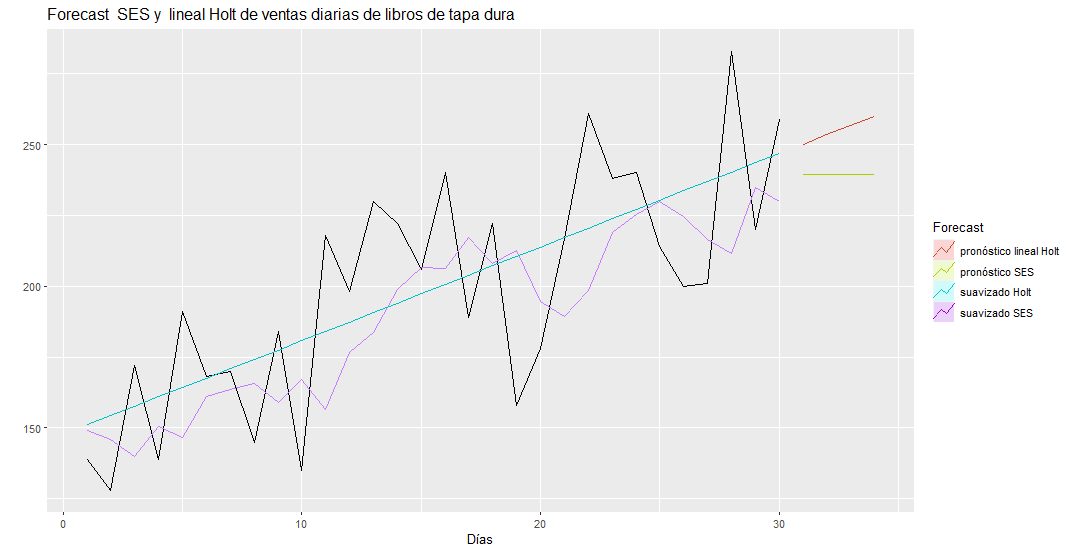
**Figura 4**

*Pronóstico SES y lineal Holt de las ventas diarias de libros de bolsillo*

****

**Figura 5**

*Pronóstico SES y lineal Holt de las ventas diarias de libros de tapa dura*



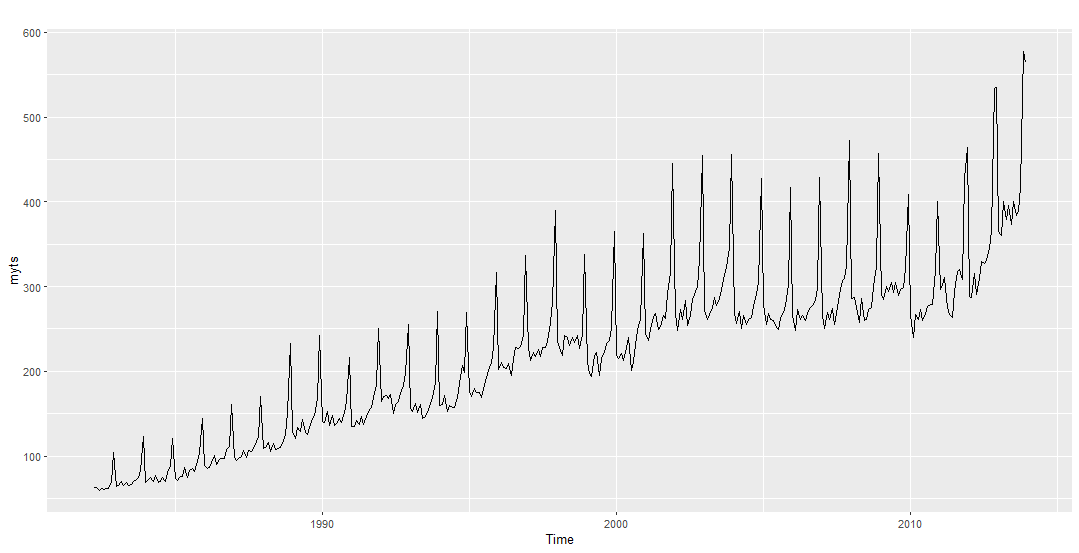
d. Calcule un intervalo de predicción del 95% para el primer pronóstico para cada serie, utilizando los valores RMSE y suponiendo errores normales. Compare sus intervalos con los producidos usando ses y holt.

4. Recuerde sus datos de series de tiempo de venta minorista.

a. ¿Por qué es necesaria la estacionalidad multiplicativa para esta serie?

**Figura 6**

*Serie de tiempo de venta minorista*



Cuando la estacionalidad de la serie aumenta proporcionalmente con el tiempo se usa la descomposición multiplicativa y cuando no es el caso, conviene utilizar la aditiva. En la figura 6 se observa la serie de tiempo minorista y el aumento de su estacional conforme pasa el tiempo, por eso es necesario usar el método multiplicativo.

b. Aplique el método multiplicativo de Holt-Winters a los datos. Experimente haciendo que la tendencia se amortigüe.

# 4.b

fit1 <- hw(myts,seasonal="multiplicative")

fit2 <- hw(subset(myts,end=length(myts)),

damped = TRUE, seasonal="multiplicative")

autoplot(myts) +

autolayer(fit1, series="HW multiplicativo ",

PI=FALSE) +

autolayer(fit2, series="HW multiplicativo amortiguado ",

PI=FALSE) +

xlab("year") +

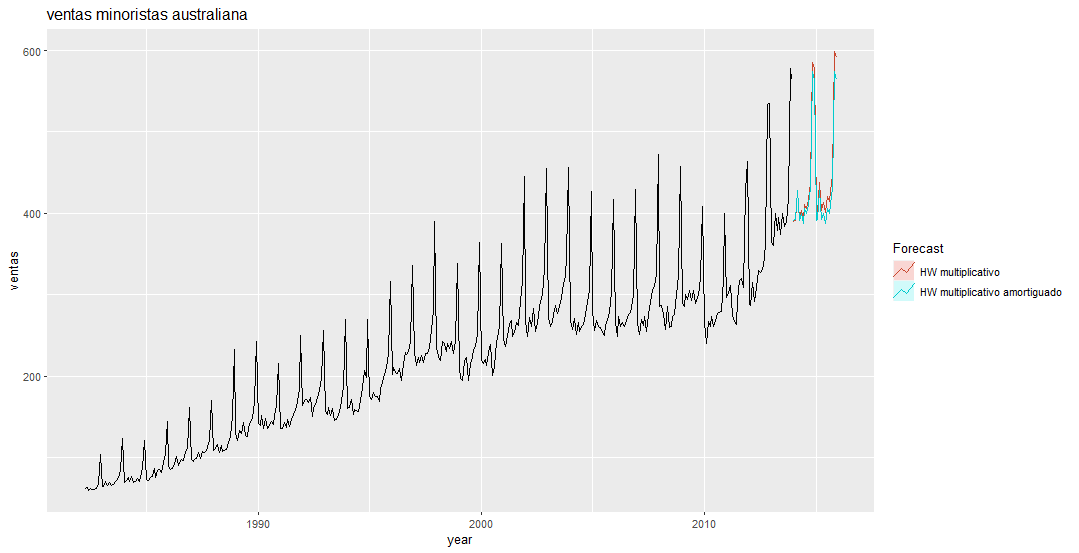
ylab("ventas") +

ggtitle("ventas minoristas australiana") +

guides(colour=guide\_legend(title="Forecast"))

