Series de tiempo

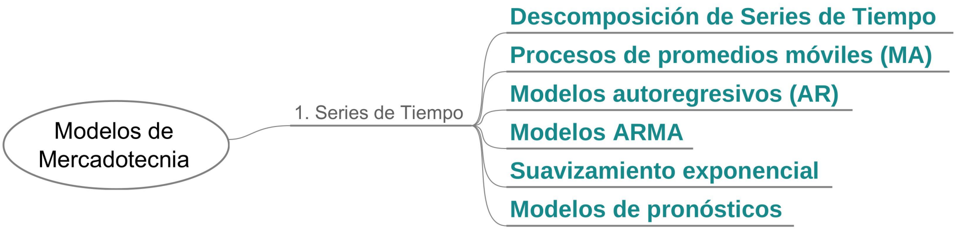
Modelos de Mercadotecnia

Jorge de la Vega Góngora

Maestría de Mercadotecnia

Sesión 2

Programa del Curso



Datos Temporales

1. Análisis de series de tiempo
   - Introducción a las series de tiempo
   - Enfoque descriptivo (Descomposición de series temporales)
   - Suavizamiento
   - Suavizamiento Avanzado

2. Modelos ARIMA
   - Metodología de Box-Jenkins
   - Análisis Box-Jenkins
   - Ruido blanco
   - Modelos AR, MA y ARMA

3. Anexo

Introducción

- Las *series de tiempo* son muy comunes en la vida cotidiana diaria contemporánea. Existe una gran cantidad de aplicaciones para las series de tiempo.

Series de tiempo

- Una **serie de tiempo** $\{y_t\}$ es una sucesión cronológica de observaciones, indexadas por un conjunto de índices $\mathcal{T}$ que usualmente representa tiempo, de un fenómeno que se mide a través una variable y_t.
- Usualmente se consideran tiempos *igualmente espaciados*, pero no es necesario. Las series que tienen observaciones que no están regularmente espaciadas se llaman *series irregulares*.

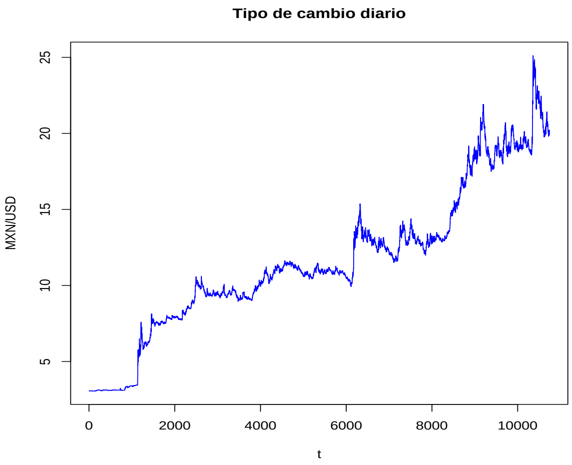
- Cualquier sucesión de observaciones puede transformarse en una serie regular simplemente adoptando una escala de tiempo suficientemente fina o suficientemente gruesa. Entre observaciones adyacentes separadas por más de una unidad en esta escala, se registran valores no disponibles (**NA** o **ND**).
- En series irregulares, se debe asegurar que se está midiendo la misma cantidad. Por ejemplo, en ventas semanales, puede que no se registre en una semana. Esto puede producir, o un dato no disponible, o un valor que se adiciona a las ventas de la siguiente semana.

Introducción

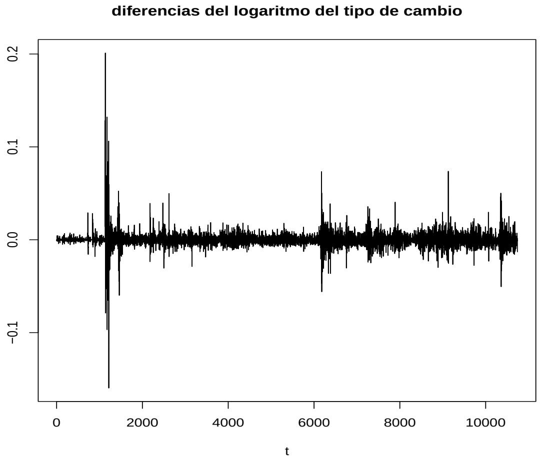
- La propiedad de *orden* es la que es crucial a los datos que conforman a la serie de tiempo. A diferencia de una muestra aleatoria de una v.a. X_n, las observaciones de una serie de tiempo están *ordenadas* por el índice temporal.
- La serie de tiempo que toma T valores se puede representar como la sucesión $y_1, y_2, \ldots, y_T$, o bien, con la siguiente expresión: $\{y_t\}_{t=1}^{T}$.
- Usualmente, una serie de tiempo se visualiza mediante una gráfica { (t, y_t) } y uniendo los puntos que son consecutivos en el tiempo.

Veamos algunos ejemplos.

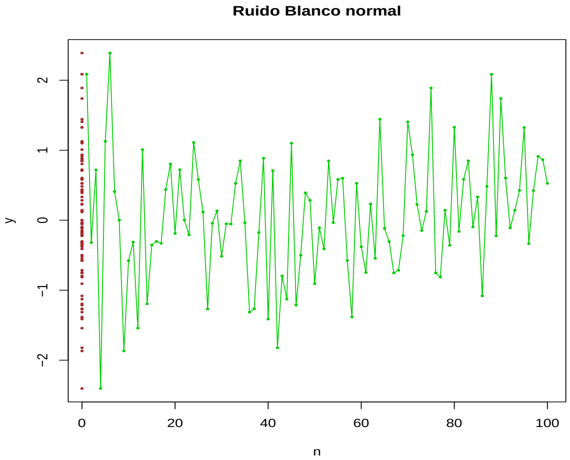
Ejemplos: tipos de cambio



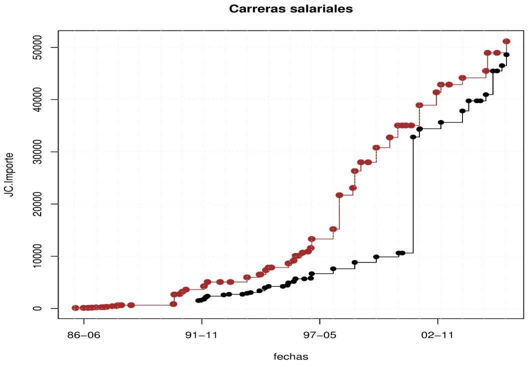
Ejemplos: Diferencias en los tipos de cambio



Ejemplos: Ruido Blanco



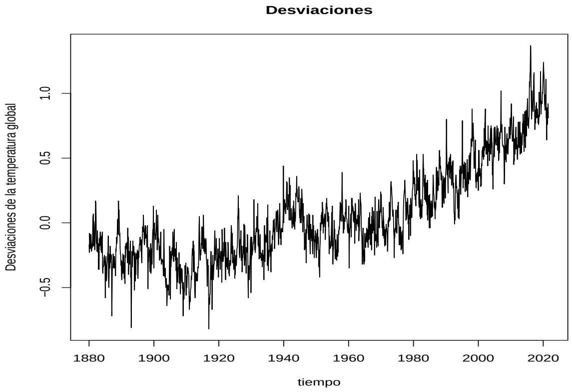
Ejemplos



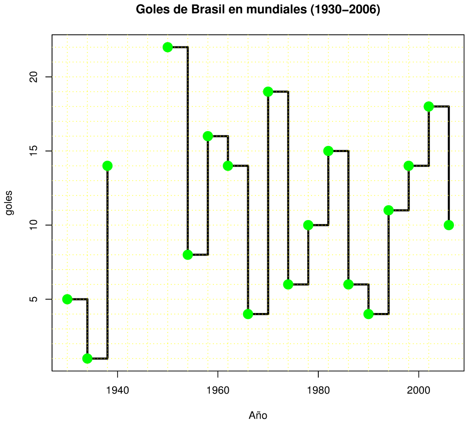
Carreras salariales
JC.Importe
0
10000
20000
30000
40000
50000
86–06
91–11
97–05
02–11
fechas

Ejemplos

```r
data <- read.csv("https://data.giss.nasa.gov/gistemp/tabledata_v4/GLB.Ts+dSST.csv",skip=1)
data <- data[,2:13]; A <- t(data)
dim(A) <- c(1704,1)
A <- ts(as.numeric(A),start = 1880,freq=12)
plot(A,main = "Desviaciones", xlab = "tiempo", ylab = "Desviaciones de la temperatura global")
```



Ejemplos



Goles de Brasil en mundiales (1930–2006)
goles
Año
5
10
15
20
1940
1960
1980
2000

Otros ejemplos de series temporales

- Otras situaciones prácticas que dan origen a las series de tiempo:
  - Ventas mensuales de vehículos de una agencia.
  - Número de desempleados en el tiempo (Índices de empleo).
  - Demanda mensual o trimestral de cualquier producto.
  - Demanda intermitente de productos, como partes de avión.
  - Población de una ciudad o un país a lo largo del tiempo.
  - Proporción de células cancerígenas en un órgano cada semana.
  - Índices de las diferentes bolsas de valores, o bien, los precios de productos financieros.
- Las fuentes de las series de tiempo pueden ser muy variadas, muchas son de origen industrial (en los procesos de producción), o bien, de naturaleza económica, física, financiera, médica, mercadotécnica, etc.

Objetivo

Una serie de tiempo se puede analizar con varios propósitos en mente. Nos interesa comprender el comportamiento de la variable y_t a lo largo del tiempo,

- para comprender sus fluctuaciones con respecto a un valor promedio, (variabilidad de la serie);
- identificar tendencia y nivel de la serie,
- para proyectar su posible comportamiento hacia el futuro, generando pronósticos y predicciones útiles para la planeación y toma de decisiones;
- para tratar de encontrar un modelo matemático que explique el *proceso generador de los datos* así como su *comportamiento*, a través de si misma o de otros posibles *predictores*.

Un supuesto fundamental de las series de tiempo es que es factible que *el comportamiento pasado de una variable nos ayude a explicar el actual o futuro*.

Series de tiempo vs. pronósticos

- Pronosticar y analizar una serie de tiempo son dos actividades distintas.
  - Un *pronóstico* es una visión de un futuro incierto.
  - *Analizar una serie de tiempo* es describir el proceso generador de los datos.
- Podemos analizar una serie de tiempo sin considerar como objetivo hacer pronósticos, así como podemos hacer pronósticos sin pensar en series de tiempo (o de hecho cualquier otro tipo de análisis).
- Los principales precursores de la actividad de pronóstico son la construcción de un modelo adecuado basado en el desarrollo histórico de la serie y la utilización de información relevante al posible desarrollo futuro de la serie.

Modelación de series temporales

Hay dos enfoques para trabajar con las series de tiempo:

1. El **enfoque descriptivo**, que describe y resume en forma concisa un conjunto de datos, mediante estadísticas sumarias y/o gráficas que puedan ayudar a tener una visión general del comportamiento de la serie, pero que no necesariamente ayuda a comprender la naturaleza de ese comportamiento. Esto incluye el *suavizamiento* de la serie, y los *métodos de descomposición*.
2. El **enfoque inferencial**, cuyo objetivo fundamental es utilizar datos de una muestra para realizar inferencias, que sean válidas para toda la población de donde se obtuvo la muestra. Incluye los *modelos de predicción*. En las series de tiempo, la población sobre la que se desea inferir depende del tipo de análisis y/o modelo que se emplee ya sea *procesos estocásticos* o bien *modelos estructurales*.

1. Análisis de series de tiempo
   - Introducción a las series de tiempo
   - Enfoque descriptivo (Descomposición de series temporales)
   - Suavizamiento
   - Suavizamiento Avanzado
2. Modelos ARIMA
   - Metodología de Box-Jenkins
   - Análisis Box-Jenkins
   - Ruido blanco
   - Modelos AR, MA y ARMA
3. Anexo

Componentes de una serie de tiempo

Una forma de modelar una serie de tiempo se obtiene al suponer que está conformada de diferentes componentes. Los componentes básicos son los siguientes:

- *Tendencia*: el movimiento hacia arriba o hacia abajo que caracteriza a una serie de tiempo en un periodo de tiempo dado. Muestra el crecimiento o decline de una serie en el largo plazo.
- *Ciclo*: se refiere a movimientos recurrentes hacia arriba o hacia abajo alrededor del nivel de tendencia. Usualmente son movimientos que se observan en periodos mayores a un año. En muchas ocasiones no se separa explícitamente y se considera parte de la tendencia.
- *Estacionalidad*: es el comportamiento que muestra una serie a lo largo de un periodo de tiempo, típicamente un año, y que se repite durante varios periodos.
- *Error*: se refiere a la parte de la serie que es completamente aleatoria; es el error estadístico que se presenta por incertidumbre.

Método clásico de descomposición. I

El método de descomposición clásico para series de tiene el objetivo de separar los componentes de una serie de tiempo. Este método supone que una serie de tiempo se puede escribir como una función de tres componentes:

$$X_t = f(TC_t, S_t, E_t), \quad \text{donde:}$$

- f es una función que relaciona los componentes,
- TC es un componente de tendencia-ciclo,
- S es un componente estacional,
- E es el componente de error o ruido de la serie.

