



Tarea 3

Modelación ARIMA

Modelos de series de Tiempo y Supervivencia

Profesor: Naranjo Albarrán Lizbeth

Adjuntos: Reyes González Belén

Rivas Godoy Yadira

Integrantes: Cuéllar Chávez Eduardo de Jesús

García Tapia Jesús Eduardo

Miranda Meraz Areli Gissell

Ramírez Maciel José Antonio

Saldaña Morales Ricardo

Grupo: 9249

Fecha: 10/NOV/2021

Analizar los datos Quarterly U.S. new plant/equip. expenditures 64 76 billions de la librería tsdl de R.

Análisis descriptivo.

Grafique los datos, describa lo que observe (varianza constante o no constante, descomposición clásica, tendencia, ciclos estacionales, periodicidad de los ciclos).

Primero carguemos la librería tsdl, ya que los datos necesarios se encuentran en dicha librería; así como otras necesarias para esta tarea.

```
library(tsd1);library(ggplot2);library(itsmr);library(forecast);library(TSA);library(lmtest)
library(timeSeries);library(timeSeries);library(astsa);
library(tseries);library(forecast);library(nortest);library(dplyr);library(imputeTS)
```

Cargamos la base de datos, nos aseguramos de que es la deseada.

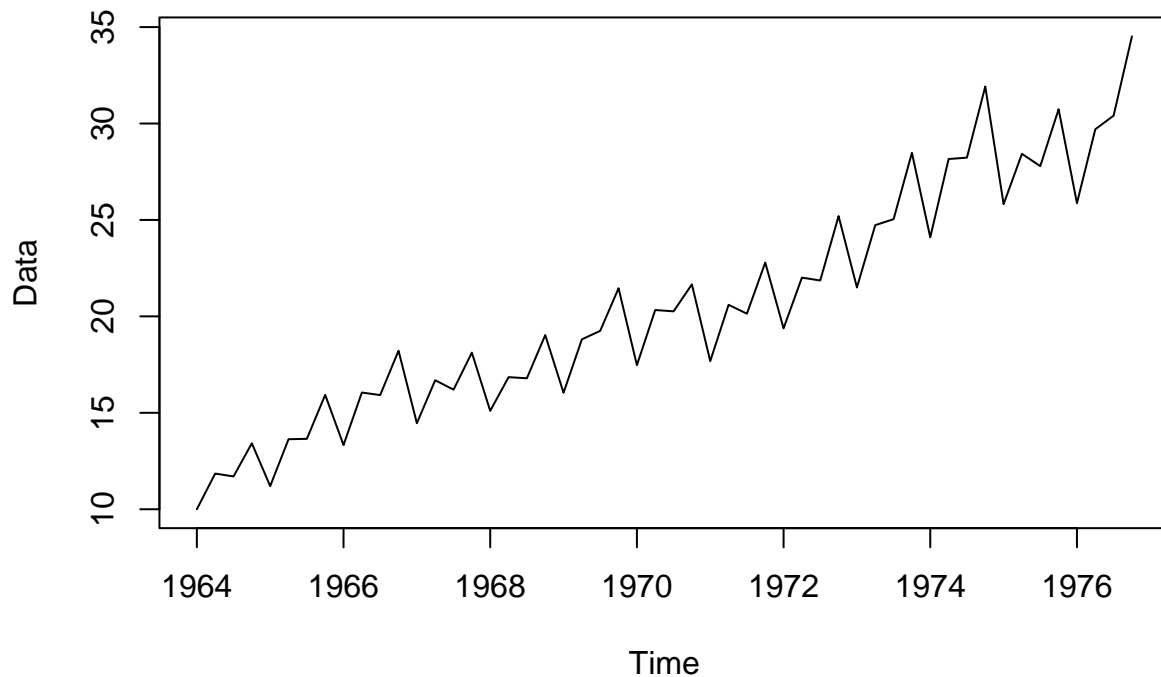
```
Data <- tsdl[[12]]
attributes(Data)

## $tsp
## [1] 1964.00 1976.75    4.00
##
## $class
## [1] "ts"
##
## $source
## [1] "Abraham & Ledolter (1983)"
##
## $description
## [1] "Quarterly U.S. new plant/equip. expenditures -64 - -76 billions"
##
## $subject
## [1] "Microeconomic"
```

Ahora que noa aseguramos de que es la base que queríamos, procedemos a graficar:

```
plot(Data, main = "Quarterly U.S. new plant/equip. expenditures \n 64 76 billions")
```

Quarterly U.S. new plant/equip. expenditures 64 76 billions



Varianza Al menos de manera gráfica, la intuición nos dice que no hay varianza constante, pero probémoslo con un test de homocedasticidad:

```
#Los pasamos a series de tiempo
Serie<-ts(data=Data,start=c(1964,01),end=c(1976,4),frequency=4)
tiempo<-seq(1964+0/4, 1976+3/4, by = 1/4)
bptest(Serie~tiempo)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  Serie ~ tiempo
## BP = 5.3713, df = 1, p-value = 0.02047
```

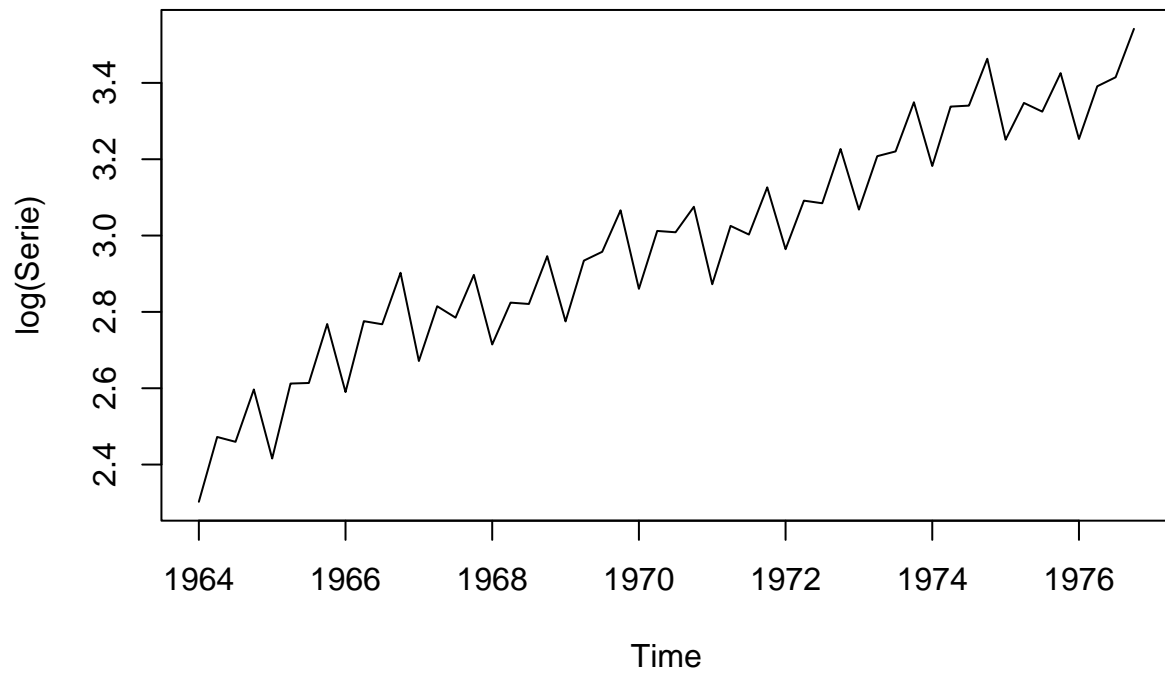
Efectivamente, no pasa el test de homocedasticidad.