**Figura 7**

*Pronóstico de venta minorista australiana usando el método Holt-Winters multiplicativo*



c. Compare el RMSE de los pronósticos de un paso de los dos métodos. ¿Cuál prefieres?

round(accuracy(fit1),2)

round(accuracy(fit2),2)

**Tabla 2**

*Evaluación de la precisión del pronóstico*

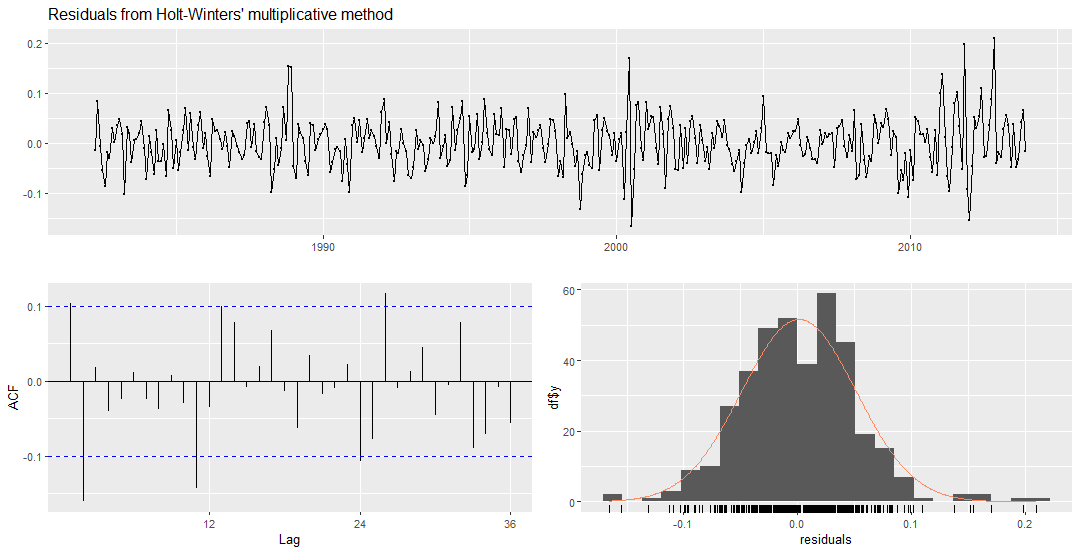
|  |  |  |
| --- | --- | --- |
| Serie | RMSE usando HW Multiplicativo no amortiguado | RMSE usando HW multiplicativo amortiguado |
| Venta Minorista australiana | 13,29 | 13,3 |

Al observar la Tabla 2 se prefiere usar el método Holt-Winter multiplicativo no amortiguado ya que su RMSE es menor, aunque ciertamente es indiferente usar ambos métodos por tener una diferencia de 0,01.

d. Verifique que los residuos del mejor método se vean como ruido blanco.

**Figura 8**

*Prueba gráfica si los residuos son ruido blanco y normalmente distribuidos*.



# Ljung-Box test

Ho: Los datos se distribuyen de forma independiente; no presentan autocorrelación (la serie es ruido blanco).

H1: Los datos no se distribuyen de forma independiente; exhiben correlación serial (la serie no es ruido blanco).

# data: Residuals from Seasonal naive method

# Q\* = 40.405, df = 8, p-value = 0,000002692

# Model df: 16. Total lags used: 24

Observando la figura 8 los residuos están correlacionados en varios rezagos (no son ruido blanco), su media es cercana a cero y aparentemente se distribuye como una normal. Además, el test de Ljung-Box, concluye que el método de pronóstico Holt-Winters multiplicativo no amortiguado no da cuenta de toda la información de los datos. Se rechaza la hipótesis nula de que los residuos se distribuyen de forma independiente; no presentan autocorrelación, es decir, los residuos no son ruido blanco.

Recordemos que:

Los residuos son útiles para comprobar si un modelo ha capturado adecuadamente la información de los datos. Un buen método de pronóstico producirá residuos con las siguientes propiedades:

Los residuos no están correlacionados. Si hay correlaciones entre los residuos, entonces queda información en los residuos que debe usarse para calcular los pronósticos.

Los residuos tienen media cero. Si los residuos tienen una media distinta de cero, entonces los pronósticos están sesgados. (Hyndman & Athanasopoulos, s. f.)

5. La producción anual de carbón bituminoso en los Estados Unidos entre 1920 y 1968 se encuentra en el conjunto de datos *bicoal.*

a. Producir un diagrama de tiempo de los datos.

## 5.a

help(bicoal)

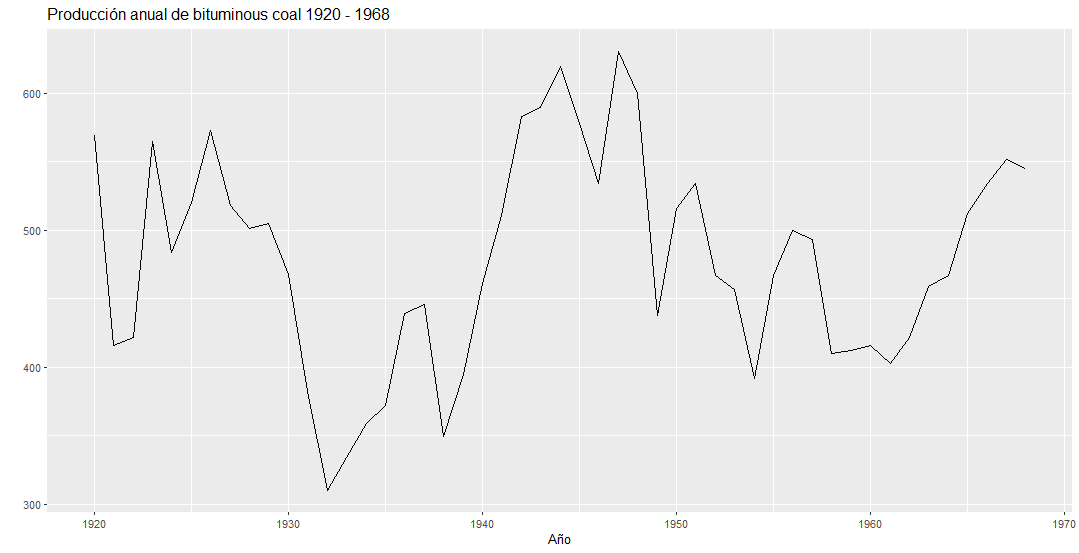
autoplot(bicoal) +

ggtitle("Producción anual de bituminous coal 1920 - 1968") +

ylab("") + xlab("Año")