Método clásico de descomposición. II

El método de descomposición se debe a varias personas, principalmente a Macauley (1930).
Aquí se supondrá que el modelo es *multiplicativo*, los componentes se multiplican (la función f es el producto):

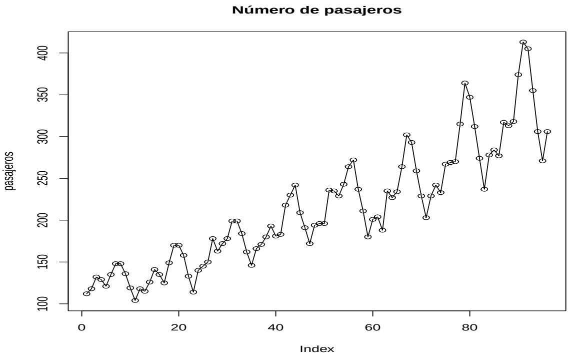
$$X_t = S_t \times TC_t \times E_t$$

Mostraré el método a través de un ejemplo. Se tienen 8 años de datos mensuales del número de pasajeros internacionales de una línea aérea. La serie se muestra en la figura siguiente.

```
pasajeros <- c(112, 118 , 132 , 129 , 121 , 135 , 148 , 148 , 136 , 119 , 104 , 118,
115 , 126 , 141 , 135 , 125 , 149 , 170 , 170 , 158 , 133 , 114 , 140,
145 , 150 , 178 , 163 , 172 , 178 , 199 , 199 , 184 , 162 , 146 , 166,
171 , 180 , 193 , 181 , 183 , 218 , 230 , 242 , 209 , 191 , 172 , 194,
196 , 196 , 236 , 235 , 229 , 243 , 264 , 272 , 237 , 211 , 180 , 201,
204 , 188 , 235 , 227 , 234 , 264 , 302 , 293 , 259 , 229 , 203 , 229,
242 , 233 , 267 , 269 , 270 , 315 , 364 , 347 , 312 , 274 , 237 , 278,
284 , 277 , 317 , 313 , 318 , 374 , 413 , 405 , 355 , 306 , 271 , 306)
n <- length(pasajeros)  #número de observaciones disponibles
n

[1] 96
```

Método clásico de descomposición. III

```
plot(pasajeros, type="o", main="Número de pasajeros")
```



Para dar formato de serie de tiempo, usamos la función `ts`:

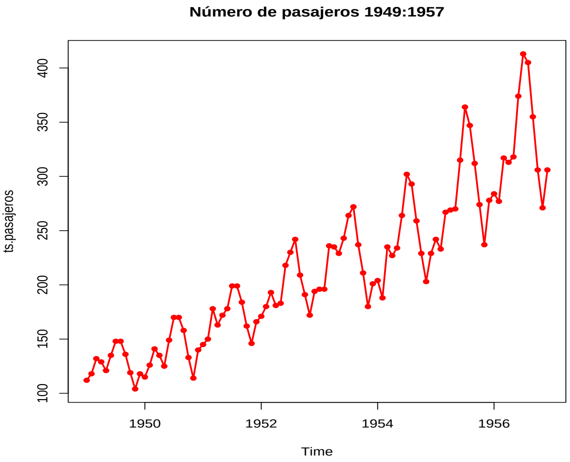
Método clásico de descomposición. IV

```r
ts.pasajeros <- ts(data = pasajeros,start = c(1949,1), frequency = 12)
ts.pasajeros
```

```
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
```

```r
plot(ts.pasajeros, main = "Número de pasajeros 1949:1957", type="o", pch=16, col="red", lwd=2)
```

Método clásico de descomposición. V



Pasos del método de descomposición I

- Aplicaremos los siguientes pasos a los datos:

Paso 1 Se calcula un promedio móvil de tamaño 12 y se coloca en el mes de julio (se centra con respecto a la unidad de tiempo, que aquí es años). La observación de julio de 1949 por ejemplo, se calcula como:

$$\frac{112 + 118 + 132 + \cdots + 119 + 104 + 118}{12} = 126.667$$

La siguiente observación para agosto es $126.67 - 112/12 + 115/12 = 126.917$ y así sucesivamente. Habrá seis valores perdidos al principio y cinco al final, por los promedios considerados. Esto genera los siguientes datos:

Pasos del método de descomposición II

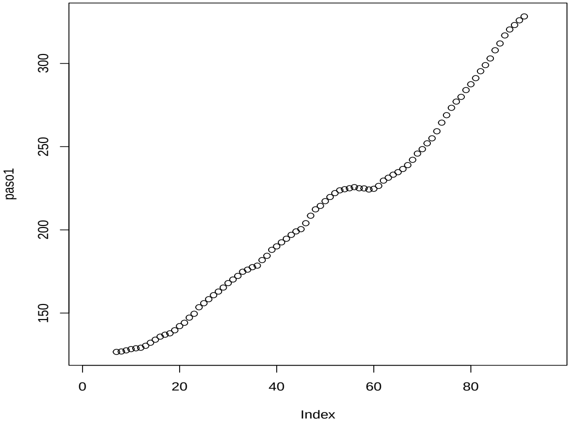
```
paso1 <- rep(NA,n)
for(i in 1:(n-6))
paso1[i+6] <- mean(ts.pasajeros[i:(i+11)])

paso1
```

```
 [1]       NA       NA       NA       NA       NA       NA 126.6667 126.9167
 [9] 127.5833 128.3333 128.8333 129.1667 130.3333 132.1667 134.0000 135.8333
[17] 137.0000 137.8333 139.6667 142.1667 144.1667 147.2500 149.5833 153.5000
[25] 155.9167 158.3333 160.7500 162.9167 165.3333 168.0000 170.1667 172.3333
[33] 174.8333 176.0833 177.5833 178.5000 181.8333 184.4167 188.0000 190.0833
[41] 192.5000 194.6667 197.0000 199.0833 200.4167 204.0000 208.5000 212.3333
[49] 214.4167 217.2500 219.7500 222.0833 223.7500 224.4167 225.0000 225.6667
[57] 225.0000 224.9167 224.2500 224.6667 226.4167 229.5833 231.3333 233.1667
[65] 234.6667 236.5833 238.9167 242.0833 245.8333 248.5000 252.0000 255.0000
[73] 259.2500 264.4167 268.9167 273.3333 277.0833 279.9167 284.0000 287.5000
[81] 291.1667 295.3333 299.0000 303.0000 307.9167 312.0000 316.8333 320.4167
[89] 323.0833 325.9167 328.2500       NA       NA       NA       NA       NA
```

Si se tuvieran datos trimestrales, el promedio móvil sería de tamaño 4 y se colocaría en el tercer trimestre. Si los datos fueran semestrales, el promedio móvil sería de tamaño dos y se colocaría en el segundo semestre. Una gráfica del paso 1 muestra una serie básicamente suave, correspondiente a la tendencia. El paso 2 ayuda a centrar los datos en el lugar correcto

```
plot(paso1)
```

Pasos del método de descomposición III



Pasos del método de descomposición IV

Paso 2 Se calcula un promedio móvil de tamaño 2 del promedio móvil del paso anterior y se coloca el primer promedio móvil en el mes de julio de 1949. Esto da un *promedio móvil centrado* de tamaño 12 al final. La primera observación en julio de 1949 es $(126.667 + 126.917)/2 = 126.792$. Al final se pierde un dato más. Este paso genera se genera de la siguiente manera:

```
paso2 <- rep(NA,n)
for(i in 1:(n-6)) paso2[6+i] <- mean(paso1[(6+i):(7+i)])
paso2
```

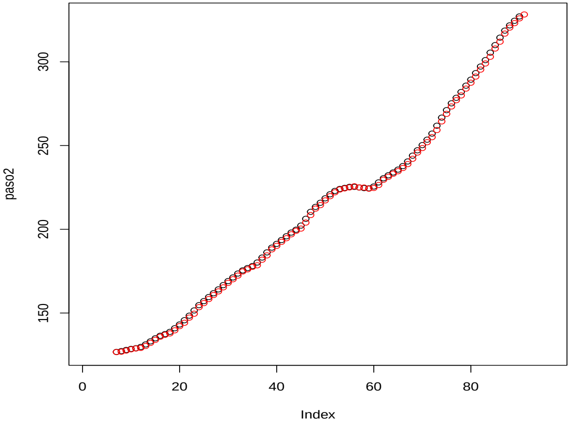
```
 [1]       NA       NA       NA       NA       NA       NA 126.7917 127.2500
 [9] 127.9583 128.5833 129.0000 129.7500 131.2500 133.0833 134.9167 136.4167
[17] 137.4167 138.7500 140.9167 143.1667 145.7083 148.4167 151.5417 154.7083
[25] 157.1250 159.5417 161.8333 164.1250 166.6667 169.0833 171.2500 173.5833
[33] 175.4583 176.8333 178.0417 180.1667 183.1250 186.2083 189.0417 191.2917
[41] 193.5833 195.8333 198.0417 199.7500 202.2083 206.2500 210.4167 213.3750
[49] 215.8333 218.5000 220.9167 222.9167 224.0833 224.7083 225.3333 225.3333
[57] 224.9583 224.5833 224.4583 225.5417 228.0000 230.4583 232.2500 233.9167
[65] 235.6250 237.7500 240.5000 243.9583 247.1667 250.2500 253.5000 257.1250
[73] 261.8333 266.6667 271.1250 275.2083 278.5000 281.9583 285.7500 289.3333
[81] 293.2500 297.1667 301.0000 305.4583 309.9583 314.4167 318.6250 321.7500
[89] 324.5000 327.0833       NA       NA       NA       NA       NA       NA
```

Si los datos fueran trimestrales o semestrales, siempre el segundo promedio móvil es de tamaño 2 y se coloca a la misma altura del primer promedio móvil.

Una gráfica del paso 2 es como la siguiente, que muestra la parte de tendencia de la serie

```
plot(paso2)
points(paso1,col="red") # solo se añade para comparar
```

Pasos del método de descomposición V



Pasos del método de descomposición VI

Paso 3 Se divide la columna generada en el paso anterior entre los datos y se multiplican por 100 para que puedan ser expresados como porcentajes. Este procedimiento separa el componente de tendencia-ciclo de los datos, pues los promedios móviles dobles eliminan el error y la estacionalidad y precisamente capturan la tendencia de los datos:

$$R_t = \frac{X_t}{TC_t} = S_t \times E_t$$

```
Rt <- 100*ts.pasajeros/paso2
```

Paso 4 Para separar el componente estacional S_t de la serie, hay que calcular los promedios de los valores R_t en cada mes.

```
St <- numeric(12)
for(i in 0:11)
St[ifelse((7+i) %% 12 == 0, 12, (7+i) %% 12)] <- mean(Rt[(1:n) %% 12 == i],na.rm = T)/100

St

 [1] 1.0988912 1.1997467 1.1879806 1.0564024 0.9200596 0.7980033 0.9052642
 [8] 0.9108800 0.9030236 1.0322468 0.9862632 0.9787807
```

Así, el factor estacional para el los meses de enero es 0.9106, para los meses de febrero es 0.8346 y así sucesivamente (recuerden que originalmente se multiplicaron los datos por 100, por eso se dividen por 100 ahora).

Pasos del método de descomposición VII

Paso 5 Por último, el componente de error se obtiene dividiendo la serie sin tendencia R_t por los factores estacionales:

$$E_t = R_t/S_t = \frac{S_t \times E_t}{S_t}$$

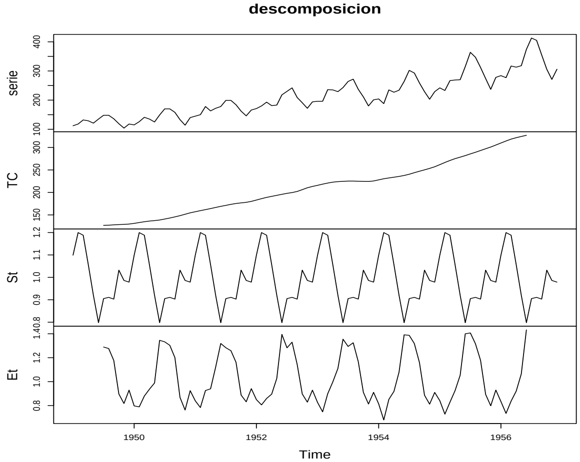
```
Et <- Rt/(100*St)
Et

           Jan       Feb       Mar       Apr       May       Jun       Jul
1949        NA        NA        NA        NA        NA        NA 1.2894237
1950 0.7973405 0.7891459 0.8797194 0.9367786 0.9886775 1.3457010 1.3326350
1951 0.8397848 0.7836598 0.9258544 0.9401205 1.1216665 1.3192117 1.2836515
1952 0.8497551 0.8057194 0.8593903 0.8956806 1.0274653 1.3949710 1.2829092
1953 0.8263858 0.7476788 0.8992370 0.9979205 1.1107337 1.3551345 1.2942052
1954 0.8142179 0.6799482 0.8517317 0.9186188 1.0793904 1.3914855 1.3871279
1955 0.8410770 0.7282787 0.8289577 0.9252548 1.0537137 1.3999770 1.4071481
1956 0.8337970 0.7343188 0.8374716 0.9208660 1.0651149 1.4328756        NA
           Aug       Sep       Oct       Nov       Dec
1949 1.2768584 1.1769858 0.8965587 0.8174304 0.9291573
1950 1.3036044 1.2008081 0.8681313 0.7627460 0.9245469
1951 1.2585888 1.1613008 0.8874979 0.8314543 0.9413438
1952 1.3300483 1.1445852 0.8971310 0.8288109 0.9289082
1953 1.3252026 1.1666677 0.9101676 0.8130999 0.9105082
1954 1.3185323 1.1604082 0.8864981 0.8119424 0.9099254
1955 1.3166484 1.1781958 0.8932374 0.7983420 0.9298383
1956        NA        NA        NA        NA        NA
```

Una gráfica con cada uno de los componentes de la serie se muestra en la figura siguiente.

Pasos del método de descomposición VIII

```r
descomposicion <- data.frame(serie = ts.pasajeros, TC = paso2, St = St, Et = Et)
plot.ts(descomposicion)
```

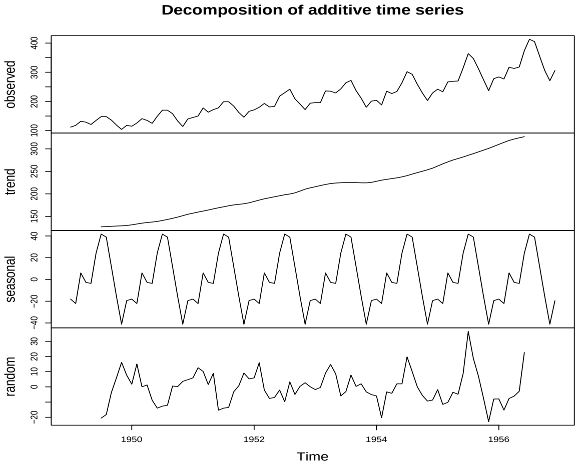
Pasos del método de descomposición IX



La función `decompose` en `R` permite realizar todos los pasos de manera eficiente y rápida.