Veamos qué pasa si aplicamos la transformación logaritmo:

```
bptest(log(Serie)~tiempo)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  log(Serie) ~ tiempo
## BP = 1.965, df = 1, p-value = 0.161
```

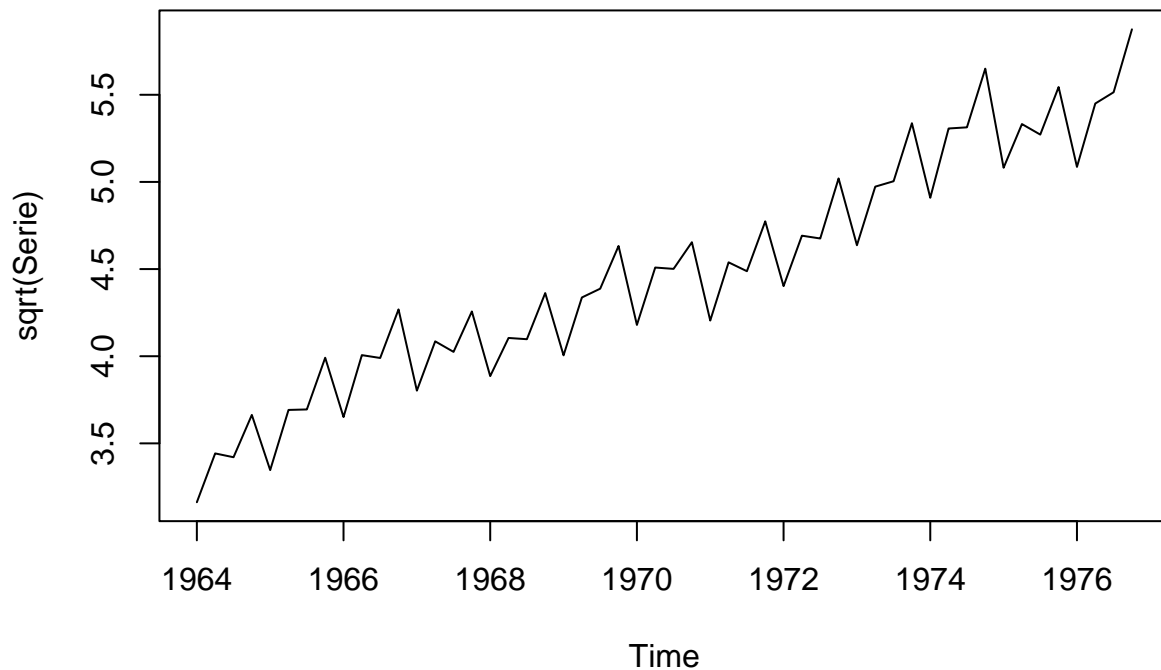
```
plot(log(Serie))
```



Tampoco ayudó, aunque mejoró un poco. Intentemos con la raíz cuadrada

```
bptest(sqrt(Serie)~tiempo)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: sqrt(Serie) ~ tiempo  
## BP = 0.8651, df = 1, p-value = 0.3523  
plot(sqrt(Serie))
```



¡Logramos estabilizarla!

Veamos si es estacionaria.

```
adf.test(sqrt(Serie))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: sqrt(Serie)
## Dickey-Fuller = -1.415, Lag order = 3, p-value = 0.8097
## alternative hypothesis: stationary
```

```
kpss.test(sqrt(Serie))
```

```
## Warning in kpss.test(sqrt(Serie)): p-value smaller than printed p-value
```

```
##
## KPSS Test for Level Stationarity
##
## data: sqrt(Serie)
## KPSS Level = 1.389, Truncation lag parameter = 3, p-value = 0.01
```

Tendencia

Podemos observar una tendencia creciente que se presenta de manera lineal (al parecer), de manera general a lo largo de la serie

Ciclos estacionales

Los ciclos están bastante marcados , y tiene sentido puesto que son los datos de gastos de una empresa en maquinaria por trimestre, y en ese contexto es lógico que se presenten ciclos: Generalmente decae del último trimestre del año anterior al primero del año siguiente, para después crecer en el segundo semestre, en el tercero se mantiene casi al mismo nivel que el segundo, pero en el último aumenta; y es un comportamiento que se repite año con año

Periodicidad de los ciclos

Como comentamos en el apartado anterior, parece (al menos de manera gráfica) que se tienen ciclos anuales.

Descomposición clásica

Descomponeremos la serie por medio de filtros lineales:

Estabilización de la varianza Aplicamos la transformación raíz cuadrada

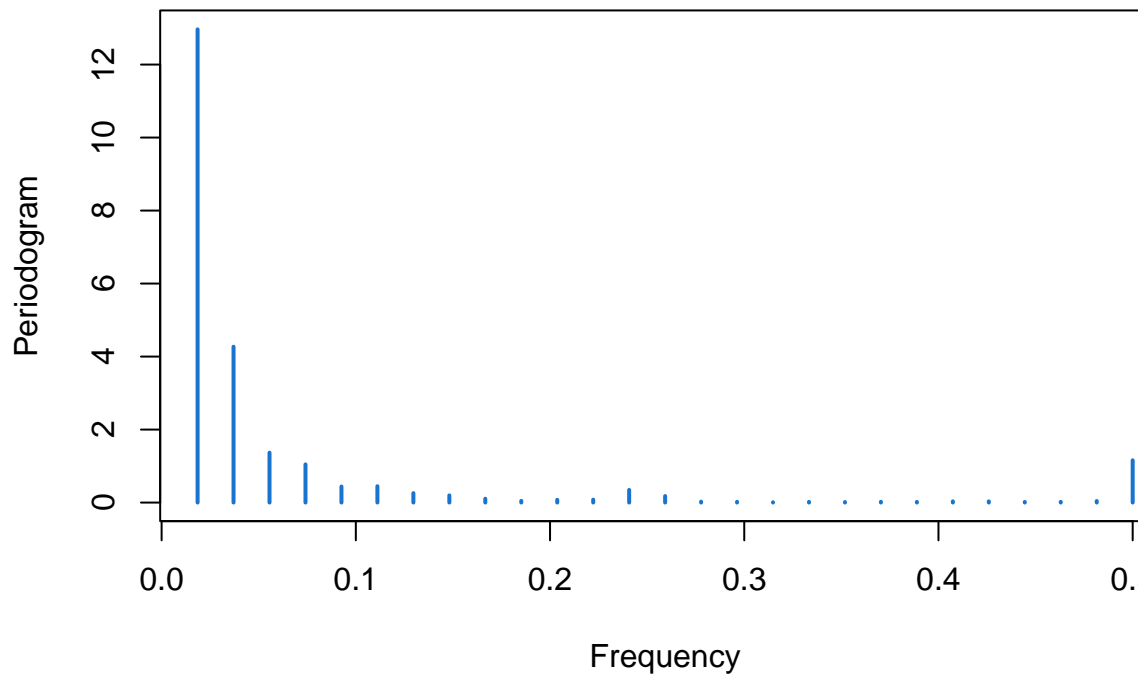
```
Serie_sq<-sqrt(Serie)
bptest(Serie_sq~tiempo)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  Serie_sq ~ tiempo
## BP = 0.8651, df = 1, p-value = 0.3523
```

Podemos asumir varianza constante

```
#Veamos la tendencia y los ciclos
Xt = Serie_sq
p = periodogram(Xt, main="Periodograma", col=4) # Obtenemos el periodograma
```