**Figura 9**

*Producción anual de bituminous coal 1920 - 1968*



b. Decide ajustar el siguiente modelo a la serie



## 5.b

arima.bicoal <- Arima( bicoal, order=c(4,0,0) )

arima.bicoal

C=MEAN∗(1−ϕ1−ϕ2−ϕ3−ϕ4)

C=481.5221\*(1−0.8334+0.3443−0.5525+0.3780) =161,98403444

C=161,98

Yt = 161,98 +0.8334\*yt−1- 0.3443\* yt−2 + 0.5525\*yt−3 - 0.3780\*yt−4

c. Explique por qué se eligió este modelo utilizando ACF y PACF

## 5.c

ggAcf(bicoal)

ggPacf(bicoal)

(arima1.bicoal <- Arima(bicoal, order=c(4,0,0)))

ndiffs(bicoal)

(arima2.bicoal <- Arima(bicoal, order=c(3,0,0)))

**Figura 10**

*ACF de la serie producción anual de bituminous coal 1920 - 1968*

Imagen que contiene Gráfico

Descripción generada automáticamente

**Figura 11**

*PACF de la serie producción anual de bituminous coal 1920 - 1968*

Gráfico

Descripción generada automáticamente

Recordemos que las gráficas de autocorrelación ACF y PACF pueden ser útiles para determinar el valor de p o q de los modelos ARIMA(p,d,0) o ARIMA(0,d,q)

Al observar las figuras 10 y 11 concluimos que el ARIMA(p,d,0) es el más adecuado ya que ocurre lo siguiente:

-El ACF es sinusoidal

-Hay un pico significativo del retraso 4 en el PACF, pero ninguno más allá del retraso 4.

Además, al observar las gráficas, proponemos dos modelos y elegimos el que tiene un menor valor de criterio de información Akaike corregido AICc. Los modelos tentativos son ARIMA(4,0,0) y ARIMA(3,0,0) en el cual el primero tiene un valor AICc1<AICc2 (538.1<541.95). Por lo expuesto el mejor modelo que elegimos es el ARIMA(4,0,0).

6. Se cuentan con datos de contaminación por Ozono en una ciudad del interior del país, (*Datos*

*processed\_o3\_data.csv*), Suponga que lo contactan como experto en estadística para

analizar y tomar medidas al respecto ante niveles altos del contaminante. No escatime en sus

análisis. Además, es de suma importancia pronosticar los valores futuros. Recuerde validar

los modelos propuestos y plantear una medida de que tan buenos son los mismos.

Los datos elevados de Ozono son causales de enfermedades respiratorias en población

vulnerable, investigue cual es la normativa vigente que rige en su país y proponga los modelos

acorde a los pronósticos que permitan hacer una detección temprana de posibles niveles

nocivos para la salud humana.

**Paso 0: Limpieza de datos**

Completamos los valores *missing* con promedios. Luego trabajamos con los datos ajustados estacionalmente

o3\_data <- read.csv(

file = "processed\_o3\_data.csv",

header = TRUE,

sep = ",",

dec = ".",

stringsAsFactors = TRUE

)

#¿Reemplazar con ceros? ¡No!

#o3\_data[is.na(o3\_data)] <- 0

#o3\_data

#¿Reemplazamos con el promedio?

o3\_data$O3\_BA[is.na(o3\_data$O3\_BA)] <- mean(o3\_data$O3\_BA, na.rm = TRUE)

#o3\_data$O3\_BA <- round(o3\_data$O3\_BA, digits = 0)

#Crear un objeto de la clase ts: serie de tiempo diario.

help(ts)

o3a <- ts(o3\_data[,1], frequency = 7, start = c(2018-12-01))

#Analizar la data ajustada estacionalmente

o3a %>% stl(s.window='periodic') %>% seasadj() -> o3aadj

autoplot(o3aadj)

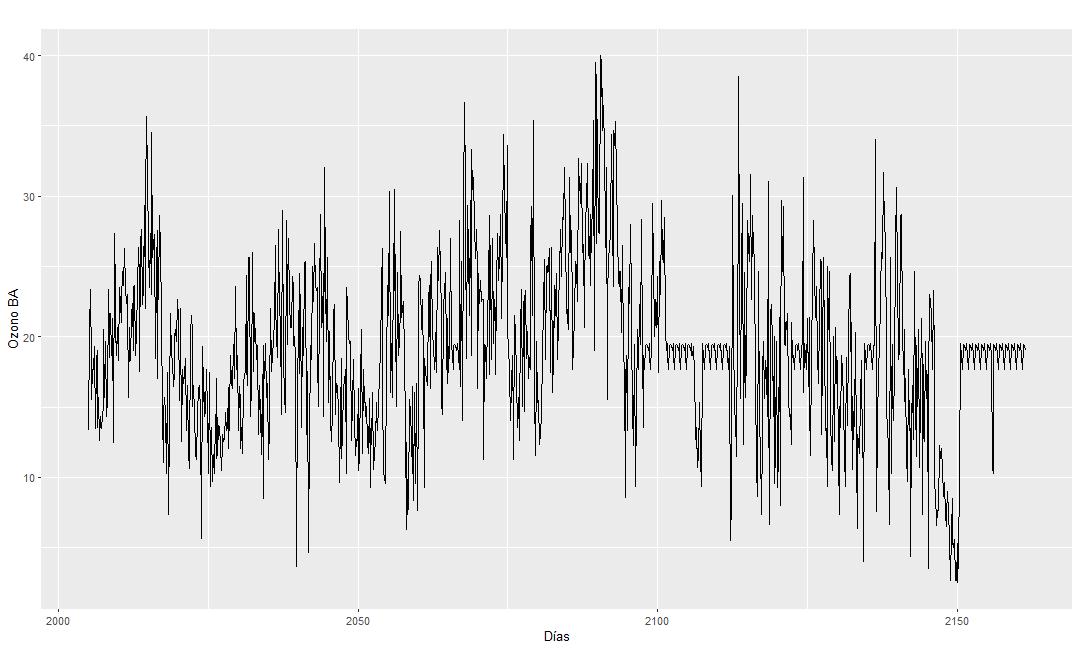
# Paso 1 : la gráfica inicial

autoplot(o3aadj) +

ylab("Ozono BA ajustada estacionalmente") + xlab("Días")

**Figura 12**

*Serie de Ozono BA diario imputado y ajustado estacionalmente del 1/1/2018 - 31/12/2020*



**Paso 2: Transformación Box Cox para estabilizar la varianza ¿Es necesario?**

(lambda <- BoxCox.lambda(o3aadj))

autoplot(BoxCox(o3aadj,lambda))

No utilizamos la transformación Box-Cox porque no simplifica los patrones de los datos históricos (tendencia, estacional y ciclo).