Pasos del método de descomposición X

```
plot(decompose(ts.pasajeros))
```



Pasos del método de descomposición XI

Con cada uno de los componentes se puede reproducir el dato original. Por ejemplo, para el dato de Enero de 1950, que es 115, los componentes son: $TC = 131.25$, $S = 1.0988912$ y $E = 0.7973405$. Entonces

$$115 = 131.25 \times 1.0988912 \times 0.7973405$$

Los componentes finalmente se ven como en la tabla que se muestra a continuación:

Pasos del método de descomposición XII

```
descomposicion$St <- St
descomposicion$Et <- Et
head(descomposicion,25)
```

```
   serie       TC        St        Et
1    112       NA 1.0988912        NA
2    118       NA 1.1997467        NA
3    132       NA 1.1879806        NA
4    129       NA 1.0564024        NA
5    121       NA 0.9200596        NA
6    135       NA 0.7980033        NA
7    148 126.7917 0.9052642 1.2894237
8    148 127.2500 0.9108800 1.2768584
9    136 127.9583 0.9030236 1.1769858
10   119 128.5833 1.0322468 0.8965587
11   104 129.0000 0.9862632 0.8174304
12   118 129.7500 0.9787807 0.9291573
13   115 131.2500 1.0988912 0.7973405
14   126 133.0833 1.1997467 0.7891459
15   141 134.9167 1.1879806 0.8797194
16   135 136.4167 1.0564024 0.9367786
17   125 137.4167 0.9200596 0.9886775
18   149 138.7500 0.7980033 1.3457010
19   170 140.9167 0.9052642 1.3326350
20   170 143.1667 0.9108800 1.3036044
21   158 145.7083 0.9030236 1.2008081
22   133 148.4167 1.0322468 0.8681313
23   114 151.5417 0.9862632 0.7627460
24   140 154.7083 0.9787807 0.9245469
25   145 157.1250 1.0988912 0.8397848
```

Comentarios al método de descomposición

- La ventaja de tener identificados los componentes de la serie es que ahora podemos pronosticar cada componente por separado.
- En el procedimiento anterior, los pasos 1 y 2 esencialmente separan el componente de tendencia-ciclo TC_t de la serie. Pero se pueden aplicar muchos métodos distintos.
- En la siguiente sección consideraremos otras opciones de suavizamiento de la serie.
- Hay otros métodos de descomposición que incorporan otros métodos más sofisticados para obtener alguno de los componentes (tendencia o estacionalidad), por ejemplo:
  - Método STL
  - Método X-12-ARIMA

Promedios móviles dobles I

- Al promedio móvil de un promedio móvil se le llama **promedio móvil doble**. El propósito de hacer promedios móviles dobles es centrar las observaciones, en el sentido de hacer los pesos simétricos.
- Un promedio móvil de orden k de un promedio móvil de orden m se denota como $k \times m - MA$.
- Los promedios móviles dobles o triples generan promedios ponderados de las observaciones con diferentes pesos. Por ejemplo, un promedio $2 \times 4 - MA$ nos da:

$$
\begin{aligned}
\hat{T}_t &= \frac{1}{2}\left[\frac{1}{4}(y_{t-2} + y_{t-1} + y_t + y_{t+1}) + \frac{1}{4}(y_{t-1} + y_t + y_{t+1} + y_{t+2})\right] \\
&= \frac{1}{8}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{4}y_{t+1} + \frac{1}{8}y_{t+2}
\end{aligned}
$$

- Estos promedios ponderados también se conocen como *filtros*, y algunos son muy utilizados, como veremos más adelante.

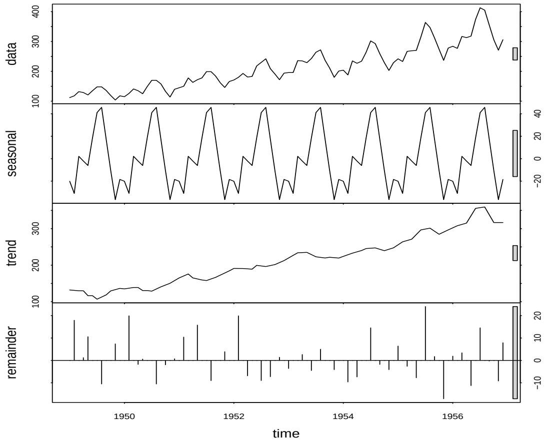
Método de descomposición STL I

- El método STL (*Seasonal and Trend decomposition using loess*) es un método robusto y versátil que utiliza un método conocido como *loess* para estimar relaciones no lineales.
- Sus principales ventajas incluyen:
  - Puede manejar cualquier periodicidad de estacionalidad
  - Se permite al componente estacional cambiar a lo largo del tiempo.
  - El nivel de suavizamiento de la tendencia se puede controlar.
  - Es robusto ante la presencia de valores extremos (outliers).
- El siguiente código considera que:
  - `s.window` es la ventana de loess para el ajuste estacional, en rezagos, o se especifica *periodic* para tomar la media.
  - `t.window` es la ventana de loess para la tendencia (en rezagos)

Método de descomposición STL II

```r
m.stl <- stl(ts.pasajeros, t.window=4, s.window="periodic", robust = T)
plot(m.stl)
```

Método de descomposición STL III



1. Análisis de series de tiempo
   - Introducción a las series de tiempo
   - Enfoque descriptivo (Descomposición de series temporales)
   - Suavizamiento
   - Suavizamiento Avanzado
2. Modelos ARIMA
   - Metodología de Box-Jenkins
   - Análisis Box-Jenkins
   - Ruido blanco
   - Modelos AR, MA y ARMA
3. Anexo

Suavizamiento

- Los métodos de suavizamiento utilizan los datos históricos para obtener un valor “suavizado”: remueve o disminuye la variabilidad de la serie, como se vió en el método de descomposición, pasos 1 y 2.
- Estos métodos se pueden utilizar como método de descomposición y como un método para obtener *pronósticos de corto plazo*.
- Dentro de los métodos de suavizamiento hay 2 clases que se utilizan mucho en la práctica:
  - *Métodos de promedio*: se calculan promedios ponderados para las observaciones de la serie.
  - *Métodos de suavizamiento exponencial*: se asignan pesos a las observaciones de tal forma que los datos más viejos tienen menor relevancia.

Datos para los ejemplos I

- En los métodos que veremos a continuación suponemos que tenemos una serie de tiempo $\{y_t\}$.
- Para propósitos ilustrativos, consideren la siguiente serie mensual de precios de algún producto. El primer dato corresponde a enero de 1980:

```r
library(forecast)
data(wineind)  # datos de ventas de vino
(yt <-wineind)
```

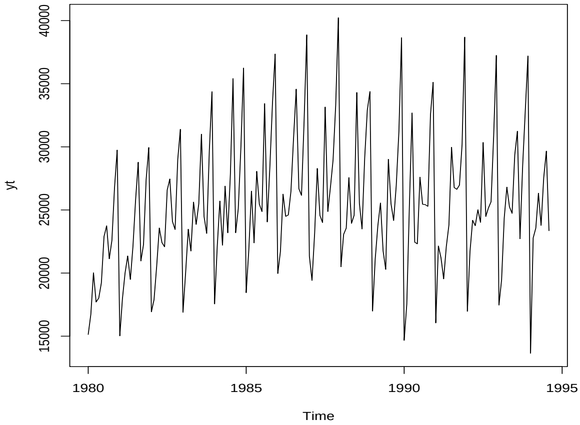
```
       Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
1980 15136 16733 20016 17708 18019 19227 22893 23739 21133 22591 26786 29740
1981 15028 17977 20008 21354 19498 22125 25817 28779 20960 22254 27392 29945
1982 16933 17892 20533 23569 22417 22084 26580 27454 24081 23451 28991 31386
1983 16896 20045 23471 21747 25621 23859 25500 30998 24475 23145 29701 34365
1984 17556 22077 25702 22214 26886 23191 27831 35406 23195 25110 30009 36242
1985 18450 21845 26488 22394 28057 25451 24872 33424 24052 28449 33533 37351
1986 19969 21701 26249 24493 24603 26485 30723 34569 26689 26157 32064 38870
1987 21337 19419 23166 28286 24570 24001 33151 24878 26804 28967 33311 40226
1988 20504 23060 23562 27562 23940 24584 34303 25517 23494 29095 32903 34379
1989 16991 21109 23740 25552 21752 20294 29009 25500 24166 26960 31222 38641
1990 14672 17543 25453 32683 22449 22316 27595 25451 25421 25288 32568 35110
1991 16052 22146 21198 19543 22084 23816 29961 26773 26635 26972 30207 38687
1992 16974 21697 24179 23757 25013 24019 30345 24488 25156 25650 30923 37240
1993 17466 19463 24352 26805 25236 24735 29356 31234 22724 28496 32857 37198
1994 13652 22784 23565 26323 23779 27549 29660 23356
```

Datos para los ejemplos II

- Para darnos una idea de los datos:

```
plot(yt)
```

Datos para los ejemplos III



- La diferencia entre el valor de la serie suavizada y la observación, es el *error*. Es similar a los residuales.

Promedios simples I

- Los promedios simples son los más sencillos de calcular.
- Este método consiste en usar como pronóstico F_t el promedio de todas las observaciones disponibles hasta el tiempo $t-1$,

$$F_t = \sum_{i=1}^{t-1} \frac{y_i}{t-1}$$

```
(Ft <- mean(yt))
```

```
[1] 25392.15
```

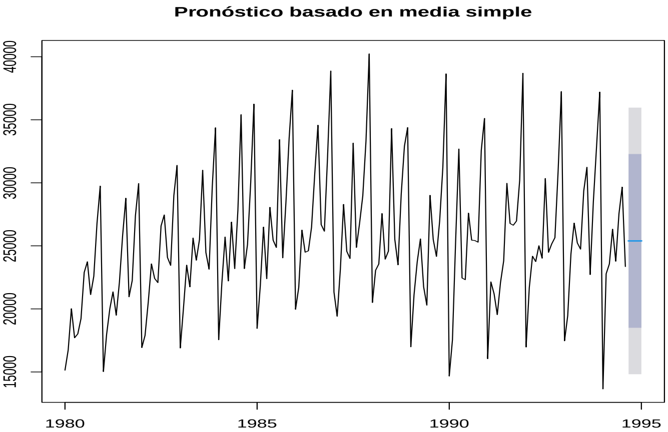
- A partir de este dato, se supone que el pronóstico h pasos adelante es constante. El paquete `forecast` hace el cálculo

```
meanf(yt,h=5) # horizonte de h=5 periodos
```

```
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Sep 1994       25392.15 18502.18 32282.11 14821.53 35962.76
Oct 1994       25392.15 18502.18 32282.11 14821.53 35962.76
Nov 1994       25392.15 18502.18 32282.11 14821.53 35962.76
Dec 1994       25392.15 18502.18 32282.11 14821.53 35962.76
Jan 1995       25392.15 18502.18 32282.11 14821.53 35962.76
```

```
plot(meanf(yt,h=5), main="Pronóstico basado en media simple")
```

Promedios simples II



Promedios simples III

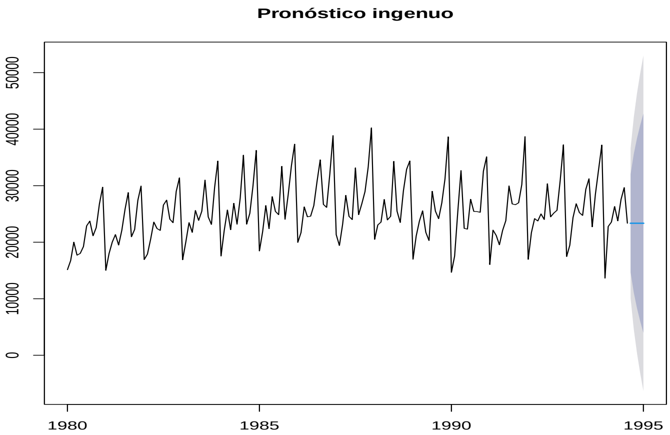
- Usualmente, este método no genera buenos pronósticos si la serie muestra una tendencia marcada. Sólo es apropiada cuando la serie es estacionaria: tiene un comportamiento muy regular en el tiempo.
- Una limitación es que le da la misma importancia a todos los puntos, inclusive los más alejados.

Pronóstico ingenuo I

- EL pronóstico ingenuo consiste simplemente en tomar como pronóstico el último dato observado. Es una alternativa al promedio simple.

```
naive(yt,h=5)

         Point Forecast     Lo 80    Hi 80      Lo 95    Hi 95
Sep 1994          23356 14678.259 32033.74 10084.5401 36627.46
Oct 1994          23356 11083.821 35628.18  4587.3214 42124.68
Nov 1994          23356  8325.711 38386.29   369.1571 46342.84
Dec 1994          23356  6000.518 40711.48 -3186.9199 49898.92
Jan 1995          23356  3951.981 42760.02 -6319.8866 53031.89

plot(naive(yt,h=5), main = "Pronóstico ingenuo")
```

Pronóstico ingenuo II



Promedios móviles (MA) I

- Un *promedio móvil simple de orden* k para el tiempo t se define como el promedio de las k observaciones anteriores a t:

$$F_t^{(k)} = \frac{y_{t-1} + y_{t-2} + \cdots + y_{t-k}}{k} = \frac{1}{k}\sum_{i=1}^{k} y_{t-i}$$

- Un promedio móvil de orden k se denota como $MA(k)$.
- La fórmula anterior se puede expresar de forma recursiva:

$$F_{t+1}^{(k)} = F_t^{(k)} + \frac{y_{t+1} - y_{t-k}}{k},$$

de este modo, cada nuevo pronóstico es un ajuste del pronóstico precedente. Este ajuste es cada vez más pequeño conforme k es mayor.