Periodograma



Periodicidad de ciclos

```
names(p)
```

```
## [1] "freq"      "spec"      "coh"      "phase"     "kernel"    "df"
## [7] "bandwidth" "n.used"    "orig.n"   "series"    "snames"    "method"
## [13] "taper"     "pad"       "detrend"  "demean"
```

```
# Ordenamos de mayor a menor las estimaciones del periodograma.
```

```
spec = sort(p$spec, decreasing = TRUE)
```

```
(spec = spec[1:10]) # Nos quedamos con los 8 coeficientes de mayor frecuencia.
```

```
## [1] 12.9646281 4.2701894 1.3676068 1.1581310 1.0460373 0.4486585
```

```
## [7] 0.4395386 0.3466661 0.2593484 0.1970454
```

```
i = match(spec, p$spec) # Buscamos sus indices en el periodograma.
```

```
d = p$freq # Vemos las frecuencias del periodograma.
```

```
d = d[i] # Nos quedamos con las frecuencias que nos interesan.
```

```
cbind(spec,d,i)#
```

```
##          spec          d  i
## [1,] 12.9646281 0.01851852 1
## [2,] 4.2701894 0.03703704 2
## [3,] 1.3676068 0.05555556 3
## [4,] 1.1581310 0.50000000 27
## [5,] 1.0460373 0.07407407 4
## [6,] 0.4486585 0.11111111 6
## [7,] 0.4395386 0.09259259 5
## [8,] 0.3466661 0.24074074 13
```

```
## [9,] 0.2593484 0.12962963 7
## [10,] 0.1970454 0.14814815 8

d = 1 / d # Obtenemos los parametros para utilizar en promedios moviles.
d = floor(d) #
(d = sort(d))

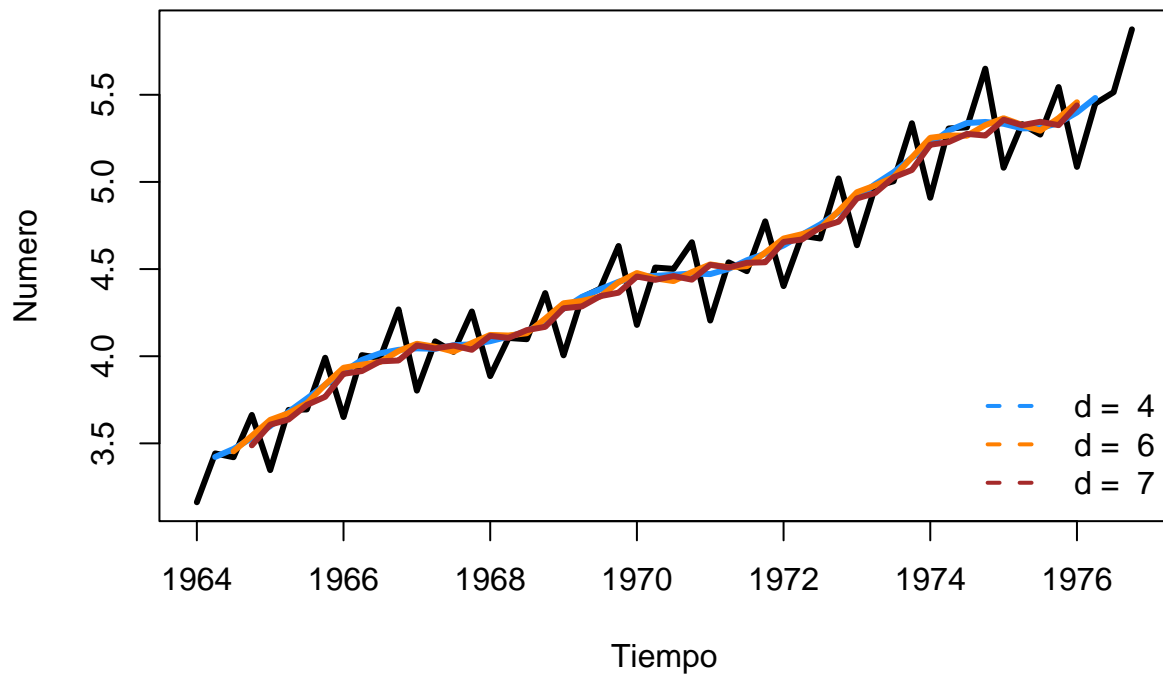
## [1] 2 4 6 7 9 10 13 18 27 54

# Quitamos los periodos mas grandes
d = d[-length(d)]
d = d[-length(d)]
# Quitamos el más pequeño
d = d[-1]
d #Posibles periodos del ciclo

## [1] 4 6 7 9 10 13 18

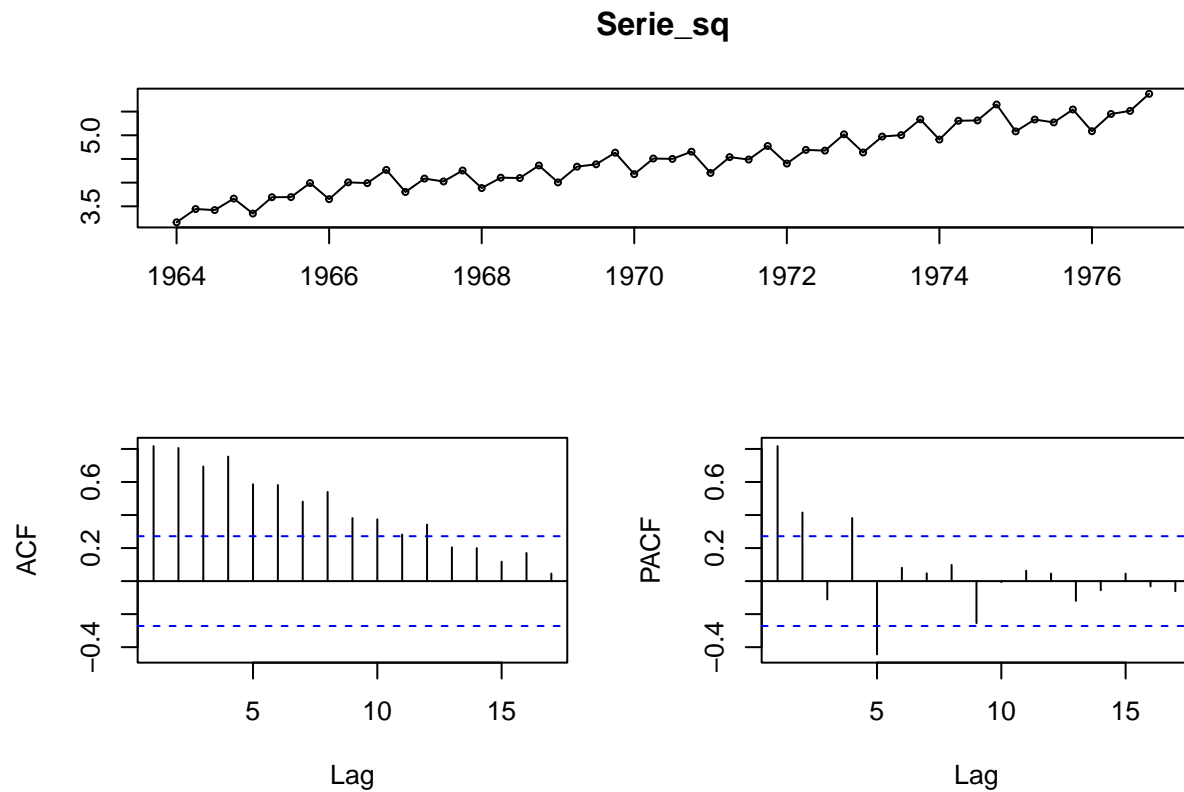
#Realizamos la grafica:
col = c("dodgerblue1", "darkorange1", "brown")
plot(Serie_sq, lwd = 3, xlab = "Tiempo", col = "gray0",
     main = "Serie con varianza Homocedastica",
     ylab = "Numero", col.main = "burlywood")
for (i in 1:3) {
  lines(tiempo, stats::filter(Serie_sq, rep(1 / d[i], d[i])), col = col[i],
    lwd = 3)
}
legend("bottomright", col = col, lty = 2, lwd = 2, bty = "n",
     legend = c(paste("d = ", d[1]), paste("d = ", d[2]),
     paste("d = ", d[3])), cex = 1)
```


Serie con varianza Homocedastica



Notemos que $d = 2$ parece sobreajustar un poco nuestra gráfica, de hecho, bastante. Sin embargo, con $d = 4$ podemos obtener un buen suavizamiento sin pagar el costo de otros 2 datos al elegir $d = 6$. Veamos el ACF y PACF:

```
tsdisplay(Serie_sq)
```