**Paso 3: ¿cuántas diferencias requiere la serie original?**

ndiffs(o3aadj)

#Sólo una diferenciación

o3aadj %>% diff() %>% ggtsdisplay(main="")

Con la primera diferencia la serie se vuelve estacionaria

**Paso 4: Observar el ACF y PACF para proponer modelos**

Al observar los gráficos ACF y PACF se proponen los modelos ARIMA(1,1,0) y el ARIMA(6,1,0). Por otro lado, el camino corto propone el ARIMA(2,1,2)

(fit\_o3a <- auto.arima(o3aadj,stepwise=FALSE,

approximation=FALSE,

seasonal=FALSE))

**Paso 5: Ajustamos los modelos propuestos y elegimos el que tiene el menor valor AICc**

(fit1 <- Arima(o3aadj, order=c(1,1,0)))

(fit2 <- Arima(o3aadj, order=c(6,1,0)))

(fit3 <- Arima(o3aadj, order=c(2,1,2)))

Elegimos el ARIMA(2,1,2) para hacer el pronóstico por tener el menor valor de citerio de información Akaike corregido AICc

**Paso 6: Test de ruido blanco para el residuo del mejor modelo**

checkresiduals(fit3)

# Ljung-Box test

Ho: Los datos se distribuyen de forma independiente; no presentan autocorrelación (la serie es ruido blanco).

H1: Los datos no se distribuyen de forma independiente; exhiben correlación serial (la serie no es ruido blanco).

# data: Residuals from from ARIMA(2,1,2)

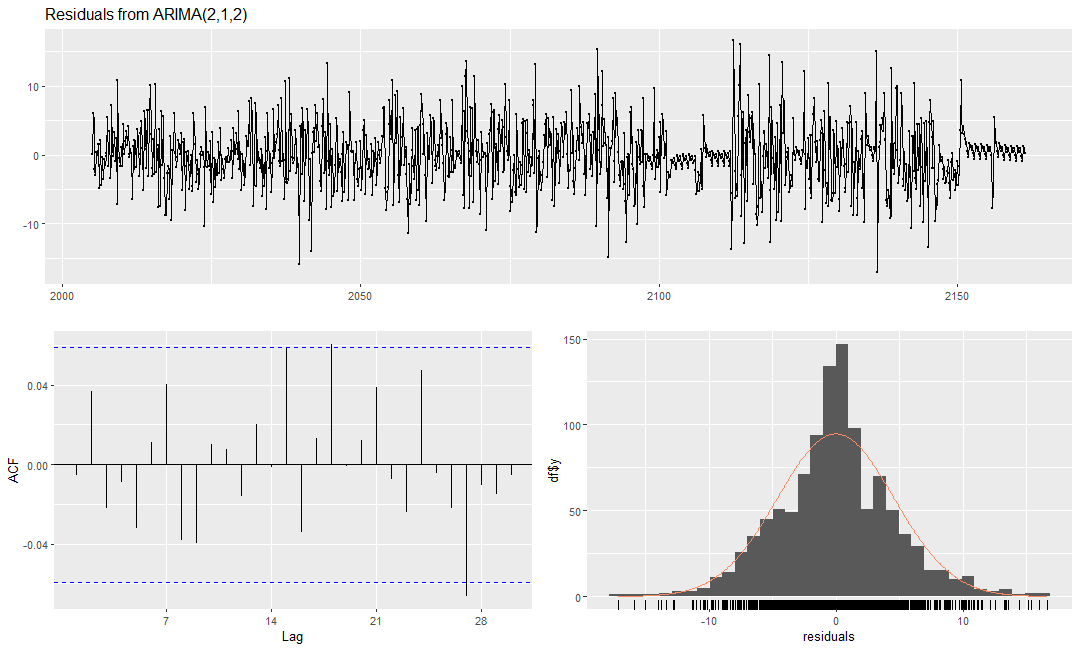
# Q\* = 9.4874, df = 10, p-value = 0.4866

# Model df: 4. Total lags used: 14

No se rechaza la hipótesis nula del test de Jung-Box de que el residuo es ruido blanco y esto se observa en el ACF de figura 13 (los rezagos están dentro de las bandas).

**Figura 13**

*Prueba gráfica si los residuos son ruido blanco y normalmente distribuidos.*

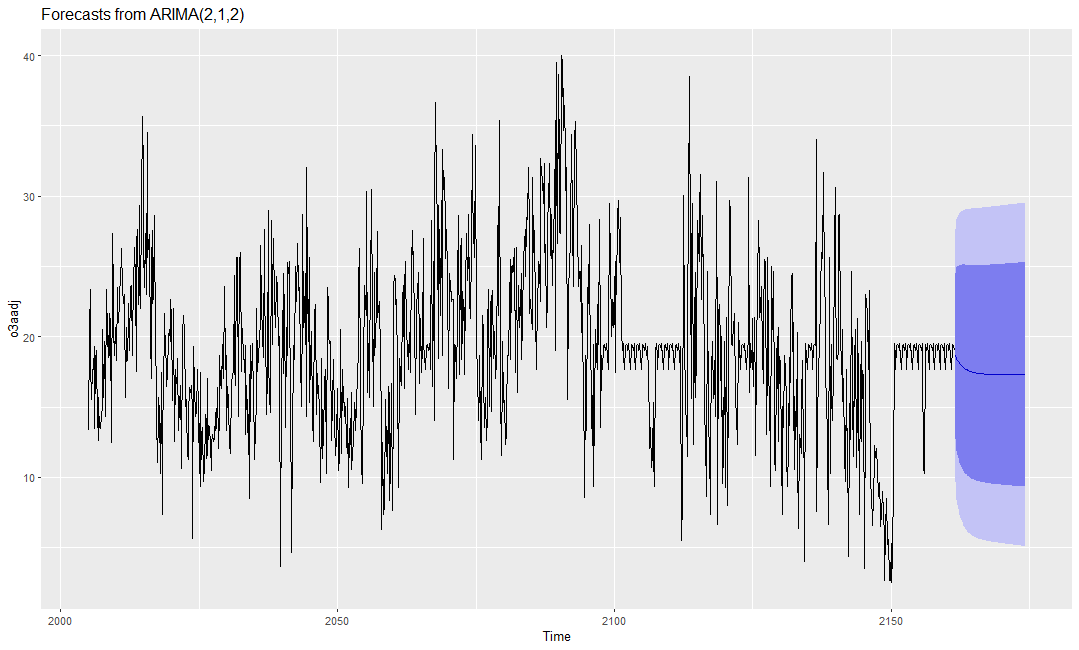


**Paso 7: Proposticar el modelo elegido**

autoplot(forecast(fit3, h=90))