Promedios móviles (MA) II

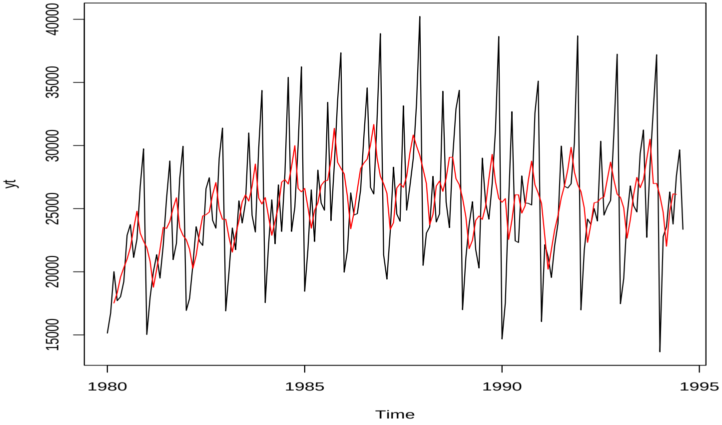
```r
options(width = 150)
(ma5 <- ma(yt, order = 5))
```

```
         Jan     Feb     Mar     Apr     May     Jun     Jul     Aug     Sep     Oct     Nov     Dec
1980      NA      NA 17522.4 18340.6 19572.6 20317.2 21002.2 21916.6 23428.4 24797.8 23055.6 22424.4
1981 21907.8 20821.4 18773.0 20192.4 21760.4 23514.6 23435.8 23987.0 25040.4 25866.0 23496.8 22883.2
1982 22539.0 21774.4 20268.8 21299.0 23036.6 24420.8 24523.2 24730.0 26111.4 27072.6 24961.0 24153.8
1983 24157.8 22709.0 21556.0 22948.6 24039.6 25545.0 26090.6 25595.4 26763.8 28536.8 25848.4 25368.8
1984 25880.2 24382.8 22887.0 24014.0 25164.8 27105.6 27301.8 26946.6 28310.2 29992.4 26601.2 26331.2
1985 26606.8 25083.8 23446.8 24847.0 25452.4 26839.6 27171.2 27249.6 28866.0 31361.8 28670.8 28200.6
1986 27760.6 25952.6 23403.0 24706.2 26510.6 28174.6 28613.8 28924.6 30040.4 31669.8 29023.4 27569.4
1987 26971.2 26215.6 23355.6 23888.4 26634.8 26977.2 26680.8 27560.2 29422.2 30837.2 29962.4 29213.6
1988 28132.6 26982.8 23725.6 24541.6 26790.2 27181.2 26367.6 27398.6 29062.4 29077.6 27372.4 26895.4
1989 25824.4 24354.2 21828.8 22489.4 24069.4 24421.4 24144.2 25185.8 27371.4 29297.8 27132.2 25807.6
1990 25506.2 25798.4 22560.0 24088.8 26099.2 26098.8 24646.4 25214.2 27264.6 28767.6 26887.8 26232.8
1991 25414.8 22809.8 20204.6 21757.4 23320.4 24435.4 25853.8 26831.4 28109.6 29854.8 27895.0 26907.4
1992 26348.8 25058.8 22324.0 23733.0 25462.6 25524.4 25804.2 25931.6 27312.4 28691.4 27287.0 26148.4
1993 25888.8 25065.2 22664.4 24118.2 26096.8 27473.2 26657.0 27309.0 28933.4 30501.8 26985.4 26997.4
1994 26011.2 24704.4 22020.6 24800.0 26175.2 26133.4      NA      NA
```

La gráfica que obtenemos es la siguiente:

```r
plot(yt)
lines(ma5,col="red")
```

Promedios móviles (MA) III

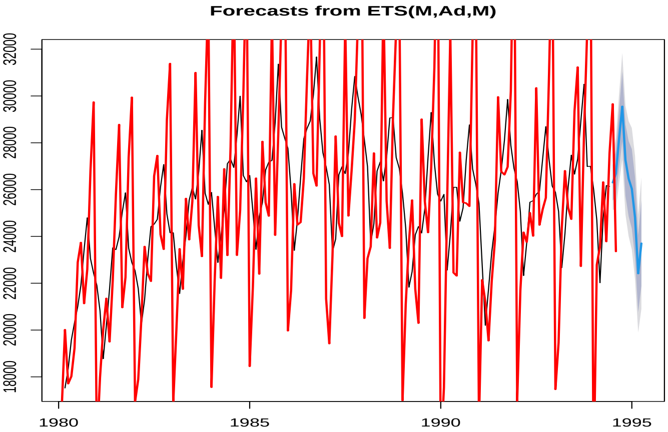


Promedios móviles (MA) IV

- Usualmente el promedio móvil sirve para suavizar una serie. Pero se puede hacer un pronóstico utilizando la serie suavizada, por ejemplo:

```
F5 <- forecast(ma(yt, order=5), h=10)
F5

         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jul 1994       26296.76 25539.29 27054.22 25138.31 27455.20
Aug 1994       26682.84 25684.16 27681.52 25155.49 28210.18
Sep 1994       28178.43 26927.06 29429.80 26264.62 30092.24
Oct 1994       29557.62 28066.89 31048.35 27277.74 31837.49
Nov 1994       27263.91 25742.16 28785.66 24936.60 29591.22
Dec 1994       26466.34 24859.23 28073.46 24008.48 28924.21
Jan 1995       26032.26 24333.30 27731.23 23433.92 28630.60
Feb 1995       24811.77 23087.10 26536.44 22174.11 27449.43
Mar 1995       22411.32 20763.80 24058.85 19891.65 24931.00
Apr 1995       23709.43 21876.44 25542.42 20906.12 26512.75

plot(F5)
lines(yt, col = "red", lwd = 2)
```

Promedios móviles (MA) V



Observaciones a los promedios móviles

- Mientras más grande es el orden k de un promedio móvil, el efecto de suavizamiento es mayor.
- ¿Cómo se escoge k? Para comparar la calidad de diferentes métodos de suavizamiento se requiere una medida de la asociación entre y_t y el pronóstico F_t.
- A continuación veremos varias medidas de calidad de un pronóstico.
- Este método es apropiado para series que tienen tendencia sin grandes variaciones.

Medidas de calidad para los pronósticos I

- Una vez que se obtienen los residuales, se pueden aplicar varias medidas para tener una forma de comparar entre métodos. Las medidas usuales son las que se describen a continuación. Se tienen $m = N - k$ residuales $e_1, e_2, \ldots, e_m$.
  1. Error medio (ME): $\bar{e} = \frac{\sum_{i=1}^m e_i}{m}$.
  2. Error medio absoluto (MAE): $\frac{\sum_{i=1}^m |e_i|}{m}$.
  3. Error cuadrático medio (MSE): $\frac{\sum_{i=1}^m e_i^2}{m}$.
  4. Error porcentual medio (MPE): $\frac{1}{m} \sum_{i=1}^m (100 \times e_i / Y_i)$.
  5. Error porcentual medio absoluto (MAPE): $\frac{1}{m} \sum_{i=1}^m 100 \times |e_i / Y_i|$.
- Podemos usar la función `accuracy` para obtener todas las medidas calculadas:

```
accuracy(F5)

                    ME     RMSE      MAE         MPE     MAPE      MASE       ACF1
Training set -16.36168 545.6951 420.4824 -0.09829481 1.653005 0.3738074 0.04089187
```

Suavizamiento exponencial simple I

- Hay situaciones en donde las observaciones más recientes contienen información más actualizada sobre lo que se espera en el futuro y por lo tanto, se les tiene que dar un peso mayor que a las observaciones más antiguas.
- El método de suavización exponencial calcula el pronóstico como una combinación entre el ponóstico del periodo anterior y la última observación disponible:

$$F_t(\alpha) = \alpha y_{t-1} + (1-\alpha)F_{t-1}$$

donde al parámetro α se le llama factor de suavizamiento, $\alpha \in (0,1)$ es un peso que pondera la relación entre ambos términos.

Suavizamiento exponencial simple II

- Haciendo una sustitución recursiva, podemos ver que:

$$
\begin{aligned}
F_{t+1} &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots \\
&= \alpha \sum_{i=0}^{\infty} (1-\alpha)^i y_{t-i}
\end{aligned}
$$

así que el suavizamiento exponencial obtiene su nombre de esta expresión.

- El pronóstico también se puede escribir como

$$F_{t+1} = F_t + \alpha(y_t - F_t) = F_t + e_t$$

así que un nuevo pronóstico es un ajuste del pronóstico anterior, agregando el error.

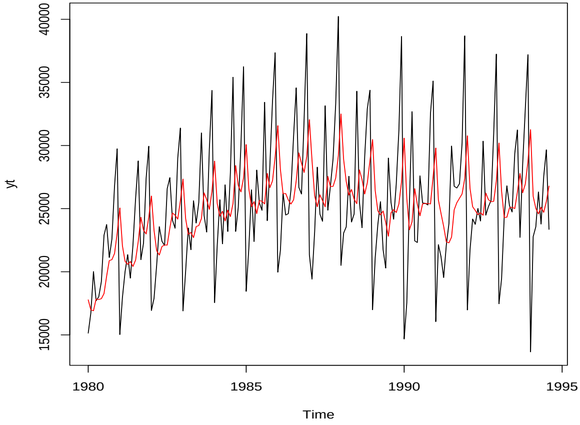
- Si $\alpha \to 0$, casi no hay ajuste y el pronóstico tiende a estar muy suavizado. Si $\alpha \to 1$, la última observación se vuelve más importante, y casi no hay suavizamiento.
- Se puede buscar un valor óptimo para α minimizando alguna de las funciones de error, como el error cuadrático medio, que definiremos a continuación.

Suavizamiento exponencial simple III

- En la práctica el suavizamiento exponencial es muy usado, ya que es un método preciso, efectivo y práctico.
- Para suavizamiento exponencial, podemos usar la función `ses` del paquete `forecast`

```
se1 <- ses(yt, alpha=0.3)
plot(yt)
lines(se1$fitted,col="red")  # valores ajustados
```

Suavizamiento exponencial simple IV

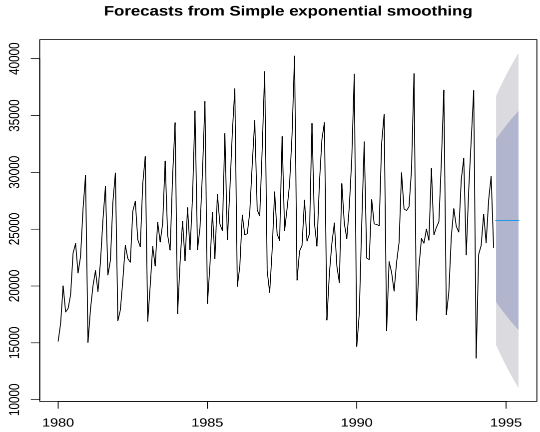


Suavizamiento exponencial simple V

- El pronóstico utilizando suavizamiento exponencial

```r
plot(se1)
```

Suavizamiento exponencial simple VI



Método lineal de Holt

- El modelo de Holt (1957) es una extensión del modelo de suavizamiento exponencial para datos con tendencia.
- En este método se plantean tres ecuaciones: una ecuación para el nivel, una para la pendiente (nivel y pendiente hacen la tendencia-ciclo) y otra de pronóstico:

$$
\begin{aligned}
L_t &= \alpha y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \\
b_t &= \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \\
F_{t+m} &= L_t + b_t m
\end{aligned}
$$

  donde L_t representa el nivel de la serie al tiempo t, y b_t es la pendiente de la tendencia. El parámetro α pondera el nivel estimado en el punto anterior y el nivel de la observación actual, mientras que β pondera a la pendiente.
- Se requiere inicializar L_t y b_t. En cada caso respectivamente, se toma $L_1 = y_1$ y $b_1 = y_2 - y_1$.

Ejemplo I

- Considerando $\alpha = 0.5$ y $\beta = 0.1$, y con los datos que se muestran a continuación, se puede construir un cuadro como el que se muestra a continuación.

t	y_t	L_t	b_t
1	200	200	-65
2	135	135	-65
3	195	132.5	-58.75
4	197.5	135.6	-52.56
5	310	196.5	-41.21

- Para cada tiempo se puede calcular una serie de pronósticos para un horizonte dado. Por ejemplo, en el primer periodo, la serie de valores pronosticados es:

$$F_1 = L1 = 200, \quad F_2 = L_1 + b_1 = 135, \quad F_3 = 70, \quad F_4 = 5, \quad F_5 = -60$$

para el segundo periodo, la serie de valores es:

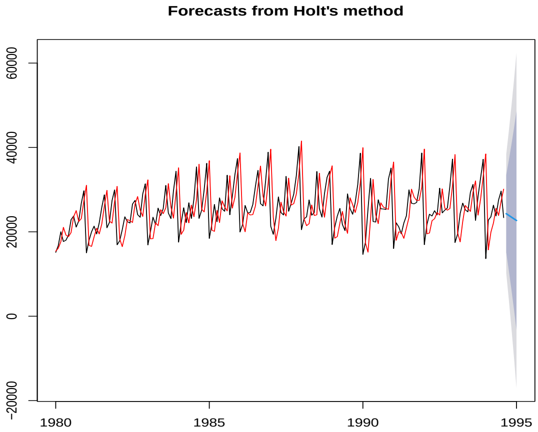
$$F_2 = L_2 = 135, \quad F_3 = L_2 + b_2 = 65, \quad F_4 = 5 \quad \ldots$$

Ejemplo II

Con una serie larga y una buena selección de parámetros, el modelo elegido mejora la calidad de los pronósticos.

```
h1 <- holt(yt, alpha=0.8,beta=0.2, h=5)
plot(h1)
lines(h1$fitted, col = "red")
```

Ejemplo III



Forecasts from Holt's method
60000
40000
20000
0
−20000
1980
1985
1990
1995

Selección de los parámetros

- Para escoger los valores más convenientes de α y β, se selecciona una función adecuada de los errores, como el error cuadrático medio, y se considera como función de los parámetros: $MSE(\alpha, \beta)$.
- Los mejores parámetros son los que minimizan tal función.
- Cuando $\alpha = \beta$, el método se conoce como *Suavizamiento exponencial doble de Brown*.

Método de Holt-Winters I

- Los modelos de suavizamiento exponencial y el modelo de Holt son típicamente adecuados para datos sin estacionalidad.
- El método de Holt-Winters está diseñado para datos estacionales.
- Winters (1960) extendió el método de Holt, y hay dos versiones dependiendo de cómo se modele la estacionalidad (aditiva o multiplicativa).
- En este caso se tienen 4 ecuaciones y 3 parámetros:

$$\begin{aligned}
\text{Nivel } L_t &= \alpha \frac{y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}) \\
\text{Tendencia } b_t &= \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \\
\text{Estacional } S_t &= \gamma \frac{y_t}{L_t} + (1-\gamma)S_{t-s} \\
\text{Pronóstico } F_{t+m} &= (L_t + b_t m)S_{t-s+m}
\end{aligned}$$

donde s es la longitud de la estacionalidad.

Método de Holt-Winters II

- Inicialización: Se necesitan valores iniciales para L_t, b_t y los índices estacionales S_t.

  S_t: requiere los datos completos de una estación, y $S_i = \frac{y_i}{L_s}$.

  L_t: Se calcula el valor para s: $L_s = \frac{1}{s}(y_1 + \cdots + y_s)$.

  b_t: Se calcula el valor para s: $b_s = \frac{1}{s}(\frac{y_{s+1}-y_1}{s} + \cdots + \frac{y_{2s}-y_s}{s})$.