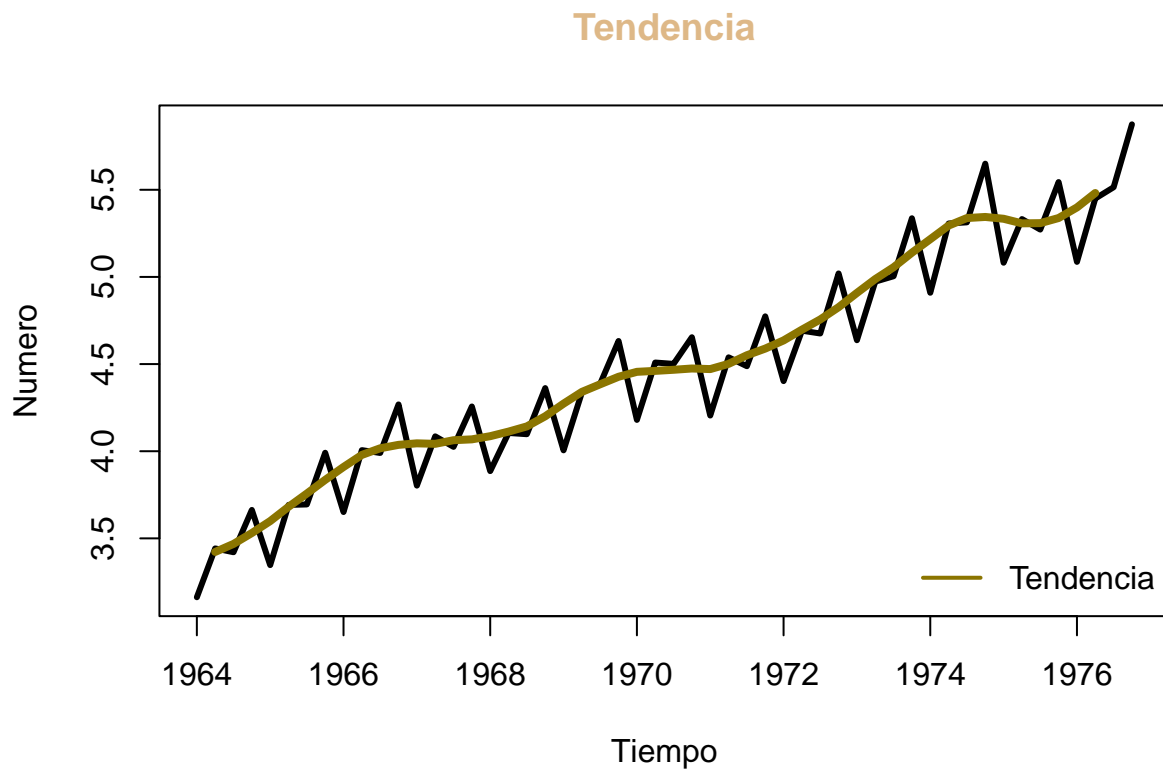


Y, junto con este último resultado, nos parece ideal concluir que el ciclo es $d = 4$

Tendencia

Ahora, aislemos la tendencia:

```
tendencia = stats::filter(Serie_sq, rep(1/4, 4))
plot(Serie_sq, lwd = 3, xlab = "Tiempo", col = "black",
     main = "Tendencia",
     ylab = "Numero", col.main = "burlywood")
lines(tendencia, col = "gold4", lwd = 4)
legend("bottomright", col = "gold4", lty = 1, lwd = 2, bty = "n",
     legend = "Tendencia", cex = 1)
```

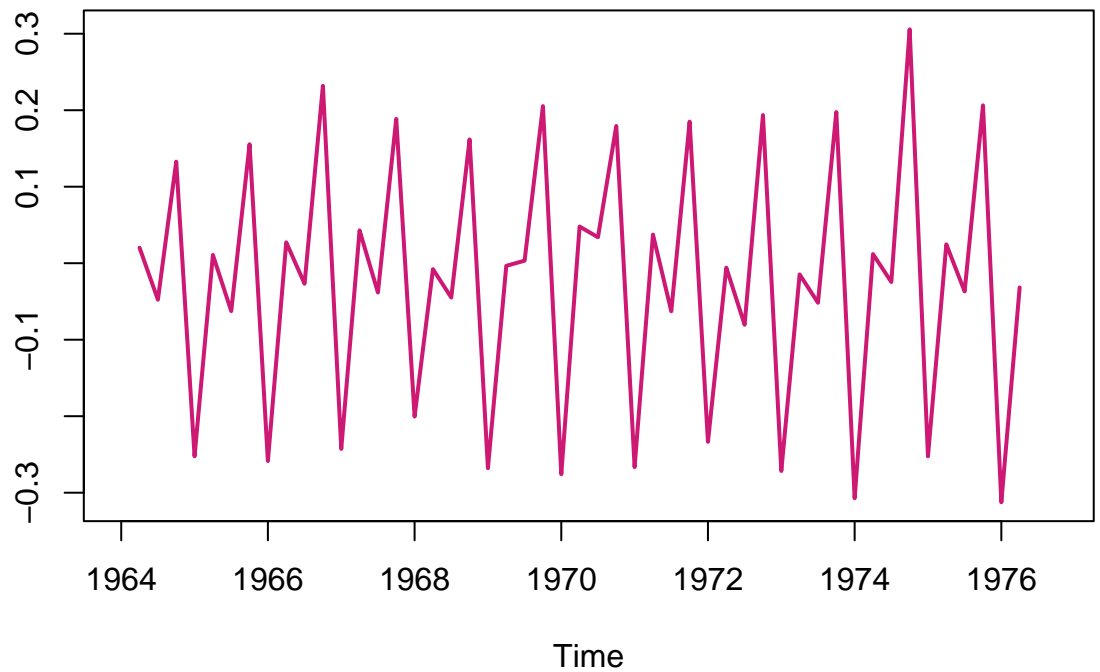


Lo que refuerza lo que creíamos: Tiene tendencia creciente casi de manera general.

```
# Quitamos la tendencia
# Solo trabajamos con la serie cuya varianza es cte.

datosSinTendencia = Serie_sq - tendencia # Serie sin tendencia
plot(datosSinTendencia, main="Serie sin tendencia", lwd=2, ylab="", col=14)
```

Serie sin tendencia



Tendencia

Convertimos datosSinTendencia en objeto TS, dado que hicimos promedios moviles

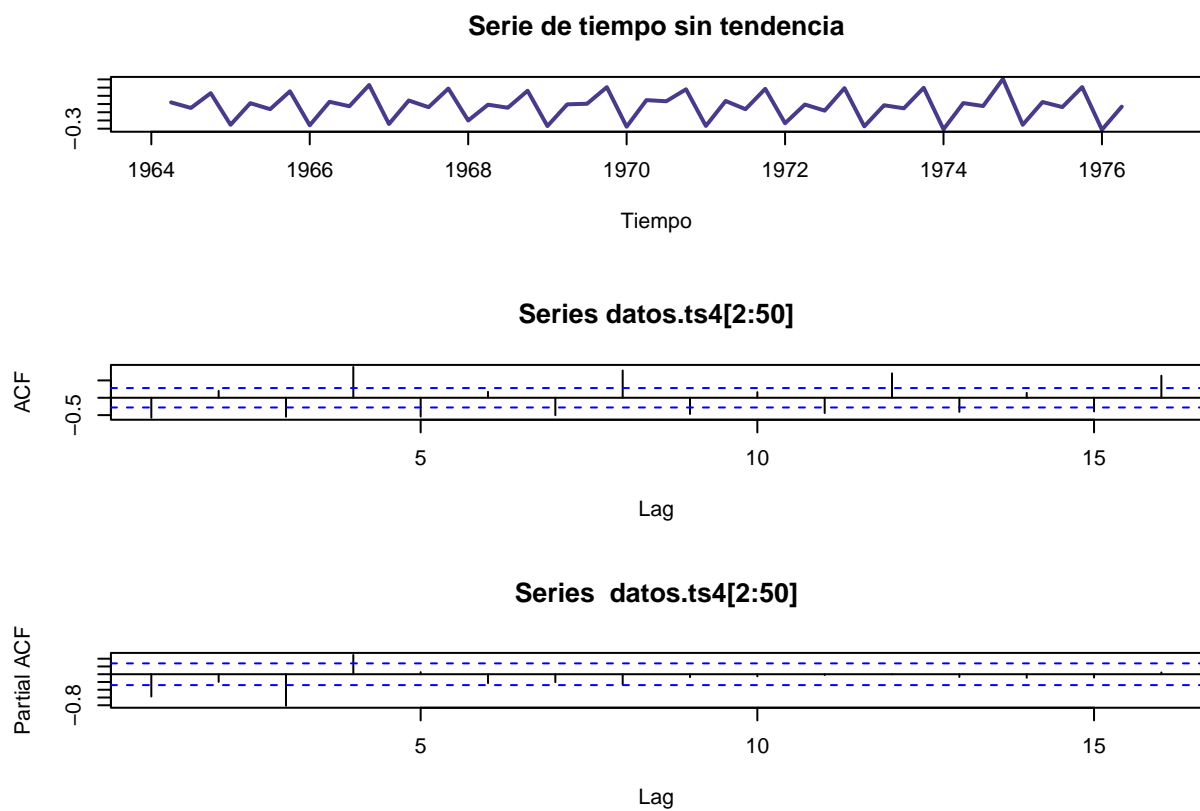
```
inicio=start(Serie_sq)
final=end(Serie_sq)
```

```
datos.ts4=ts(datosSinTendencia, frequency = 4, start=inicio,end=final)
which(is.na(datos.ts4)==T)
```