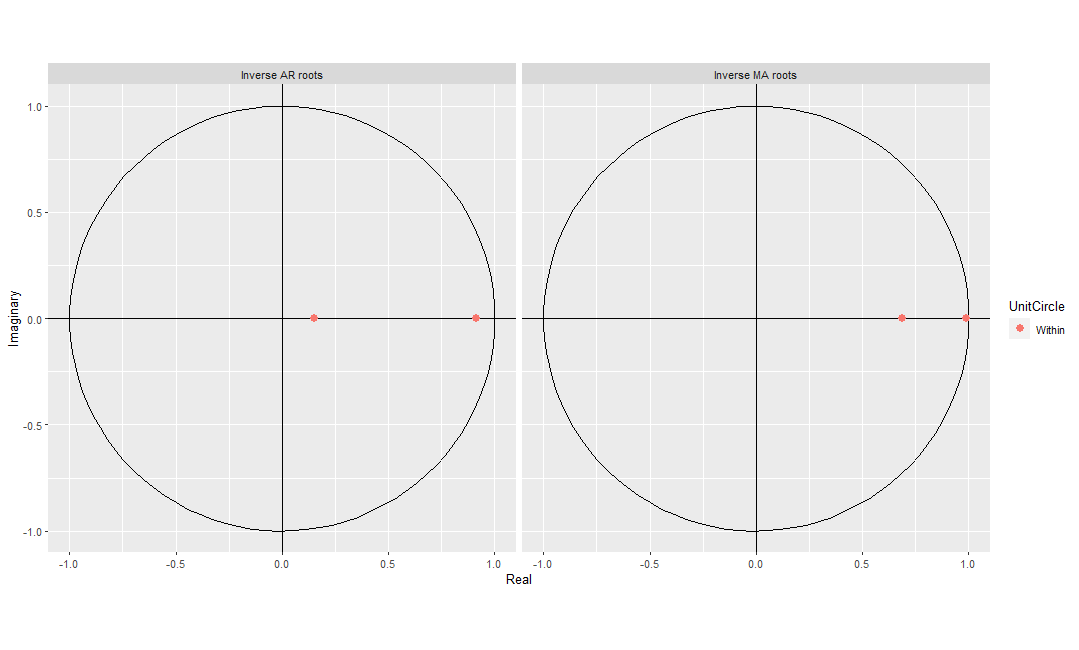
**Figura 14**

*Pronóstico del Ozono BA diario ajustado estacionalmente 1/1/2021 - 30/3/2021*



**Figura 15**

*Raíces características inversas para el modelo ARIMA(2,1,2) ajustado a los datos del ozono BA desestacionalizado.*



Los dos puntos rojos en la gráfica de la izquierda corresponden a las raíces de los polinomios Φ(B), mientras que los puntos rojos en la gráfica de la derecha corresponden a la raíz de θ(B). Todos están dentro del círculo unitario, como es de esperar, porque R asegura que el modelo ajustado sea tanto estacionario como invertible.

7. En la Ciudad de Lima se ha hecho trazabilidad de los casos de cáncer (Datos *Cases.csv*); es de suma importancia conocer y explorar las futuras tasas, con la finalidad de proyectar las ayudas y cobertura del sistema de salud ante los nuevos casos. Proponga un modelo o forma de pronosticar los casos en donde el margen de error sea estadísticamente razonable, por otro lado ¿es importante para el municipio contar con una herramienta que permita actualizar los datos y seguir proyectando las tasas. En caso de detectar valores que no corresponden a la realidad debe tomar una decisión acorde a mejorar los modelos propuestos.

**Paso 0: Limpieza de datos y gráfico inicial**

##Paso 0: Carga de la data y limpieza

df\_cancer <- read.csv(

file = "Cases.csv",

header = TRUE,

sep = ",",

dec = ".",

stringsAsFactors = TRUE

)

# Detectar y contar los valores faltantes

which(is.na(df\_cancer$Freq))

sum(is.na(df\_cancer$Freq))

#Crear un objeto de la clase ts: serie de tiempo.

help(ts)

cancer0 <- ts(df\_cancer[,3], frequency = 1, start = 1958)

#Observamos la data del 2018 muy por debajo de la realidad. Lo quitamos del análisis

df2\_cancer <- subset(df\_cancer,Year != 2018 )

cancer <- ts(df2\_cancer[,3], frequency = 1, start = 1958)

par(mfrow = c(1,2))

plot(cancer0, main="Serie de cáncer en Lima")

plot(cancer, main="Serie de cáncer en Lima, sin el 2018")

#b. Dividimos el conjunto de datos en un conjunto de entrenamiento

#y un conjunto de prueba, donde el conjunto de prueba son los

#últimos 12 años de datos.

cancer\_train <- window(cancer,start=c(1958),end=c(2005))

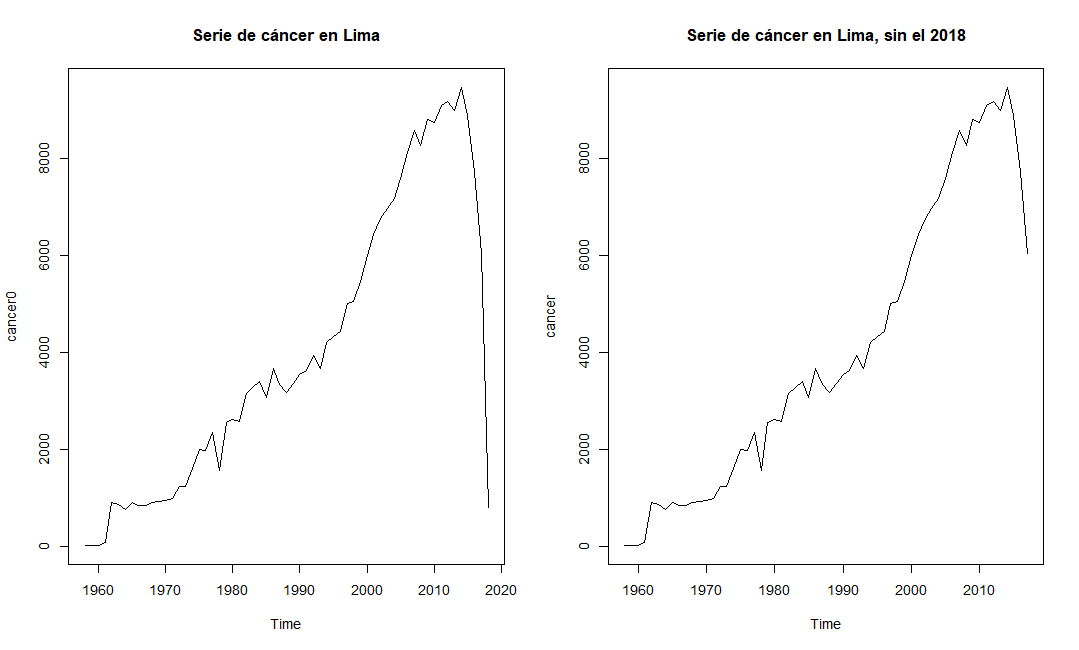
cancer\_test <- window(cancer,start=c(1958))