- Los valores de α, β y γ se eligen de tal forma que minimizen $MSE(\alpha, \beta, \gamma)$.
- En `R`, la función `HoltWinters` permite definir un ajuste considerando un modelo aditivo o multiplicativo, e inlcuye como casos particulares el modelo de Holt ($\gamma = 0$) y el modelo de suavizamiento exponencial (con $\gamma = 0$, $\beta = 0$). Aquí `R` encuentra los valores óptimos si no son dados.

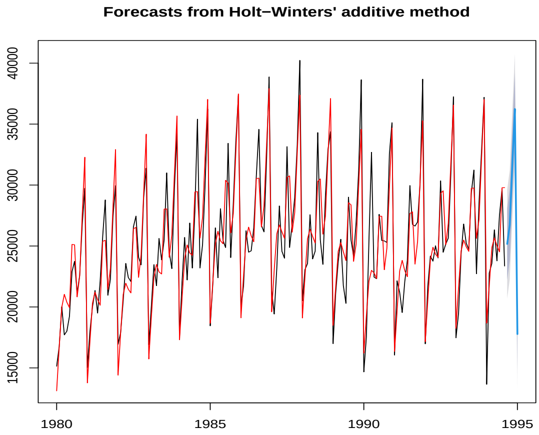
Ejemplos de Holt-Winters

```
(m <- hw(yt,h=5)) # aditivo por default, elige valores óptimos opr default

         Point Forecast     Lo 80    Hi 80    Lo 95    Hi 95
Sep 1994       25158.19 22234.62 28081.75 20686.98 29629.39
Oct 1994       26727.87 23799.28 29656.47 22248.97 31206.78
Nov 1994       31594.46 28658.18 34530.74 27103.81 36085.11
Dec 1994       36249.81 33302.67 39196.96 31742.54 40757.08
Jan 1995       17765.61 14803.89 20727.33 13236.05 22295.18
```

Ejemplos de Holt-Winters I

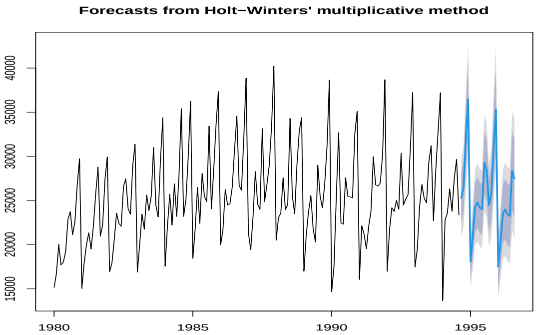
```
plot(m)
lines(m$fitted,col="red")
```

Ejemplos de Holt-Winters II



Ejemplos de Holt-Winters III

```r
# Ahora consideramos un modelo multiplicativo
m <- hw(yt, seasonal = "mult")
plot(m)
```



1 Análisis de series de tiempo
- Introducción a las series de tiempo
- Enfoque descriptivo (Descomposición de series temporales)
- Suavizamiento
- Suavizamiento Avanzado

2 Modelos ARIMA
- Metodología de Box-Jenkins
- Análisis Box-Jenkins
- Ruido blanco
- Modelos AR, MA y ARMA

3 Anexo

Otros comentarios sobre la estacionalidad I

- Hasta ahora hemos visto que una serie de tiempo puede descomponerse en:
  1. Tendencia T_t, que es el movimiento suave que indica la dirección de movimiento de la serie en el largo plazo
  2. Ciclo C_t, que indica el movimiento en forma de onda, sobrepuesto a la tendencia y cuya duración típicamente es mayor a un año.
  3. la Estacionalidad S_t, que es un movimiento periódico que se da en un año y (iv) el Error E_t que representa la parte irregular y aleatoria de la serie.
- El interés de descomponer una serie de tiempo desde el punto de vista económico es tratar de aislar cada componente y poder interpretar y predecir su comportamiento de manera separada. Usualmente se considera que las posibles causas que generan la estacionalidad, por ejemplo, no son de naturaleza económica.
- Granger identificó cuatro posibles causas de las fluctuaciones estacionales:
  - Calendario: algunas fiestas sociales se celebran en fechas específicas, los meses tienen diferente número de días.
  - Actividades fijadas por instituciones: pago de impuestos, periodos vacacionales, etc.
  - Clima: las estaciones del año determinan las cosechas, la cobertura de ciertos riesgos, etc.
  - Expectativas de fluctuaciones estacionales: la Navidad genera una expectativa de venta de juguetes, por ejemplo.

Otros comentarios sobre la estacionalidad II

- Es por las razones previas que desde un punto de vista económico, es conveniente desestacionalizar las series o ajustarlas estacionalmente.

Ejemplos de promedios móviles ponderados $WMA(k)$

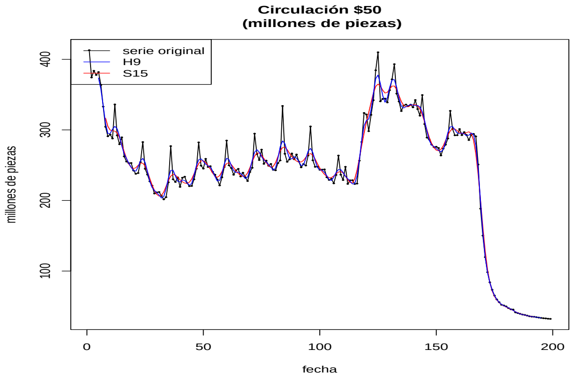
Algunos promedios móviles ponderados especiales han sido de utilidad para separar el componente de tendencia-ciclo:

- **filtro de Spencer** (1904) propuso un $5 \times 4 \times 4MA$ seguido de un promedio móvil ponderado con pesos $a_{-2} = a_2 = -\frac{3}{4}, a_0 = 1, a_{-1} = a_1 = \frac{3}{4}$. Esta combinación es equivalente al promedio $WMA(15)$ con pesos $\frac{1}{320}(-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3)$. Este filtro se conoce como $S15$.
- $3 \times 5MA$ es equivalente a un $WMA(4)$ con pesos $(0.200, 0.200, 0.133, 0.067)$.
- **filtro de Henderson** (1914), $H9 = (0.33, 0.267, 0.119, -0.010, -0.041)$.

La función `filter` en `R` permite suavizar una serie de tiempo calculando el promedio de las observaciones con los pesos que se le dan como argumento.

Ejemplo: Ajustes de Promedios ponderados especiales

```
den <- read.csv(paste0(ruta,"denom.csv"))
plot.ts(den$x50, ylab = "millones de piezas", xlab = "fecha")
points(den$x50,pch=16,cex=0.4)
S15 <- c(-3,-6,-5,3,21,46,67,74,67,46,21,3,-5,-6,-3)/320
H9 <- c(-0.04072, -0.00987,  0.11847,  0.26656,  0.33114,  0.26656, 0.11847, -0.00987, -0.04072)
lines(filter(den$x50,filter=S15),col="red")
lines(filter(den$x50,filter=H9),col="blue")
title("Circulación $50\n (millones de piezas)")
legend("topleft",legend=c("serie original","H9","S15"),col=c("black","blue","red"),lty=c(1,1))
```



Suavizamiento de regresión local

- Este modelo ajusta rectas a secciones de los datos. La tendencia en t está dada por $T_t = a_t + b_t t$ donde

$$(a_t, b_t) = arg_{min} \left\{ \sum_{j=-m}^{m} a_j (Y_{t+j} - a - b(t+j))^2 \right\}$$

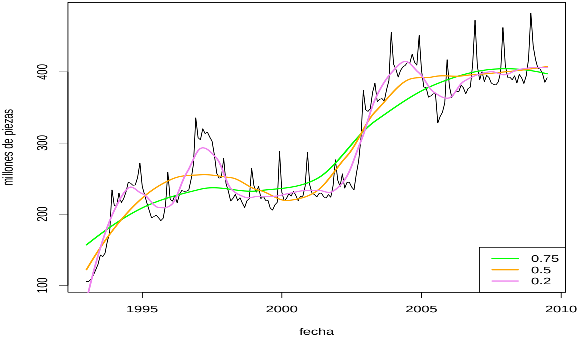
  los pesos a_j determinan el peso de cada observación, y $m = \frac{k-1}{2}$
- En cada punto t se ajusta un valor. En los extremos se requiere hacer modificaciones para evitar sesgos.
- En este caso no mostraré un ejemplo porque la programación es tediosa y este es un paso intermedio para llegar a un mejor estimador, que es *loess*, como veremos adelante.

loess (*local estimate scatterplot smoothing*)

- Es una implementación del suavizamiento de regresión local que es robusto ante valores extremos (outliers), que se obtiene de realizar varias iteraciones:
    i. Calcula una regresión local $\hat{T}_t$
    ii. Estima el error de ajuste $\hat{E}_t = Y_t - \hat{T}_t$
    iii. Ajusta los pesos a_j de tal manera que las observaciones con errores grandes E_t, reciban un peso menor y tengan menos influencia sobre el ajuste.
    iv. Repetir el paso (i) y los que siguen hasta convergencia.

Ejemplo de ajuste loess

```
z <- loess(den$x100 ~ time(den$x100))
plot.ts(ts(den$x100,start = c(1993,01), freq=12), ylab = "millones de piezas", xlab = "fecha")
lines(ts(z$fit, start = c(1993,01), freq = 12), col = "green", lwd = 2)
lines(ts(loess(den$x100 ~ time(den$x100), span = 0.5)$fit, start = c(1993, 01), freq = 12), col = "orange", lwd = 2)
lines(ts(loess(den$x100 ~ time(den$x100), span = 0.2)$fit, start = c(1993, 01), freq = 12), col = "violet", lwd = 2)
legend("bottomright", legend = c(0.75, 0.5, 0.2), col = c("green", "orange", "violet"), lty = c(1,1,1),lwd=c(2,2,2))
```



Métodos del Bureau of Census de los EU

- Como refinación a los métodos presentados previamente, y considerando que en estructuras más complejas la descomposición es más compleja, surge un procedimiento en donde se aplican de manera iterativa promedios móviles ponderados.
- El método Census II surgió en 1955 y ha tenido varias variantes experimentales, llamadas X-1 a X-13. Estos métodos siguen siendo utilizados por varias agencias experimentales en todo el mundo.
- El sitio http://www.seasonal.website presenta las facilidades completas para justas series de tiempo con este método.

Modelos ARIMA

1 Análisis de series de tiempo
- Introducción a las series de tiempo
- Enfoque descriptivo (Descomposición de series temporales)
- Suavizamiento
- Suavizamiento Avanzado

2 Modelos ARIMA
- Metodología de Box-Jenkins
- Análisis Box-Jenkins
- Ruido blanco
- Modelos AR, MA y ARMA

3 Anexo

- Como se ha mencionado, hay dos aspectos a estudiar en las series de tiempo: uno es el análisis, y el otro es la modelación.
- Los modelos más comunes de series de tiempo son conocidos como **modelos ARIMA**, que significa: *Autoregresive, Integrated, Moving Average*. Estos modelos describen *series estacionarias*.
- Estos modelos fueron propuestos en los 70's por George Box y Gwilym Jenkins y son esenciales para entender otros modelos mucho más complicados que se utilizan en la práctica, como los modelos ARCH y GARCH, los modelos de espacio de estados, los modelos de cadenas de Markov ocultos, los modelos dinámicos lineales, etc.
- Robert F. Engle y Clive Granger obtuvieron el premio Nobel 2004 "for methods of analyzing economic time series with time-varying volatility (ARCH)" y "for methods of analyzing economic time series with common trends (cointegration)" respectivamente. Este comentario es sólo para enfatizar la importancia de los modelos ARIMA.

Metodología de Box y Jenkins

- Lo dividiremos en 4 fases:
  I. *Identificación:* preparación de los datos y selección del modelo. Utiliza los datos históricos para identificar un modelo apropiado.
  II. *Estimación:* estima los parámetros de modelos potenciales y selecciona el mejor modelo.
  III. *Diagnósticos:* se utilizan varios diagnósticos para evaluar si el modelo es adecuado.
  IV. *Aplicación:* se usa el modelo para hacer pronósticos y otros estudios.
- Las primeras dos fases pueden repetirse si los diagnósticos indican que los modelos no son adecuados.
- En la sesión de hoy, seguiremos el desarrollo de estos pasos a través de ejemplos, y se irán introduciendo los conceptos necesarios conforme se utilicen.

1. Análisis de series de tiempo
   - Introducción a las series de tiempo
   - Enfoque descriptivo (Descomposición de series temporales)
   - Suavizamiento
   - Suavizamiento Avanzado
2. Modelos ARIMA
   - Metodología de Box-Jenkins
   - Análisis Box-Jenkins
   - Ruido blanco
   - Modelos AR, MA y ARMA
3. Anexo

Lista de herramientas

Aquí agrego un resumen de las herramientas que son necesarias para hacer un análisis de series de tiempo tipo Box-Jenkins. Algunas son necesarias para construir el modelo y otras son necesarias para diagnosticar el modelo.

1. Gráficas de las series de tiempo.
2. Concepto de estacionariedad, operador diferencia y prueba de estacionariedad de Dickey-Fuller.
3. Conceptos de autocorrelación y autocorrelación parcial y sus respectivas funciones y gráficas (`acf`, `pacf`)
4. Concepto y gráficas de ruido blanco.
5. Pruebas estadísticas para coeficientes de autocorrelación de Bartlett, y de Ljung-Box.
6. Definiciones de procesos: autoregresivo (AR), de promedios móviles (MA).

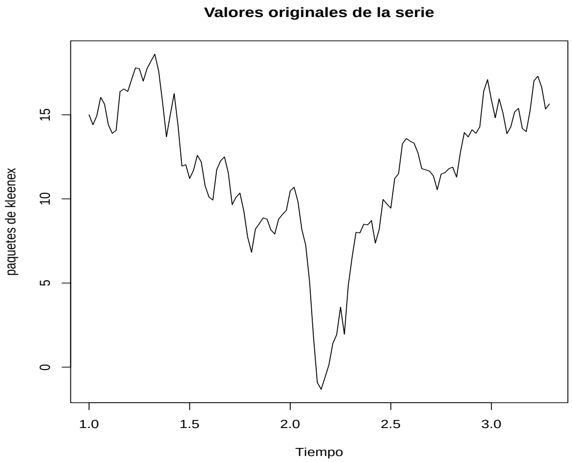
En lo que sigue, se dará un repaso muy general de estos conceptos, y veremos varios ejemplos para ir aplicando los conceptos.