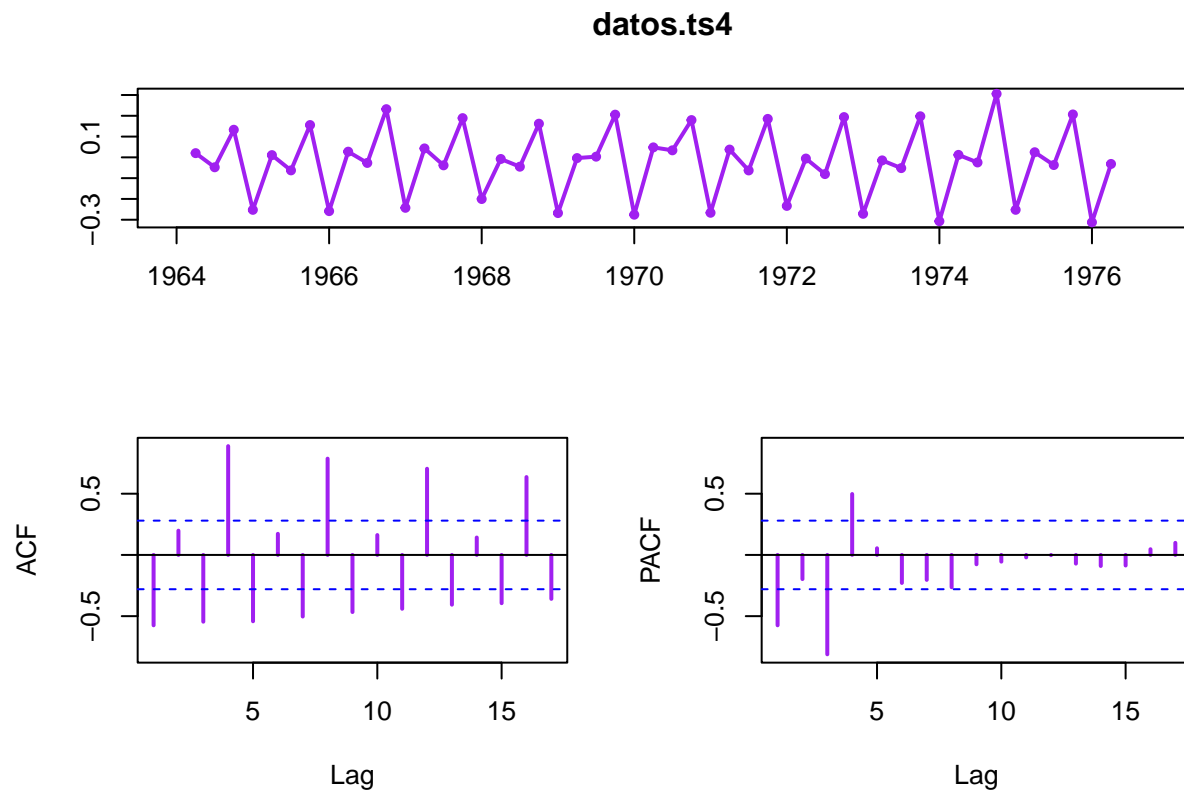
```
## [1] 1 51 52
```

```
par(mfrow = c(3,1))
plot(datos.ts4, col = "slateblue4", lwd = 2, ylab = " ", type = "l",
     main = "Serie de tiempo sin tendencia", xlab = "Tiempo")
```

```
acf(datos.ts4[2:50])
pacf(datos.ts4[2:50])
```



```
par(mfrow = c(1,1))
tsdisplay(datos.ts4, col="purple", lwd=2)
```



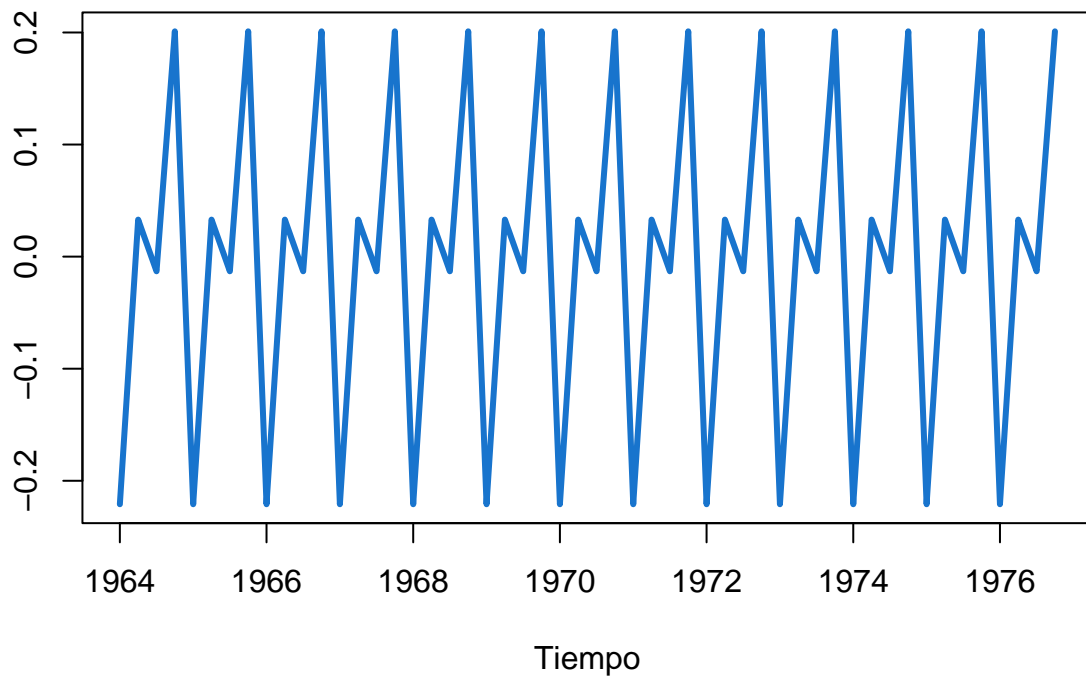
Parece que eliminamos la tendencia, ahora tratemos de verificar los ciclos

#Ciclos o parte estacional Ahora, estimaremos la parte estacional. Tenemos que $d = 4$. Originalmente contábamos con 52 datos, pero ahora tenemos 48 (por los NA), entonces $\frac{48}{4} = 12$ ciclos.

```
# Creamos un ciclo promedio que estime la parte estacional,
# usando la serie sin tendencia.
d = 4
k = length(datos.ts4) / d # Numero de ciclos de la serie sin tendencia
w = rep(0, 4)
# Para el resto de los trimestres
for (i in 1:4)
  w[i] = sum(datos.ts4[d * (0:(k-1)) + i], na.rm = TRUE) / k

# Ahora, ajustamos el ciclo obtenido
ciclo = w - mean(w)
ciclo = ts(rep(ciclo, times = k), start = start(Serie_sq),
           frequency = frequency(Serie_sq))
par(mfrow = c(1, 1))
plot(ciclo, col = 20, lwd = 3, ylab = " ", xlab = "Tiempo",
     main = "Ciclos de la serie")# Es el ciclo de la serie
```