**Paso 1 : Analizamos la data de entrenamiento**

Observamos que la data del año 2018 está muy por debajo de la realidad. La quitamos del análisis. Además, dividimos el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba, donde el conjunto de prueba son los últimos 12 años de datos.

**Figura 16**

*Número de personas con cáncer en Lima*



**Paso 2 ¿Transformación Box Cox para estabilizar la varianza?**

# Paso 2: Transformación Box Cox para estabilizar la varianza

(lambda <- BoxCox.lambda(cancer\_train))

cancer\_boxcox <- BoxCox(cancer\_train,lambda)

autoplot(cancer\_boxcox)

#lambda = 1.259332

par(mfrow = c(1,2))

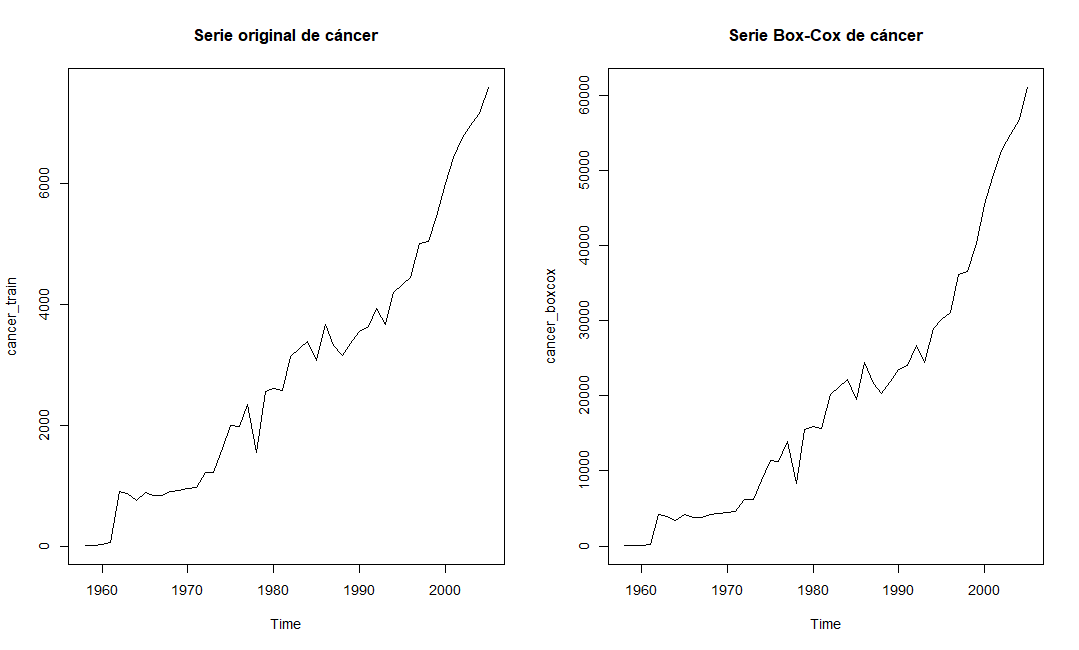
plot(cancer\_train, main="Serie original de cáncer")

plot(cancer\_boxcox, main="Serie Box-Cox de cáncer")

No utilizamos la transformación Box-Cox porque no simplifica los patrones de los datos históricos y no estabiliza la varianza.

**Figura 16**

*Serie original y transformada Box-Cox de los datos cáncer en Lima*



**Paso 3 ¿cuántas diferencias requiere la serie transformada de cáncer?**

#Paso 3

#¿cuántas diferencias requiere la serie transformada de cáncer?

cancer\_train %>% ur.kpss() %>% summary()

cancer\_train %>% diff() %>% ur.kpss() %>% summary()

cancer\_train %>% diff() %>% diff() %>% ur.kpss() %>% summary()

#Según esta prueba se debe hacer dos diferencias para que la serie sea estacionaria

#Secuencia de KPSS

ndiffs(cancer\_train)

#Según la secuencia del test KPSS se requiere una diferencia para que la serie

#transformada box cox de cáncer se vuelva estacionaria.

# Test de Dickey Fuller aumentado (prueba de raíz unitaria)

adf.test(cancer\_train)

adf.test(diff(cancer\_train))

adf.test(diff(diff(cancer\_train)))

#Según el test Dickey Fuller se requiere dos diferencias para que la serie

#transformada se vuelva estacionaria.