Ejemplo: Kleenex I

Para conocer el margen de mercado que tiene Kimberly-Clark con la marca kleenex, se le han pedido su producción semanal, que reporta para las 120 semanas anteriores a ésta y en unidades de 10,000 paquetes. Se requiere un modelo para analizar su comportamiento y hacer un pronóstico para saber si no está inundando el mercado.

```r
kleenex <- read.table(file = paste0(ruta,"kleenex.txt"))   # lee los datos
kleenex <- ts(kleenex, start = 1, freq = 52)  # convierte a una serie con frecuencia semanal.
plot(kleenex, main = "Valores originales de la serie", xlab = "Tiempo", ylab = "paquetes de kleenex")
```

Ejemplo: Kleenex II



La primera pregunta que nos debemos hacer es si esta serie tiene un comportamiento *estacionario*

Series estacionarias I

- Intuitivamente, una serie es estacionaria si ciertas propiedades estadísticas, por ejemplo la media y la varianza, son esencialmente constantes en el tiempo.
- Si una gráfica de la serie de n puntos no muestra que los valores $y_1, y_2, \ldots, y_n$ sean no estacionarios, a veces se puede transformar la serie no estacionaria en estacionaria *diferenciando la serie*:

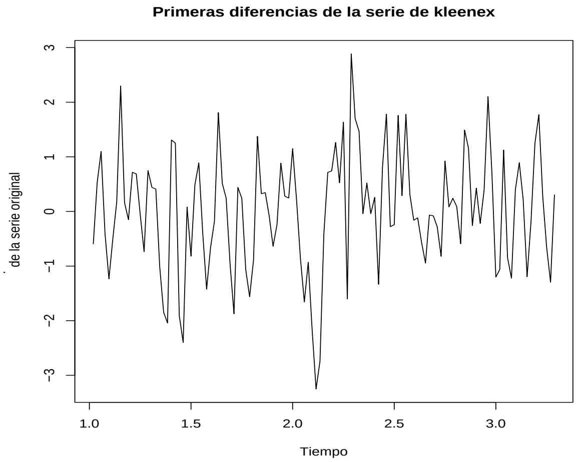
$$z_t = y_t - y_{t-1} \quad \text{para} \quad t = 2, \ldots, n$$

```
dkleenex <- diff(kleenex, differences = 1)
head(dkleenex)
```

```
Time Series:
Start = c(1, 2)
End = c(1, 7)
Frequency = 52
          V1
[1,] -0.5936
[2,]  0.5319
[3,]  1.0991
[4,] -0.4054
[5,] -1.2345
[6,] -0.5016
```

```
plot(dkleenex, main = "Primeras diferencias de la serie de kleenex", xlab = "Tiempo",
ylab = "primeras diferencias\n de la serie original")
```

Series estacionarias II



Series estacionarias III

- Ahora la serie parece fluctuar alrededor de una media constante, por lo que la primera diferencia parece ser estacionaria.
- En ocasiones no es suficiente tomar sólo una diferencia. Si la primera diferencia no resulta estacionaria, entonces debemos tomar una segunda diferencia. Esto es tan común que conviene introducir un poco de notación.
- También hay que identificar observaciones inusuales (*outliers*) y Decidir si se requiere de una transformación de la serie de tiempo para estabilizar la varianza. Si se requiere, transforma la serie para hacer que la varianza parezca más constante (aplicando las transformaciones usuales, como obtener logaritmos, raíces cuadradadas, Box-Cox, etc.)

Operadores rezago y diferencia I

- Definan al operador rezago como la función L (de *lag*) tal que $La_t = a_{t-1}$
- En general, se puede tener un rezago de orden j: $L^j a_t = a_{t-j}$. Por convención, se toma $L^0 = 1$
- De la misma forma, podemos definir el operador diferencia como $\Delta a_t = a_t - a_{t-1}$
- De hecho, notemos que ambos operadores, rezago y diferencia, están relacionados a través de la fórmula:

$$\Delta = 1 - L$$

  ya que $(1-L)a_t = a_t - La_t = a_t - a_{t-1} = \Delta a_t$.
- Podemos tomar también la segunda diferencia, la tercera diferencia, etc. de la siguente manera:

1a. dif $\Delta y_t = y_t - y_{t-1}$

2a. dif $\Delta^2 y_t = \Delta(\Delta y_t) = \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2}$.

3a. dif $\Delta^3 y_t = \Delta(\Delta^2 y_t) = y_t - 2y_{t-1} + y_{t-2} - (y_{t-1} - 2y_{t-2} + y_{t-3}) = y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3}$.

- En la práctica, si la serie original $\{y_t\}$ *no es estacionaria* y *no tiene un patrón estacional* entonces tomando la primera diferencia Δy_t o la segunda diferencia $\Delta^2 y_t$ usualmente hará la serie estacionaria.
- Sin embargo, si la serie original *no es estacionaria* y *tiene un patrón estacional*, entonces requerirá transformaciones más complejas, utilizando el operador de diferencia estacional.

Ejemplo I

```
x <- ts(1:5)
cbind(x,lag(x, 1), lag(x, -1))

Time Series:
Start = 0
End = 6
Frequency = 1
   x lag(x, 1) lag(x, -1)
0 NA         1         NA
1  1         2         NA
2  2         3          1
3  3         4          2
4  4         5          3
5  5        NA          4
6 NA        NA          5
```

Noten que el valor por omisión de `lag(x)` es `lag(x,1)`. Así que en `R` si tiene una serie x_t, entonces la serie $y_t = lag(x_t) = lag(x_t, 1) = x_{t+1}$, que es un valor adelantado y no x_{t-1}, así que para obtener Lx_t debemos usar `lag(x, -1)`.

Por otra parte, para la diferencias:

Ejemplo II

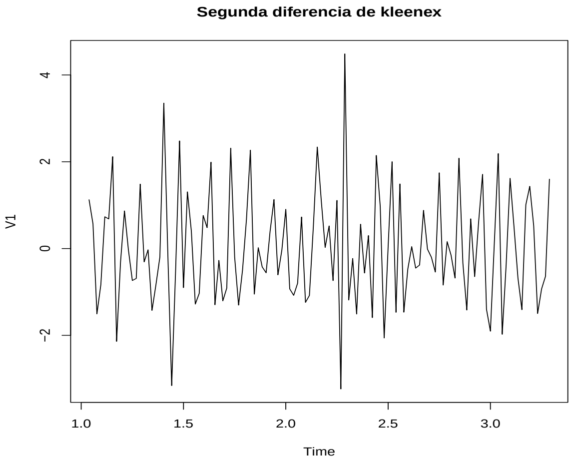
```
x <- ts(seq(2,10,by=2))
x

Time Series:
Start = 1
End = 5
Frequency = 1
[1]  2  4  6  8 10

cbind(x,diff(x,1), diff(x,2), x-lag(x,-2))

Time Series:
Start = 1
End = 5
Frequency = 1
   x diff(x, 1) diff(x, 2) x - lag(x, -2)
1  2         NA         NA             NA
2  4          2         NA             NA
3  6          2          4              4
4  8          2          4              4
5 10          2          4              4
```

El operador diferencia sí nos da los valores esperados $1-L, (1-L)^2$.

Ejemplo de segunda diferencia I

```
d2kleenex <- diff(kleenex, difference = 2)
head(d2kleenex)

Time Series:
Start = c(1, 3)
End = c(1, 8)
Frequency = 52
          V1
[1,]  1.1255
[2,]  0.5672
[3,] -1.5045
[4,] -0.8291
[5,]  0.7329
[6,]  0.6822

plot(d2kleenex, main = "Segunda diferencia de kleenex")
```

Ejemplo de segunda diferencia II



Identificando estacionariedad I

- Necesitamos un mecanismo más sólido para verificar si una serie es estacionaria. Lo haremos a través de las funciones acf y pacf.
- También se pueden considerar las pruebas más formales de Dickey-Fuller o de Phillips-Perron, que para definirlas se requiere un poco más de infraestructura. No las consideraremos aquí, pero en `R`, son las funciones `adf.test` y `pp.test` en el paquete `tseries`.
- La *función de autocorrelación* (acf). Un coeficiente de autocorrelación r_k describe la relación entre valores de la serie de tiempo que están separados por k periodos.

función de autocorrelación

- El coeficiente de autocorrelación se define como:

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

  donde $\hat{\gamma}_k$ es la covarianza estimada entre y_t y y_{t-k}, y $\hat{\gamma}_0$ es la varianza estimada de la serie.

- r_1 indica cómo se relacionan valores sucesivos de y_t entre sí; r_2 indica cómo se relacionan valores y_t que están separados por dos periodos, etc.
- La gráfica de r_k vs. k se conoce como *correlograma*, y es una gráfica de la función de autocorrelación.

Identificando estacionariedad III

- Otra herramienta importante es la *función de autocorrelación parcial*.

función de autocorrelación parcial

- Las *autocorrelaciones parciales* se usan para medir el grado de asociación entre y_t y y_{t-k}, cuando se ha eliminado el efecto de los rezagos intermedios $1, 2, 3, \ldots, k-1$.
- El coeficiente de autocorrelación parcial de orden k se denota por α_k y se calcula haciendo la regresión de y_t contra los rezagos $y_{t-1}, \ldots, y_{t-k}$:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_k y_{t-k}$$

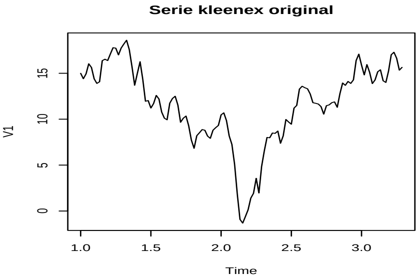
- La función de autocorrelación parcial es la grafica de α_k vs. k.

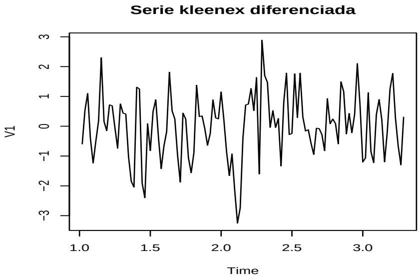
- Para la serie de Kimberly, por ejemplo, podemos ver los dos casos, con la serie original y la serie diferenciada.

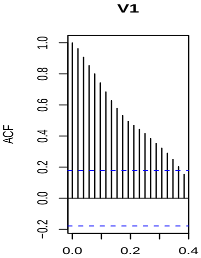
Identificando estacionariedad IV

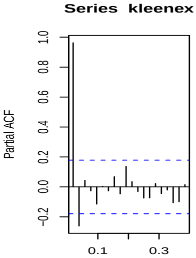
```
layout(matrix(c(1,1,2,2,3,4,5,6),nrow=2,byrow=T))
plot(kleenex, main = "Serie kleenex original")
plot(dkleenex, main = "Serie kleenex diferenciada")
acf(kleenex)
pacf(kleenex)
acf(dkleenex)
pacf(dkleenex)
```

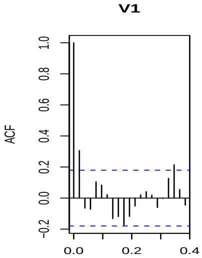
Identificando estacionariedad V

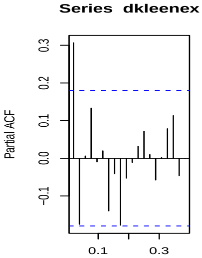












Interpretación del correlograma I

- El correlograma sirve para explorar una serie de tiempo: verificar si la serie es estacionaria, si tiene estacionalidades, ciclos,tendencias, etc.
- Es importante validar si las autocorrelaciones son *estadísticamente significativas*, ya que cualquier $\hat{\rho}_k$ de la muestra puede diferir de cero sólo por variación en la muestra.
- Podemos darnos una idea de la importancia de la estadística muestral comparándola contra su error estándar. Un error estándar aproximado para $\hat{\rho}_k$ se obtiene de la siguiente fórmula:

$$se(\hat{\rho}(k)) = \sqrt{\frac{1 + 2\sum_{j=1}^{k-1} \hat{\rho}_j^2}{n}}$$

Esta fórmula se conoce como la *fórmula de Barttlet*.

- En las gráficas de los correlogramas, se incluyen lineas azules punteadas que identifican un intervalo de 95 % para las autocorrelaciones. Si la serie de tiempo tiene *observaciones normales*, entonces también las autocorrelaciones son normales. En este caso, para probar que $\rho_k = 0$ se usa la estadística $\sqrt{T}\hat{\rho}_k$, donde T es la longitud de la serie. Se puede ver que $\sqrt{T}\hat{\rho}_k \sim \mathcal{N}(0,1)$. Los intervalos que se muestran en el correlograma se calculan con esta estadística.
- No siempre hay que confiar en estas líneas porque son aproximadas. Se sugiere hacer una prueba formal. Las opciones son las pruebas de Box-Pierce o Ljung-Box.
- Estas estadísticas prueban la hipótesis de que las observaciones son independientes hasta el rezago h:

$$H_0 : \rho_1 = \cdots = \rho_h = 0$$

Pruebas para autocorrelaciones II

```
Box.test(dkleenex, lag = 1, type = "Box") # h= 1 en este ejemplo

Box-Pierce test

data:  dkleenex
X-squared = 11.189, df = 1, p-value = 0.000823
Box.test(dkleenex, lag = 1, type = "Ljung")

Box-Ljung test

data:  dkleenex
X-squared = 11.473, df = 1, p-value = 0.0007061
```

- La prueba de Ljung-Box es simplemente una corrección a la prueba de Box-Pierce que es más robusta bajo series de tiempo cortas. Ambas se distribuyen como χ^2.
- La fórmula de Bartlett es la correcta.
- Usualmente se aplica esta prueba a los residuales de un modelo ajustado.

Interpretación del correlograma III

- Si la `acf` de una serie de tiempo decae relativamente rápido, sin importar el patrón que siga, entonces la serie de tiempo se considera estacionaria
- Si la `acf` no decae lentamente, entonces debe considerarse como no estacionaria.

En nuestro ejemplo, de acuerdo al criterio establecido, podemos ver que la serie original $y_t =$ `kleenex` no es estacionaria, pero la serie diferenciada $\Delta y_t =$ `dkleenex` si puede considerarse estacionaria.