Ciclos de la serie



```
# Ciclos anuales
```

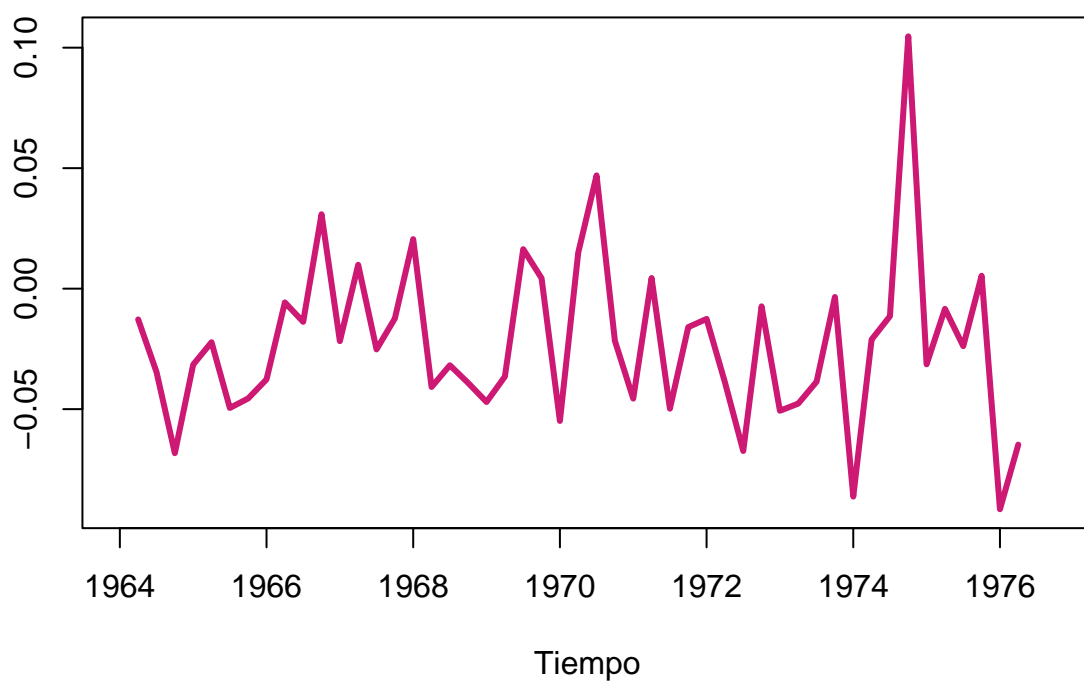
Ahora verifiquemos de manera gráfica

```
# Calculamos la parte aleatoria
```

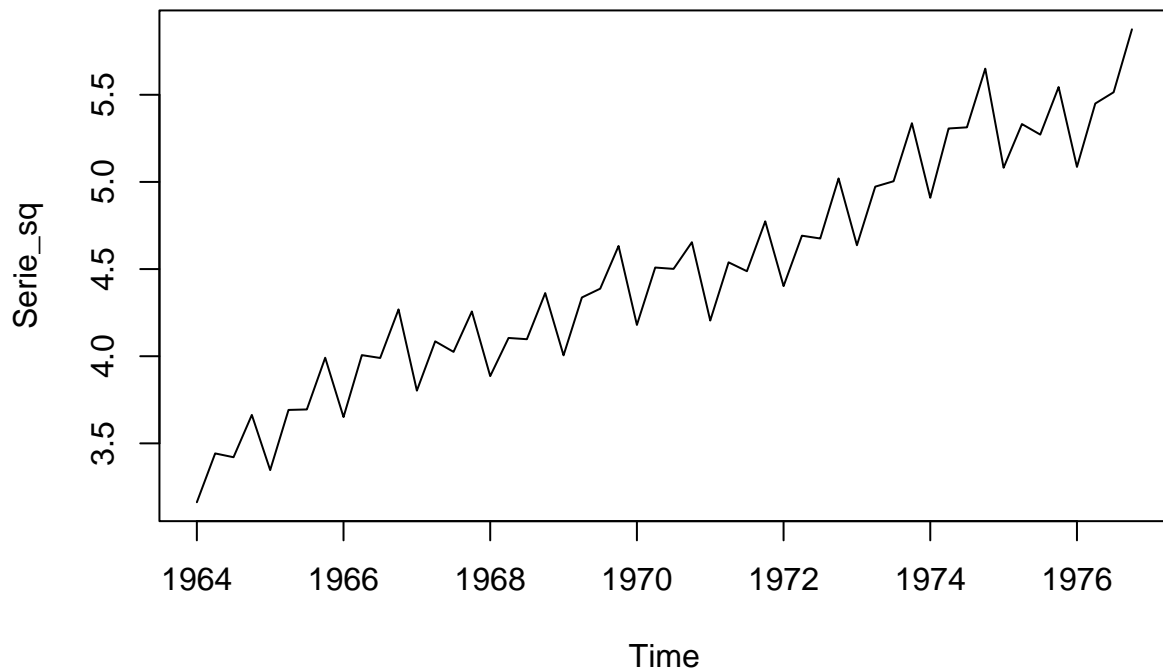
```
parte_aleatoria = datos.ts4 - ciclo
```

```
plot(parte_aleatoria, main = "Parte aleatoria",  
     col = 30, lwd = 3, xlab = "Tiempo", ylab = "")
```

Parte aleatoria



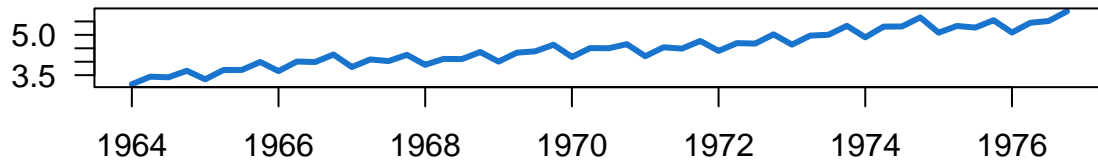
```
plot(Serie_sq)
```

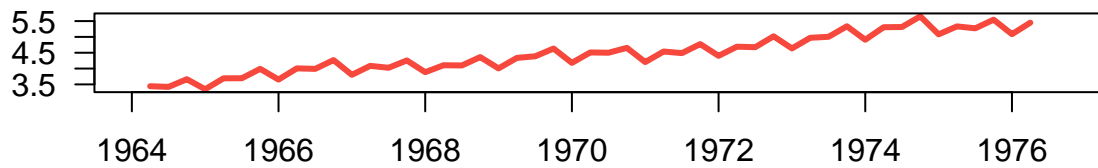
Con esto, ya tenemos nuestras series

```
componentes = tendencia + ciclo+parte_aleatoria
componentes = ts(componentes, start = start(Serie_sq), frequency = 4)
par(mfrow = c(2,1))
plot(Serie_sq, col=28, las=1, main="Serie con varianza constante", lwd=3, xlab="",ylab="")
plot(componentes, col = 18, lwd = 3, las=1, main="Yt=tendencia+ciclos+aleatoria", xlab="",ylab="")
```

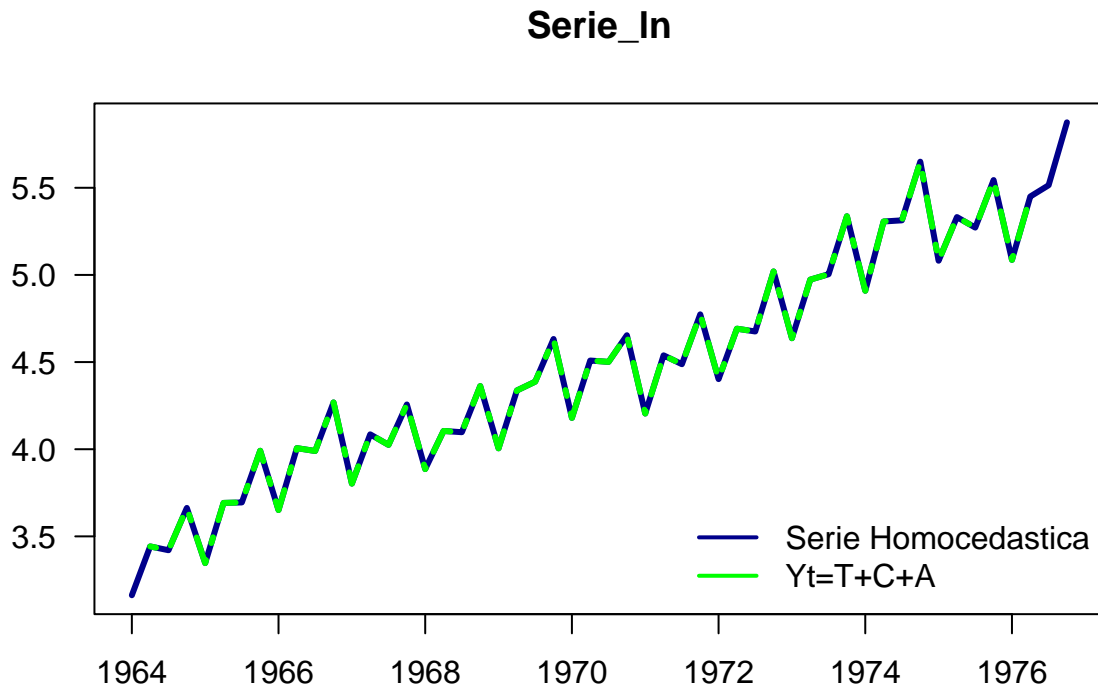
Serie con varianza costante



$Y_t = \text{tendencia} + \text{ciclos} + \text{aleatoria}$



```
par(mfrow = c(1,1))
plot(Serie_sq,col="darkblue", las=1, lwd=3,main="Serie_ln", ylab="",xlab="")
invisible(lines(componentes, type="l", lwd=3, col="green",lty=6))
legend("bottomright", col = c("darkblue","green"), lty = 1, lwd = 2, bty = "n",
      legend = c("Serie Homocedastica","Yt=T+C+A"), cex = 1)
```



¡Logramos identificar los componentes de la serie!