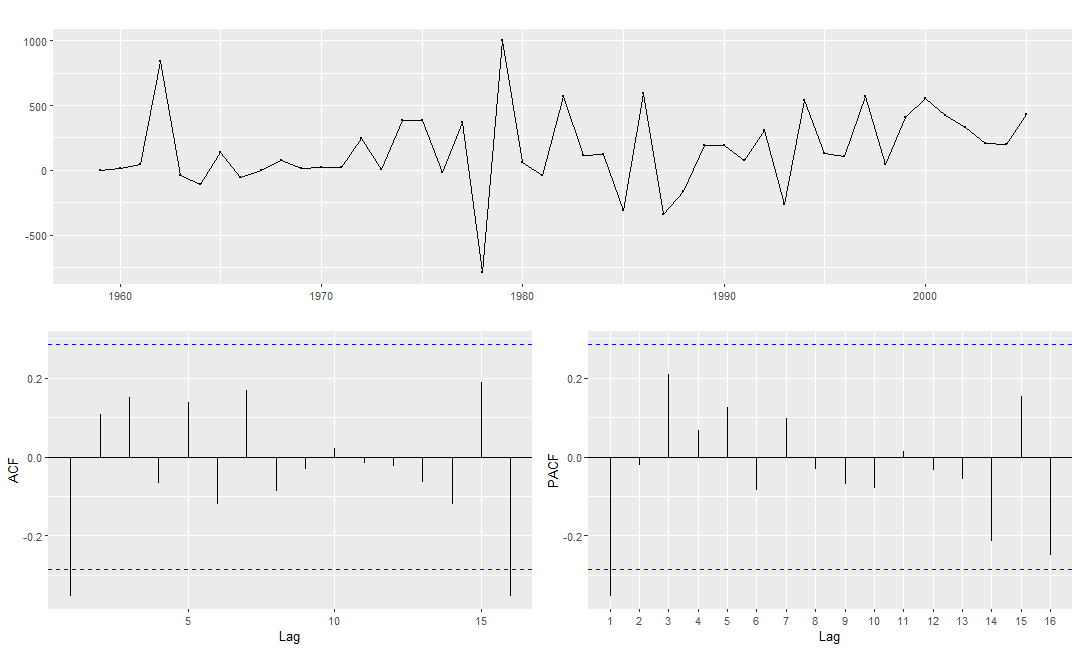
#Visualmente

cancer\_train %>% diff() %>% ggtsdisplay(main="")

Al observar la figura 17 sólo se necesita una diferencia para que la serie se vuelva estacionaria tal como nos sugiere la función *ndiffs.*

**Figura 17**

*ACF y PACF de la primera diferencia de los datos de cáncer*



**Paso 4: Observar el ACF y PACF para proponer modelos ARIMA. Además usaremos el *auto.arima***

(fit\_cancer <- auto.arima(cancer\_train,

stepwise=FALSE,

approximation=FALSE,

seasonal=FALSE))

Al observar la figura 17, las gráficas del ACF y PACF proponen el ARIMA(1,1,0) y ARIMA(1,1,1). Por otro lado, el camino corto del *auto.arima* propone el ARIMA(1,1,0)

**Paso 5: Ajustamos los modelos propuestos y elegimos el que tiene el menor valor AICc**

(fit1 <- Arima(cancer\_train , order=c(1,1,1)))

(fit2 <- Arima(cancer\_train , order=c(1,1,0)))

Elegimos el ARIMA(1,1,0) para hacer el pronóstico por tener menor valor en el criterio de información Akaike corregido AICc.

**Paso 6: Test de ruido blanco para los residuos del mejor modelo**

cancer\_fit <- forecast(fit2,h=17)

checkresiduals(cancer\_fit)

checkresiduals(fit2)

# Ljung-Box test

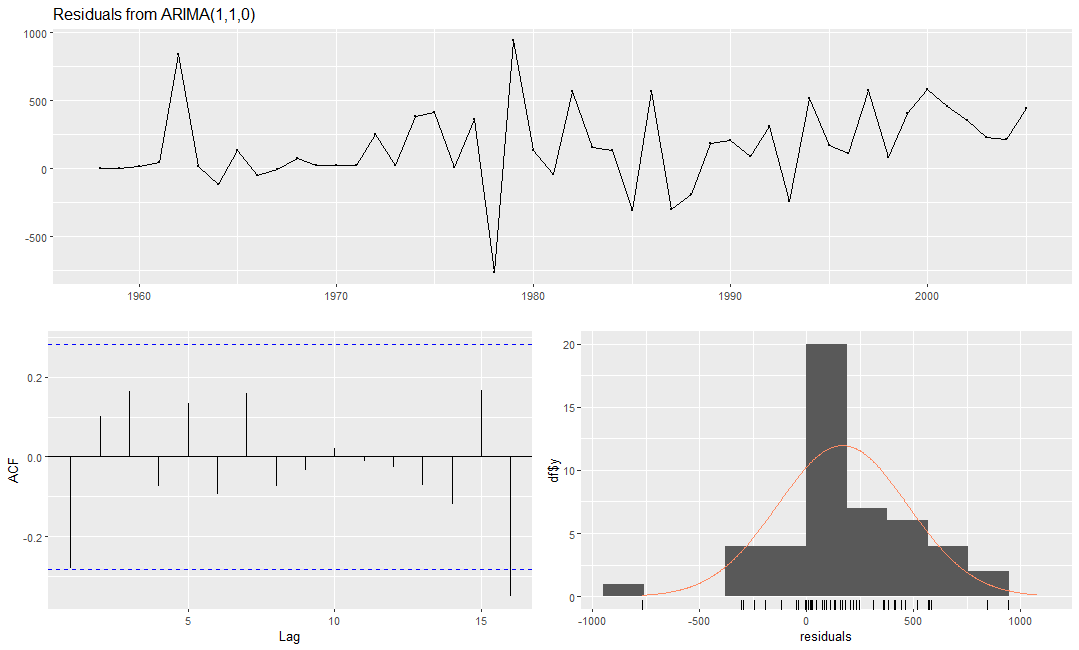
# data: Residuals from naive method

Q\* = 9.7357, df = 8, p-value < 0.3723

# Model df: 1. Total lags used: 10

**Figura 18**

*Prueba gráfica y test de ljung-box para los casos de cáncer estimados*

****

Al observar el correlograma ACF y el test de Ljung se acepta la hipótesis nula de ruido blanco de los errores de los pronósticos.

**Paso 7: Pronosticar el modelo elegido**

autoplot(cancer\_fit) +

autolayer(cancer\_fit, series = "Pronóstico ARIMA(1,1,0)") +

autolayer(cancer\_test, series = "casos de cáncer") +

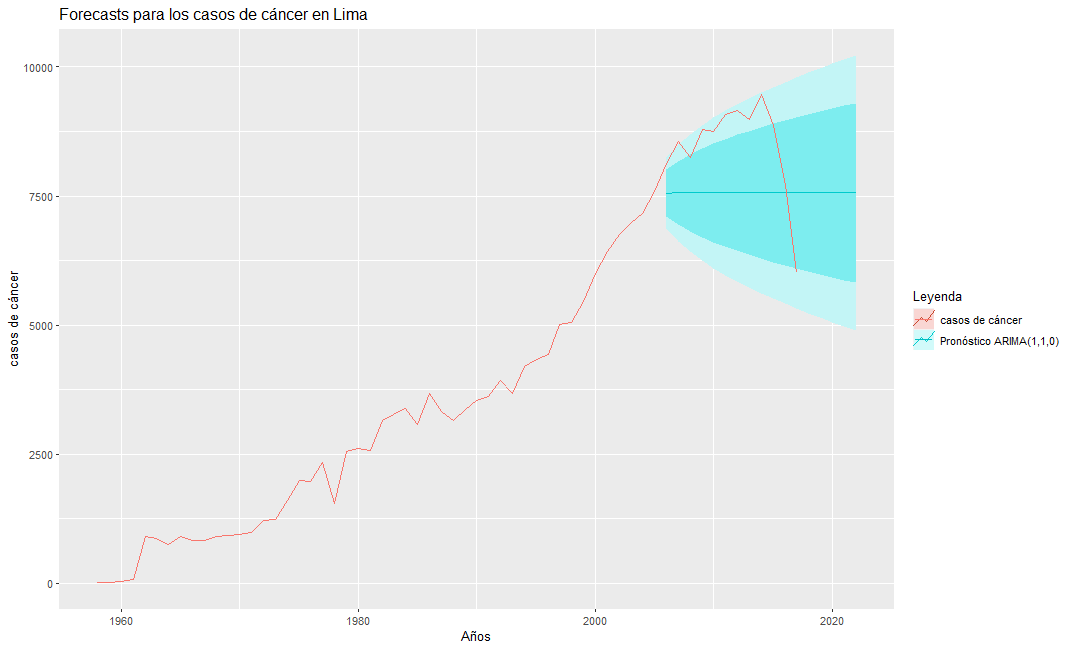
xlab("Años") + ylab("casos de cáncer") +