1 Análisis de series de tiempo
- Introducción a las series de tiempo
- Enfoque descriptivo (Descomposición de series temporales)
- Suavizamiento
- Suavizamiento Avanzado

2 Modelos ARIMA
- Metodología de Box-Jenkins
- Análisis Box-Jenkins
- Ruido blanco
- Modelos AR, MA y ARMA

3 Anexo

Ejemplo: modelo de ruido blanco I

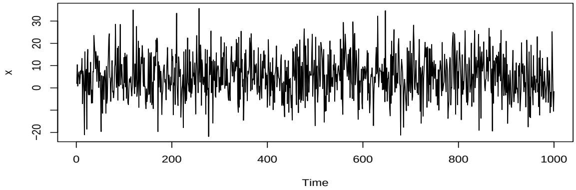
- Un modelo en el que cada observación se compone de dos partes, un nivel constante c y un componente de error ϵ_t que es independiente de cualquier periodo,

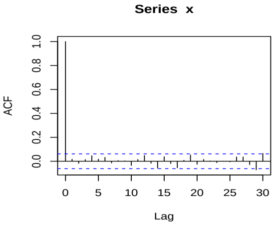
$$y_t = c + \epsilon_t$$

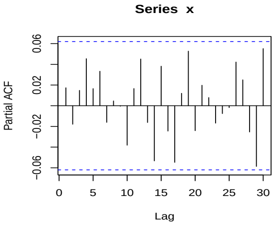
se le llama *ruido blanco*.
- Este modelo puede ser *simulado* usando un generador de números aleatorios normales, como la función `rnorm` en R.

```
x <- rnorm(1000,mean=5, sd=10)
layout(matrix(c(1,1,2,3),nrow=2,byrow=T))
plot.ts(x)
acf(x)
pacf(x)
```

Ejemplo: modelo de ruido blanco II







Ejemplo: modelo de ruido blanco III

- Usualmente después de ajustar un modelo, los residuales deberían comportarse similar a ruido blanco para considerar que el modelo ha sido adecuadamente ajustado.

1 Análisis de series de tiempo
- Introducción a las series de tiempo
- Enfoque descriptivo (Descomposición de series temporales)
- Suavizamiento
- Suavizamiento Avanzado

2 Modelos ARIMA
- Metodología de Box-Jenkins
- Análisis Box-Jenkins
- Ruido blanco
- Modelos AR, MA y ARMA

3 Anexo

Modelos AR

Un modelo autorregresivo de orden p, que se denota por $AR(p)$ es un modelo de la forma:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

donde $\epsilon_t \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$, y los errores son independientes.

- En este modelo los predictores de la observación y_t son sus propios rezagos en el tiempo.
- Los modelos AR pueden ser estacionarios o no estacionarios, dependiendo de las restricciones que se impongan sobre los pesos del modelo.

Modelos MA

Modelos MA

Un modelo de promedios móviles de orden q es un modelo de la forma:

$$y_t = \theta_0 + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

donde $\{\epsilon_t\}$ es una serie de ruido blanco.

- A este proceso se le llama promedio móvil porque es una especie de promedio móvil en la serie $\{\epsilon_t\}$, con pesos dados por los coeficientes $\theta_1, \theta_2, \ldots, \theta_q$.

Operador rezago

- Los modelos AR y MA pueden simplificarse usando los operadores rezago y diferencia, ya que podemos escribirlos como *polinomios* en L.
- Un polinomio en L de orden k es de la forma

$$\Phi_k(L) = \phi_0 L^0 + \phi_1 L^1 + \cdots \phi_k L^k,$$

donde $L^0 = 1$.

Simplificando notación

- Un $MA(q)$, se puede escribir como
$$y_t = \mu + \Theta_p(L)\epsilon_t$$
donde $\Theta_p(L) = \theta_0 L^0 + \theta_1 L^1 + \theta_2 L^2 + \cdots + \theta_q L^q$ y $\theta_0 = 1$.
- Un $AR(p)$ se puede escribir como
$$\Phi_p(L)y_t = \alpha_0 + \epsilon_t$$
donde $\Phi_p(L) = 1 - \phi_1 L^1 - \phi_2 L^2 - \cdots - \phi_p L^p$.

Modelos ARMA y ARIMA

- Los modelos ARMA combinan modelos autorregresivos con modelos de promedios móviles para llegar a modelos más generales.
- Un modelo ARMA(p, q) se escribe, usando notación de polinomios, como

$$\Phi_p(L)y_t = \mu + \Psi_q(L)\epsilon_t$$

- Los modelos ARMA sólo se pueden aplicar a series estacionarias. Para extender los modelos a series no estacionarias, se requiere diferenciar la serie. Esto da origen a los modelos ARIMA.
- Un modelo ARIMA(p, d, q) es un modelo ARMA(p, q) en donde la serie original y_t se reemplaza por la serie diferenciada $z_t = \Delta^d y_t$.
- En la práctica, d usualmente toma valores en $\{0, 1, 2\}$ y p y q toman valores no mayores a 4.
- Los modelos AR y MA son casos particulares de modelos ARIMA, ya que por ejemplo, un AR(p) es lo mismo que un ARIMA($p, 0, 0$), o un MA(q) es un ARIMA($0, 0, q$).

Estimación

- Dada la falta de tiempo para ver en detalle las características de las funciones acf y pacf para identificar modelos, consideraremos un enfoque de *fuerza bruta* para ajustar los modelos y utilizar el criterio de Akaike para seleccionar el modelo.
- Podemos considerar ajustar todos los modelos para las combinaciones de valores p, d, q en los conjuntos $p \in \{0, 1, 2, 3, 4\}$, $d \in \{0, 1, 2\}$, y $q \in \{0, 1, 2, 3, 4\}$.
- Sin embargo, este método no funcionará para series con un fuerte componente estacional. En caso de contar con componente estacional se recomienda primero eliminar la estacionalidad con los métodos vistos previamente.
- También se pueden extender los modelos a modelos SARIMA(p, d, q)(P, D, Q), que contemplan un componente estacional, pero esos modelos son más complicados y no se verán aquí.

Ejemplo: ajuste por fuerza bruta para los datos de kleenex I

```
AIC <- array(NA, dim=c(4,2,4), dimnames = list(paste("p=",0:3), paste("d=",0:1), paste("q=",0:3)))
for(p in 1:4) for(q in 1:4)     for(d in 1:2) AIC[p,d,q] <- arima(kleenex, order = c(p-1,d-1,q-1))$aic
AIC[,1,]
```

```
         q= 0     q= 1     q= 2     q= 3
p= 0 698.9005 568.3708 475.5712 437.5709
p= 1 369.6550 356.3556 358.1165 358.6551
p= 2 358.3822 358.2153 357.4390 359.2192
p= 3 357.5850 359.5559 359.2680 360.9870
```

```
AIC[,2,]
```

```
         q= 0     q= 1     q= 2     q= 3
p= 0 362.2711 349.9821 351.9519 351.8757
p= 1 352.5789 351.9658 351.3134 352.6993
p= 2 350.7612 352.7601 353.5758 353.2524
p= 3 352.7582 352.4818 349.9126 350.9945
```

Ejemplo: ajuste por fuerza bruta para los datos de kleenex II

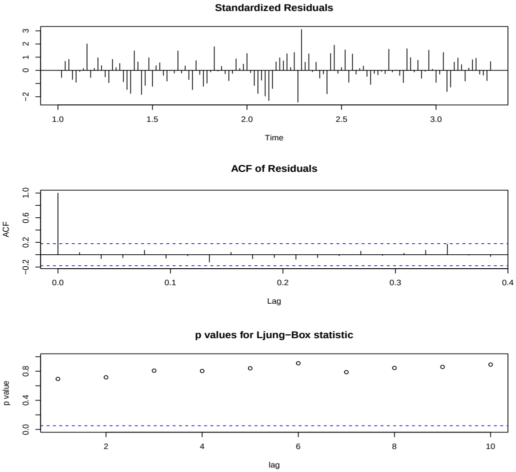
El modelo que mejor ajusta estos datos está dado por el caso con el *menor* AIC, que corresponde a los valores $p = 3$, $d = 1$ y $q = 2$, sugiriendo un ajuste de la forma ARIMA$(3, 1, 2)$. No es muy distinto al valor que se tiene para el modelo con $p = 0, d = 1, q = 1$ que también puede ser un buen candidato. Entonces los modelos que se pueden proponer para valorar en los diagnósticos son los modelos ARIMA$(3, 1, 2)$ y ARIMA$(0, 1, 1)$. Usualmente, si se tienen dos o más modelos que no difieren mucho en la calidad del ajuste, siempre se escoge el más sencillo. Este es el *principio de parsimonia*.

Ejemplo: ajuste por fuerza bruta para los datos de kleenex III

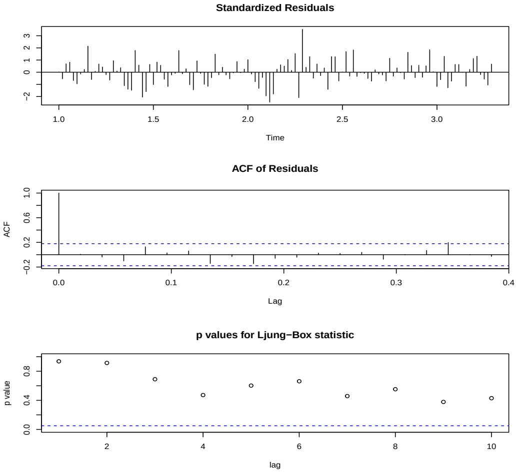
```
mod1 <- arima(kleenex,order = c(3,1,2)); mod1


Call:
arima(x = kleenex, order = c(3, 1, 2))

Coefficients:
         ar1      ar2     ar3      ma1     ma2
      1.5308  -1.3686  0.3847  -1.2517  0.9587
s.e.  0.0876   0.1047  0.0894   0.0463  0.0711

sigma^2 estimated as 0.9869:  log likelihood = -168.96,  aic = 349.91

mod2 <- arima(kleenex, order = c(0,1,1)); mod2


Call:
arima(x = kleenex, order = c(0, 1, 1))

Coefficients:
         ma1
      0.3518
s.e.  0.0800

sigma^2 estimated as 1.071:  log likelihood = -172.99,  aic = 349.98
```

Diagnósticos en series de tiempo

```r
tsdiag(mod1)
```



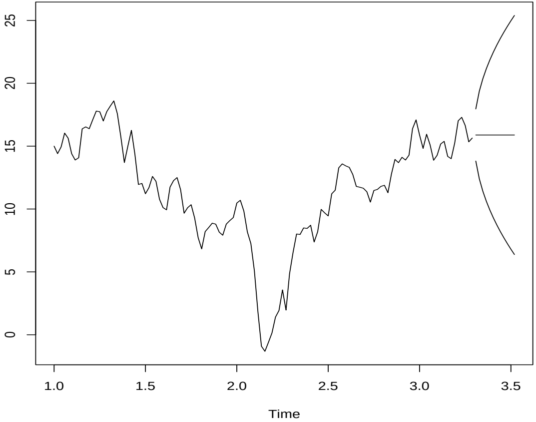

```r
tsdiag(mod2)
```



Pronósticos con el modelo parsimonioso I

```
pron2 <- predict(mod2,12)  # Haz una predicción de 12 periodos.
pron2

$pred
Time Series:
Start = c(3, 17)
End = c(3, 28)
Frequency = 52
 [1] 15.88725 15.88725 15.88725 15.88725 15.88725 15.88725 15.88725 15.88725 15.88725 15.88725 15.88725 15.88725

$se
Time Series:
Start = c(3, 17)
End = c(3, 28)
Frequency = 52
 [1] 1.034805 1.739997 2.232568 2.634602 2.982935 3.294643 3.579307 3.842943 4.089618 4.322239 4.542964 4.753450

ts.plot(kleenex, pron2$pred, pron2$pred +2*pron2$se,
pron2$pred -2*pron2$se)
```

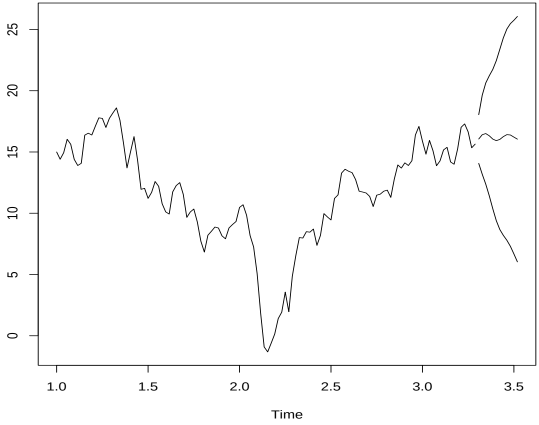
Pronósticos con el modelo parsimonioso II



Pronósticos con el modelo complicado I

```
pron1 <- predict(mod1,12)  # Haz una predicción de 12 periodos.
pron1

$pred
Time Series:
Start = c(3, 17)
End = c(3, 28)
Frequency = 52
 [1] 16.05671 16.40928 16.50297 16.32213 16.05270 15.92379 16.02564 16.25434 16.41544 16.38824 16.21409 16.04669

$se
Time Series:
Start = c(3, 17)
End = c(3, 28)
Frequency = 52
 [1] 0.9936002 1.6133586 2.0646103 2.4489185 2.8430614 3.2654087 3.6773496 4.0299156 4.3102502 4.5436417 4.7676635 5.0083769

ts.plot(kleenex, pron1$pred, pron1$pred +2*pron1$se,
pron1$pred -2*pron1$se)
```

Pronósticos con el modelo complicado II



Anexo

Prueba de Dickey-Fuller

- Es importante probar formalmente que una serie es estacionaria, particularmente en econometría. Las pruebas de estacionariedad se conocen como *prueba de raíz unitaria*. La prueba de Dickey-Fuller es una prueba de este tipo.
- La prueba consiste en estimar la regresión:

$$\nabla y_t = \phi y_{t-1} + \beta_1 \nabla y_{t-1} + \beta_2 \nabla y_{t-2} + \cdots + \beta_p \nabla y_{t-p} + u_t$$

  donde u_t es un componente de error que se supone tiene media 0.
- Si la serie no es estacionaria, entonces se espera que $\hat{\phi} \approx 0$. Si la serie es estacionaria, entonces $\hat{\phi} < 0$.
- Como los errores de este modelo están correlacionados, los supuestos típicos de regresión lineal no se cumplen, por lo que hay que probar la hipótesis $H_0 : \phi = 0$ vs. $H_a : \phi < 0$ utilizando una prueba diferente a la prueba de t que se usa en regresión.