Missing data

Suponga que las observaciones de 1971 Qtr1, 1973 Qtr2 y 1973 Qtr3, son datos faltantes NA, es decir, sustituya estas observaciones por NA.

Use al menos dos métodos de imputación de la paquetería `imputeTS`. ¿Cuál método es adecuado para estos datos? (Note que el valor imputado debe aproximarse al valor omitido).

Vamos a poner los datos que se piden, como NA:

```
Original=Serie[29]
Original[2]=Serie[38]
Original[3]=Serie[39]
Serie_2=Serie
Serie_2[29]=NA
Serie_2[38:39]=NA
Serie_2
```

```
##      Qtr1 Qtr2 Qtr3 Qtr4
## 1964 10.00 11.85 11.70 13.42
## 1965 11.20 13.63 13.65 15.93
## 1966 13.33 16.05 15.92 18.22
## 1967 14.46 16.69 16.20 18.12
## 1968 15.10 16.85 16.79 19.03
## 1969 16.04 18.81 19.25 21.46
## 1970 17.47 20.33 20.26 21.66
```

```
## 1971    NA 20.60 20.14 22.79
## 1972 19.38 22.01 21.86 25.20
## 1973 21.50    NA    NA 28.48
## 1974 24.10 28.16 28.23 31.92
## 1975 25.82 28.43 27.79 30.74
## 1976 25.87 29.70 30.41 34.52
```

Ahora, veamos método por método:

Un vistazo a la documentación de la paquetería mencionada nos menciona los siguientes métodos:

na_interpolation:

Missing Value Imputation by Interpolation. Acepta 3 tipos:

“linear” - for linear interpolation using approx (default choice)

```
Comparaciones=data.frame(Original)
na_interpolation_lineal=na_interpolation(Serie_2,option='linear')[29]-Original[1]
na_interpolation_lineal[2:3]=na_interpolation(Serie_2,option='linear')[38:39]-Original[2:3]
Comparaciones['na_interpolation_lineal']=na_interpolation_lineal
Comparaciones
```

```
##    Original na_interpolation_lineal
## 1      17.68                3.4500000
## 2      24.73               -0.9033333
## 3      25.04                1.1133333
```

“spline” - for spline interpolation using spline

```
na_interpolation(Serie_2,option='spline')
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 21.91352 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 22.89979 27.71251 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“stine” - for Stineman interpolation using stinterp

```
na_interpolation(Serie_2,option='stine')
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 21.13000 20.60000 20.14000 22.79000
```

```
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 23.10575 26.86069 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

na_kalman:

Missing Value Imputation by Kalman Smoothing and State Space Models. Acepta los siguientes modelos:
`\begin{enumerate}` “auto.arima” - For using the state space representation of arima model (using auto.arima)
 (default choice)

```
na_kalman(Serie_2,model='auto.arima')
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 17.88402 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 24.99381 25.24268 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“StructTS” - For using a structural model fitted by maximum likelihood (using StructTS)

```
na_kalman(Serie_2,model='StructTS')
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 17.98114 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 24.82428 24.92793 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

na_locf:

Missing Value Imputation by Last Observation Carried Forward. Acepta los siguientes métodos:

“locf” - for Last Observation Carried Forward (default choice)

```
na_locf(Serie_2,option = 'locf')
```

```
##          Qtr1  Qtr2  Qtr3  Qtr4
## 1964 10.00 11.85 11.70 13.42
## 1965 11.20 13.63 13.65 15.93
```

```
## 1966 13.33 16.05 15.92 18.22
## 1967 14.46 16.69 16.20 18.12
## 1968 15.10 16.85 16.79 19.03
## 1969 16.04 18.81 19.25 21.46
## 1970 17.47 20.33 20.26 21.66
## 1971 21.66 20.60 20.14 22.79
## 1972 19.38 22.01 21.86 25.20
## 1973 21.50 21.50 21.50 28.48
## 1974 24.10 28.16 28.23 31.92
## 1975 25.82 28.43 27.79 30.74
## 1976 25.87 29.70 30.41 34.52
```

“nocb” - for Next Observation Carried Backward

```
na_locf(Serie_2,option = 'nocb')
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
## 1964 10.00 11.85 11.70 13.42
## 1965 11.20 13.63 13.65 15.93
## 1966 13.33 16.05 15.92 18.22
## 1967 14.46 16.69 16.20 18.12
## 1968 15.10 16.85 16.79 19.03
## 1969 16.04 18.81 19.25 21.46
## 1970 17.47 20.33 20.26 21.66
## 1971 20.60 20.60 20.14 22.79
## 1972 19.38 22.01 21.86 25.20
## 1973 21.50 28.48 28.48 28.48
## 1974 24.10 28.16 28.23 31.92
## 1975 25.82 28.43 27.79 30.74
## 1976 25.87 29.70 30.41 34.52
```

na_ma Missing:

Value Imputation by Weighted Moving Average. Acepta los siguientes métodos:

“simple” - Simple Moving Average (SMA)

```
na_ma(Serie_2,weighting = 'simple')
```

```
##      Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 20.32875 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 24.47286 25.36143 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“linear” - Linear Weighted Moving Average (LWMA)

```
na_ma(Serie_2,weighting = 'linear')
```

```
##      Qtr1      Qtr2      Qtr3      Qtr4
```

```
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 20.55065 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 24.27452 25.54742 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“exponential” - Exponential Weighted Moving Average (EWMA) (default choice)

```
na_ma(Serie_2,weighting = 'exponential')
```

```
##          Qtr1    Qtr2    Qtr3    Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 20.75900 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 24.03682 25.77500 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

na__mean Missing:

Value Imputation by Mean Value. Acepta los siguientes métodos:

“mean” - take the mean for imputation (default choice)

```
na_mean(Serie_2,option = 'mean')
```

```
##          Qtr1  Qtr2  Qtr3  Qtr4
## 1964 10.00 11.85 11.70 13.42
## 1965 11.20 13.63 13.65 15.93
## 1966 13.33 16.05 15.92 18.22
## 1967 14.46 16.69 16.20 18.12
## 1968 15.10 16.85 16.79 19.03
## 1969 16.04 18.81 19.25 21.46
## 1970 17.47 20.33 20.26 21.66
## 1971 20.43 20.60 20.14 22.79
## 1972 19.38 22.01 21.86 25.20
## 1973 21.50 20.43 20.43 28.48
## 1974 24.10 28.16 28.23 31.92
## 1975 25.82 28.43 27.79 30.74
## 1976 25.87 29.70 30.41 34.52
```

“median” - take the median for imputation

```
na_mean(Serie_2,option = 'median')
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
## 1964 10.00 11.85 11.70 13.42
## 1965 11.20 13.63 13.65 15.93
## 1966 13.33 16.05 15.92 18.22
## 1967 14.46 16.69 16.20 18.12
## 1968 15.10 16.85 16.79 19.03
## 1969 16.04 18.81 19.25 21.46
## 1970 17.47 20.33 20.26 21.66
## 1971 19.38 20.60 20.14 22.79
## 1972 19.38 22.01 21.86 25.20
## 1973 21.50 19.38 19.38 28.48
## 1974 24.10 28.16 28.23 31.92
## 1975 25.82 28.43 27.79 30.74
## 1976 25.87 29.70 30.41 34.52
```

“mode” - take the mode for imputation