ggtitle("Forecasts para los casos de cáncer en Lima") +

guides(colour=guide\_legend(title="Leyenda"))

**Figura 19**

*Pronóstico de casos de cáncer en Lima por el método ARIMA(1,1,0)*

****

8.En la ciudad de Lima una empresa del sector energético desea modelar y pronosticar la demanda de energía en KW/h (Datos *MERCADO\_REGULADO.csv*) total a 15 días, con la finalidad de planear la compra total a los proveedores internacionales. Plantee una forma eficiente de realizar dichos pronósticos de tal forma que en la empresa se puede actualizar y estimar cada vez que se cuenta con nueva información.

#### Método de Holt-Winters con datos diarios ######

##Paso 0: Carga de la data y limpieza

df\_mercado <- read.csv(

file = "MERCADO\_REGULADO.csv",

header = TRUE,

sep = ";",

dec = ".",

stringsAsFactors = TRUE

)

# Detectar y contar los valores faltantes

which(is.na(df\_mercado$ENERGIA))

sum(is.na(df\_mercado$ENERGIA))

mercado0 <- ts(df\_mercado[,3], frequency = 7 , start =c(2019,60))

autoplot(mercado0)

fc <- hw(subset(mercado0,end=length(mercado0)-35),

damped = TRUE, seasonal="multiplicative", h=50)

autoplot(fc, PI=FALSE) +

autolayer(fc, series = "HW multi damped", PI=FALSE) +

autolayer(mercado0, series = "consumo de energía KW/h") +

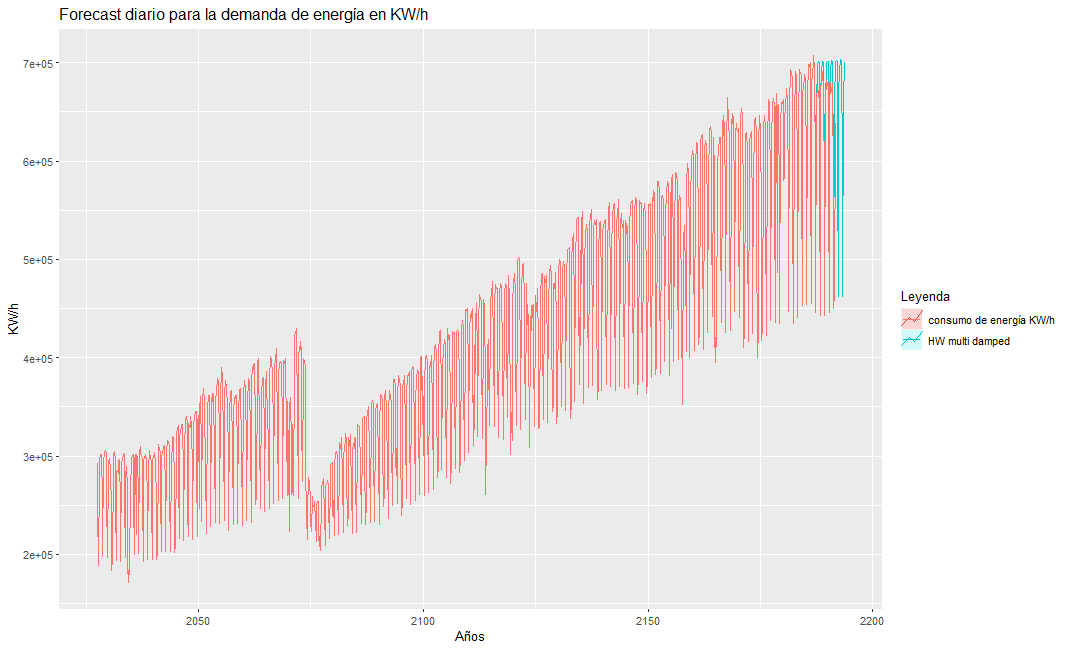
xlab("Años") + ylab("KW/h") +

ggtitle("Forecast diario para la demanda de energía en KW/h") +

guides(colour=guide\_legend(title="Leyenda"))

**Figura 20**

*Pronóstico diario para la demanda de energía en KW/h*



**Tabla 3**

*Pronóstico del consumo de energía 15 días después*

|  |  |  |  |
| --- | --- | --- | --- |
| **Pronóstico 15 días** | **Point Forescast** | **Lo 95** | **Hi 95** |
| 1 | 699 877,9 | 555 197,7 | 844 558,1 |
| 2 | 702 467,4 | 556 711,4 | 848 223,3 |
| 3 | 701 667,7 | 555 539,4 | 847 795,9 |
| 4 | 596 278,3 | 471 642,1 | 720 914,4 |
| 5 | 462 501,7 | 365 475,3 | 559 528,1 |
| 6 | 665 288,8 | 525 214,5 | 805 363,1 |
| 7 | 695 541,5 | 548 570,2 | 842 512,9 |
| 8 | 700 258,6 | 546 039,8 | 854 477,5 |
| 9 | 702 848,3 | 547 547,8 | 858 148,9 |
| 10 | 702 047,0 | 546 414,0 | 857 680,1 |
| 11 | 596 599,7 | 463 910,9 | 729 288,6 |
| 12 | 462 750,3 | 359 496,6 | 566 004,0 |
| 13 | 665 645,4 | 516 640,2 | 814 650,6 |
| 14 | 695 913,3 | 539 632,7 | 852 193,8 |
| 15 | 700 632,4 | 537 368,3 | 863 896,5 |

**Referencias**

Hyndman, y Athanasopoulos. *Forecasting: Principles and Practice (2nd ed)*. Accedido 21 de diciembre de 2022. <https://otexts.com/fpp2/>.