Prueba de Dickey-Fuller I

- La prueba se puede obtener con la función `df.test` en el paquete `tseries` de R.

```
library(tseries)
adf.test(AirPassengers, k = 0)

Augmented Dickey-Fuller Test

data:  AirPassengers
Dickey-Fuller = -4.6392, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

```
adf.test(diff(AirPassengers,1), k = 0)

Augmented Dickey-Fuller Test

data:  diff(AirPassengers, 1)
Dickey-Fuller = -8.5472, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

```
adf.test(diff(AirPassengers,2), k = 0)

Augmented Dickey-Fuller Test

data:  diff(AirPassengers, 2)
Dickey-Fuller = -6.0136, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Prueba de Dickey-Fuller II

- En la salida previa, k es el número p de componentes incluidos en la ecuación de regresión. La hipótesis alternativa dice `stationary`, ya que es el caso en el que $\phi < 0$. El p-value determina si la serie es o no estacionaria.

Identificación de modelos más generales.

- Se presenta un resumen de las características generales de los modelos AR(p) y MA(q) que se compararán con las acf y pacf muestrales de una serie en particular para poder *identificar* y asociar su comportamiento a algún modelo en particular.

$AR(1)$ I

- El modelo es $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$.
- En lo que sigue, supondremos que la media de y_t es $E(y_t) = \mu$.
- Para que el proceso sea estacionario, se requiere que $|\phi_1| < 1$.
- La media, autocovarianza y autocorrelación son, respectivamente:
    - $\mu = E(y_t) = \frac{\phi_0}{1-\phi_1}$.
    - $\gamma_k = \phi_1^k \frac{\sigma_\epsilon^2}{1-\phi_1^2}$ para $k = 0, 1, 2, \ldots$.
    - $\rho_k = \phi_1^k$ para $k = 0, 1, 2, \ldots$.
- la función de autocorrelación es de la forma:

$$\rho_k = \phi_1^k, \qquad k = 0, 1, 2, \ldots$$

- fac decae exponencialmente cuando $\phi_1 > 0$ y decae exponencialmente pero oscilando en signo cuando $\phi_1 < 0$.

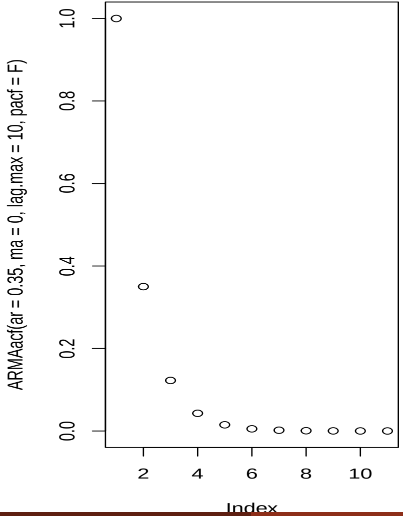
$AR(1)$ II

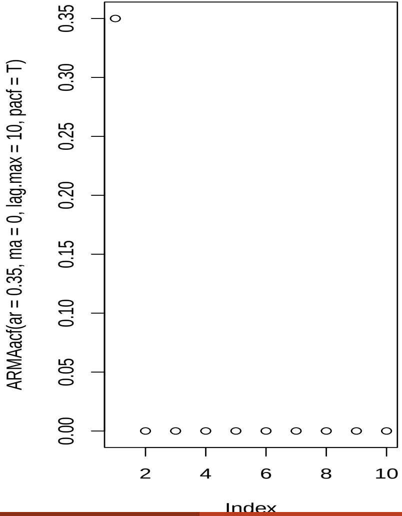
- la función de autocorrelación parcial es de la forma:

$$\alpha_k = \begin{cases} \phi_1, & k = 1 \\ 0, & k > 1 \end{cases}$$

ya que en la regresión, y_t sólo depende de y_{t-1} si el proceso es estacionario.

```
par(mfrow=c(1,2))
plot(ARMAacf(ar = 0.35, ma = 0, lag.max = 10, pacf = F))
plot(ARMAacf(ar = 0.35, ma = 0, lag.max = 10, pacf = T))
```


ARMAacf(ar = 0.35, ma = 0, lag.max = 10, pacf = F)
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
Index


ARMAacf(ar = 0.35, ma = 0, lag.max = 10, pacf = T)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
2
4
6
8
10
Index

AR(2) I

- El modelo es $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$.
- Para que el proceso sea estacionario, se requiere que se cumplan estas ecuaciones simultáneamente: $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ y $|\phi_2| < 1$.
- La media del proceso es $\mu = E(y_t) = \frac{\phi_0}{1-\phi_1-\phi_2}$.
- El cálculo de la autocovarianza teórica es complicado. Las fórmulas son recursivas (se conocen como ecuaciones de Yule-Walker):

$$\gamma_k = \begin{cases} \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\epsilon^2 & \text{si } k = 0 \\ \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} & \text{si } k > 0 \end{cases}$$

- Las autocorrelaciones se obtienen usando la ecuación $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$, para $k \geq 3$.

AR(2) II

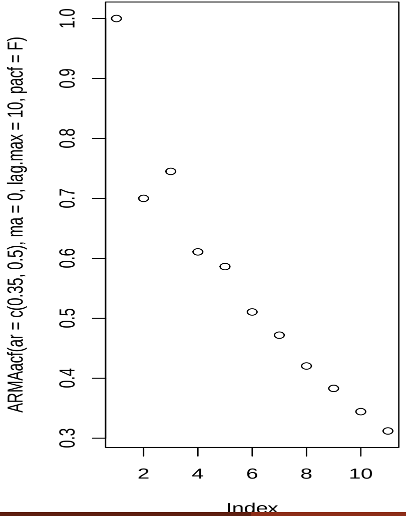
- Los valores de ρ_1 y ρ_2 se obtienen de resolver el sistema de ecuaciones (una vez que se conocen los valores de ϕ_1 y ϕ_2):

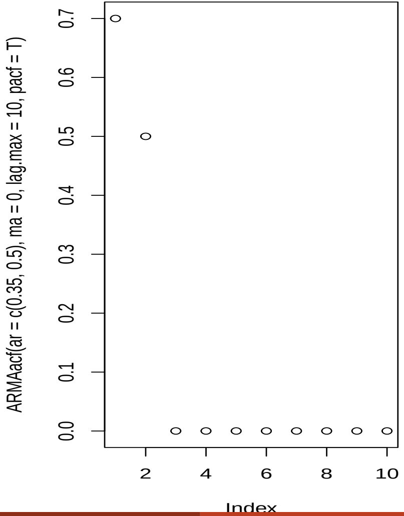
$$
\begin{aligned}
\rho_1 &= \phi_1 + \phi_2 \rho_1 \\
\rho_2 &= \phi_1 \rho_1 + \phi_2
\end{aligned}
$$

- Si $\phi_1^2 + 4\phi_2 \geq 0$, la acf es una mezcla de exponenciales decrecientes, y si $\phi_1^2 + 4\phi_2 < 0$, la autocorrelación es una onda sinusoidal decreciente.
- la pacf tiene picos en los rezagos 1 y 2, luego son ceros.

```
par(mfrow=c(1,2))
plot(ARMAacf(ar = c(0.35,0.5), ma = 0, lag.max = 10, pacf = F))
plot(ARMAacf(ar = c(0.35,0.5), ma = 0, lag.max = 10, pacf = T))
```

AR(2) III


ARMAacf(ar = c(0.35, 0.5), ma = 0, lag.max = 10, pacf = F)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
2
4
6
8
10
Index


ARMAacf(ar = c(0.35, 0.5), ma = 0, lag.max = 10, pacf = T)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2
4
6
8
10
Index

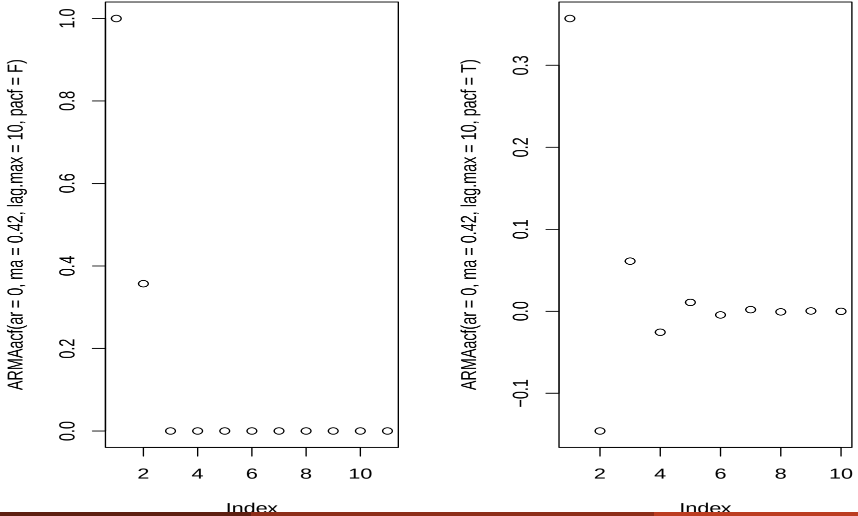
MA(1) I

- El modelo es $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$.
- El proceso es estacionario para cualquier valor de θ_1.
- La media, autocovarianza y autocorrelación son, respectivamente:
  - $E(y_t) = \mu$.
  - $\gamma_0 = \sigma_\epsilon^2(1 + \theta_1^2)$ y
  - $\rho_k = \begin{cases} \frac{\theta_1}{1+\theta_1^2} & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases}$
- La `acf` tiene un pico en $k = 1$ y después es 0, positivo si $\theta_1 > 0$, negativo si $\theta_1 < 0$
- La `pacf` tiene Decae exponencialmente: del lado negativo si $\theta_1 < 0$ y alternando en signo empezando en el lado positivo si $\theta_1 > 0$.

MA(1) II

```
par(mfrow=c(1,2))
plot(ARMAacf(ar = 0, ma = 0.42, lag.max = 10, pacf = F))
plot(ARMAacf(ar = 0, ma = 0.42, lag.max = 10, pacf = T))
```

MA(1) III


ARMAacf(ar = 0, ma = 0.42, lag.max = 10, pacf = F)
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
Index
ARMAacf(ar = 0, ma = 0.42, lag.max = 10, pacf = T)
−0.1
0.0
0.1
0.2
0.3
2
4
6
8
10
Index

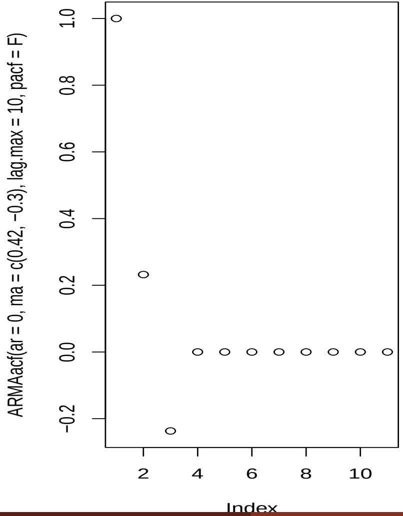
MA(2) I

- El modelo es $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$.
- El proceso es estacionario para cualquier valor de θ_1 y θ_2.
- La media, autocovarianza y autocorrelación son, respectivamente:
  - $E(y_t) = \mu$.
  - $\gamma_0 = \sigma_\epsilon^2(1 + \theta_1^2 + \theta_2^2)$ y
  - $\rho_k = \begin{cases} \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_1^2} & \text{si } k = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_1^2} & \text{si } k = 2 \\ 0 & \text{si } k > 2 \end{cases}$
- La `acf` tiene un picos en los rezagos 1 y 2 y después es 0.
- Decae exponencialmente en forma de función sinoidal decreciente. El patrón exacto depende de los signos y tamaños de θ_1 y θ_2.

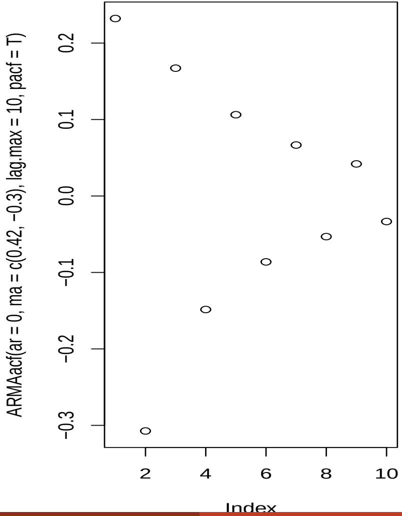
MA(2) II

```r
par(mfrow=c(1,2))
plot(ARMAacf(ar = 0, ma = c(0.42,-0.3), lag.max = 10, pacf = F))
plot(ARMAacf(ar = 0, ma = c(0.42,-0.3), lag.max = 10, pacf = T))
```

MA(2) III



ARMAacf(ar = 0, ma = c(0.42, −0.3), lag.max = 10, pacf = F)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
2
4
6
8
10
Index



ARMAacf(ar = 0, ma = c(0.42, −0.3), lag.max = 10, pacf = T)
0.2
0.1
0.0
−0.1
−0.2
−0.3
2
4
6
8
10
Index

Patrones generales de `acf` y `pacf`

- La siguiente tabla resume los casos más importantes a considerar para los patrones esperados en la acf y pacf para los modelos simples AR y MA:

Proceso	acf	pacf
AR(1)	Decae exponencialmente: en el lado positivo si $\phi_1 > 0$ y alternando signo comenzando en el lado negativo si $\phi_1 < 0$	Un pico en el rezago 1, y luego 0. Pico positivo si $\phi_1 > 0$ y negativo si $\phi_1 < 0$
AR(p)	Decae exponencialmente o como en forma sinusoidal. El patrón exacto depende de los signos y magnitudes de $\phi_1, \ldots, \phi_p$.	Picos de 1 a p, luego 0.
MA(1)	Pico en 1 y luego se va a 0. Pico positivo si $\theta_1 < 0$ y negativo si $\theta_1 > 0$	Decae exponencialmente: en el lado negativo si $\theta_1 > 0$ y alternando en signo comenzando en positivo si $\theta_1 < 0$.
MA(q)	Picos en los rezagos de 1 a q y luego 0.	Decae exponencialmente o como una función sinusoidal. El patrón exacto depende de los signos y tamaños de $\theta_1, \ldots, \theta_q$.