```
na_mean(Serie_2,option = 'mode')
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
## 1964 10.00 11.85 11.70 13.42
## 1965 11.20 13.63 13.65 15.93
## 1966 13.33 16.05 15.92 18.22
## 1967 14.46 16.69 16.20 18.12
## 1968 15.10 16.85 16.79 19.03
## 1969 16.04 18.81 19.25 21.46
## 1970 17.47 20.33 20.26 21.66
## 1971 10.00 20.60 20.14 22.79
## 1972 19.38 22.01 21.86 25.20
## 1973 21.50 10.00 10.00 28.48
## 1974 24.10 28.16 28.23 31.92
## 1975 25.82 28.43 27.79 30.74
## 1976 25.87 29.70 30.41 34.52
```

“harmonic” - take the harmonic mean

```
na_mean(Serie_2,option = 'harmonic')
```

```
##      Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 18.67181 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 18.67181 18.67181 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“geometric” - take the geometric mean


```
na_mean(Serie_2,option = 'geometric')
```

##		Qtr1	Qtr2	Qtr3	Qtr4
##	1964	10.00000	11.85000	11.70000	13.42000
##	1965	11.20000	13.63000	13.65000	15.93000
##	1966	13.33000	16.05000	15.92000	18.22000
##	1967	14.46000	16.69000	16.20000	18.12000
##	1968	15.10000	16.85000	16.79000	19.03000
##	1969	16.04000	18.81000	19.25000	21.46000
##	1970	17.47000	20.33000	20.26000	21.66000
##	1971	19.54094	20.60000	20.14000	22.79000
##	1972	19.38000	22.01000	21.86000	25.20000
##	1973	21.50000	19.54094	19.54094	28.48000
##	1974	24.10000	28.16000	28.23000	31.92000
##	1975	25.82000	28.43000	27.79000	30.74000
##	1976	25.87000	29.70000	30.41000	34.52000

na_random:

Missing Value Imputation by Random Sample

```
na_random(Serie_2)
```

##		Qtr1	Qtr2	Qtr3	Qtr4
##	1964	10.00000	11.85000	11.70000	13.42000
##	1965	11.20000	13.63000	13.65000	15.93000
##	1966	13.33000	16.05000	15.92000	18.22000
##	1967	14.46000	16.69000	16.20000	18.12000
##	1968	15.10000	16.85000	16.79000	19.03000
##	1969	16.04000	18.81000	19.25000	21.46000
##	1970	17.47000	20.33000	20.26000	21.66000
##	1971	33.81456	20.60000	20.14000	22.79000
##	1972	19.38000	22.01000	21.86000	25.20000
##	1973	21.50000	11.53727	17.69228	28.48000
##	1974	24.10000	28.16000	28.23000	31.92000
##	1975	25.82000	28.43000	27.79000	30.74000
##	1976	25.87000	29.70000	30.41000	34.52000

na_seadec:

Seasonally Decomposed Missing Value Imputation. Admite los siguiente métodos:

“interpolation” - Imputation by Interpolation (default choice)

```
na_seadec(Serie_2,algorithm = 'interpolation')
```

##		Qtr1	Qtr2	Qtr3	Qtr4
##	1964	10.00000	11.85000	11.70000	13.42000
##	1965	11.20000	13.63000	13.65000	15.93000
##	1966	13.33000	16.05000	15.92000	18.22000
##	1967	14.46000	16.69000	16.20000	18.12000
##	1968	15.10000	16.85000	16.79000	19.03000
##	1969	16.04000	18.81000	19.25000	21.46000
##	1970	17.47000	20.33000	20.26000	21.66000
##	1971	18.02499	20.60000	20.14000	22.79000
##	1972	19.38000	22.01000	21.86000	25.20000
##	1973	21.50000	24.90596	25.54880	28.48000
##	1974	24.10000	28.16000	28.23000	31.92000

```
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“locf” - Imputation by Last Observation Carried Forward

```
na_seadec(Serie_2,algorithm = 'locf')
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 17.78802 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 23.93409 23.60507 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“mean” - Imputation by Mean Value

```
na_seadec(Serie_2,algorithm = 'mean')
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 18.33885 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 20.67408 20.34506 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“random” - Imputation by Random Sample

```
na_seadec(Serie_2,algorithm = 'random')
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 23.24167 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 29.38255 20.78890 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
```

```
## 1976 25.87000 29.70000 30.41000 34.52000
```

“kalman” - Imputation by Kalman Smoothing and State Space Models

```
na_seadec(Serie_2,algorithm = 'kalman')
```

```
##          Qtr1    Qtr2    Qtr3    Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 17.86519 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 24.88983 24.98804 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“ma” - Imputation by Weighted Moving Average

```
na_seadec(Serie_2,algorithm = 'ma')
```

```
##          Qtr1    Qtr2    Qtr3    Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 18.16021 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 24.58540 25.62559 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

na_seasplit:

Seasonally Splitted Missing Value Imputation. Admite los siguiente métodos:

“interpolation” - Imputation by Interpolation (default choice)

```
na_seasplit(Serie_2,algorithm = 'interpolation')
```

```
##          Qtr1    Qtr2    Qtr3    Qtr4
## 1964 10.000 11.850 11.700 13.420
## 1965 11.200 13.630 13.650 15.930
## 1966 13.330 16.050 15.920 18.220
## 1967 14.460 16.690 16.200 18.120
## 1968 15.100 16.850 16.790 19.030
## 1969 16.040 18.810 19.250 21.460
## 1970 17.470 20.330 20.260 21.660
## 1971 18.425 20.600 20.140 22.790
## 1972 19.380 22.010 21.860 25.200
## 1973 21.500 25.085 25.045 28.480
```

```
## 1974 24.100 28.160 28.230 31.920
## 1975 25.820 28.430 27.790 30.740
## 1976 25.870 29.700 30.410 34.520
```

“locf” - Imputation by Last Observation Carried Forward

```
na_seasplit(Serie_2,algorithm = 'locf')
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
## 1964 10.00 11.85 11.70 13.42
## 1965 11.20 13.63 13.65 15.93
## 1966 13.33 16.05 15.92 18.22
## 1967 14.46 16.69 16.20 18.12
## 1968 15.10 16.85 16.79 19.03
## 1969 16.04 18.81 19.25 21.46
## 1970 17.47 20.33 20.26 21.66
## 1971 17.47 20.60 20.14 22.79
## 1972 19.38 22.01 21.86 25.20
## 1973 21.50 22.01 21.86 28.48
## 1974 24.10 28.16 28.23 31.92
## 1975 25.82 28.43 27.79 30.74
## 1976 25.87 29.70 30.41 34.52
```

“mean” - Imputation by Mean Value

```
na_seasplit(Serie_2,algorithm = 'mean')
```

```
##      Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 17.85583 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 20.25917 20.18333 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“random” - Imputation by Random Sample

```
na_seasplit(Serie_2,algorithm = 'random')
```

```
##      Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 20.39834 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 12.87674 22.31445 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
```

```
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“kalman” - Imputation by Kalman Smoothing and State Space Models

```
na_seasplit(Serie_2,algorithm = 'kalman')
```

```
##           Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 18.37287 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 25.07742 24.68497 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

“ma” - Imputation by Weighted Moving Average

```
na_seasplit(Serie_2,algorithm = 'ma')
```

```
##           Qtr1      Qtr2      Qtr3      Qtr4
## 1964 10.00000 11.85000 11.70000 13.42000
## 1965 11.20000 13.63000 13.65000 15.93000
## 1966 13.33000 16.05000 15.92000 18.22000
## 1967 14.46000 16.69000 16.20000 18.12000
## 1968 15.10000 16.85000 16.79000 19.03000
## 1969 16.04000 18.81000 19.25000 21.46000
## 1970 17.47000 20.33000 20.26000 21.66000
## 1971 18.78800 20.60000 20.14000 22.79000
## 1972 19.38000 22.01000 21.86000 25.20000
## 1973 21.50000 24.70172 24.58724 28.48000
## 1974 24.10000 28.16000 28.23000 31.92000
## 1975 25.82000 28.43000 27.79000 30.74000
## 1976 25.87000 29.70000 30.41000 34.52000
```

Ajuste

Con los datos observados completos, ajuste un modelo ARIMA o SARIMA adecuado.

Obtenga correlogramas, revise si los parámetros son significativos, compruebe los supuestos (que los residuales sean ruido blanco con distribución normal). Obtenga dos o más posibles modelos, realice análisis de residuales y calcule medidas de bondad de ajuste. Haga la comparación para decidir cuál modelo sería el más adecuado.

Forecasting

Con el modelo estimado, pronostique $n_{new} = 8$ (2 años) valores futuros (obtenga intervalos de predicción).