proyecto01

April 6, 2022

1 Proyecto 1 - Series de Tiempo I

1.0.1 Bruno Martinez - Cristóbal Cancino

1.0.2 Prof. Ronny Vallejos

1.1 ### 06 de abril de 2022

[1]:
```
library("zoo")
library("tseries")
library("tidyverse")
library("fpp2")
```

```
Attaching package: 'zoo'


The following objects are masked from 'package:base':

    as.Date, as.Date.numeric


Registered S3 method overwritten by 'xts':
  method     from
  as.zoo.xts zoo

Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo

  Attaching packages

tidyverse 1.3.1

  ggplot2 3.3.5      purrr   0.3.4
  tibble  3.1.6      dplyr   1.0.8
  tidyr   1.2.0      stringr 1.4.0
  readr   2.1.2      forcats 0.5.1
```

Conflicts

```
tidyverse_conflicts()
 dplyr::filter() masks stats::filter()
 dplyr::lag()    masks stats::lag()
```

Attaching packages

```
fpp2 2.4

 forecast  8.16      expsmooth 2.3
 fma       2.4
```

---

2 Problema 1

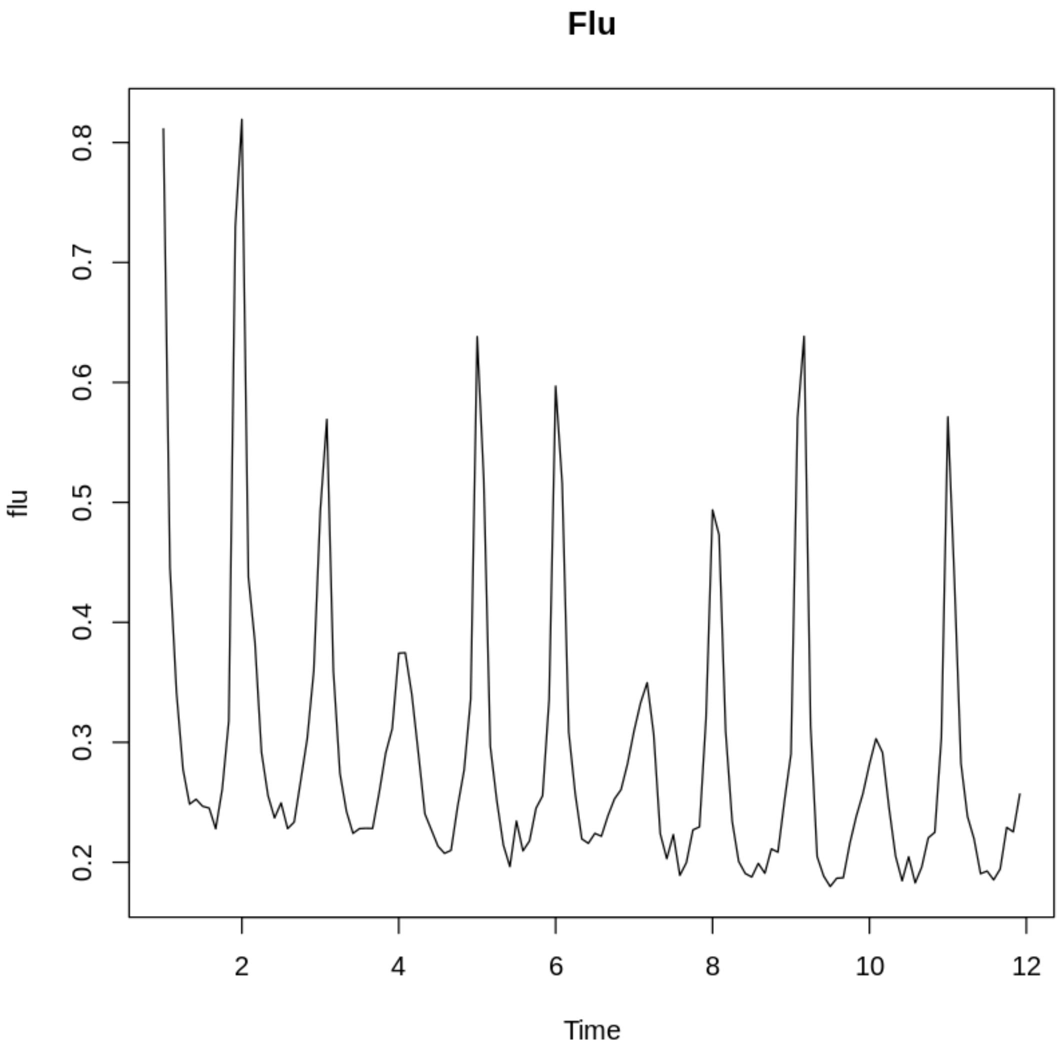
Considere la serie `flu.dat` que puede ser obtenida en el sitio http://www.stat.pitt.edu/stoffer/tsa2/tsa2.html.

1. Transforme la serie adecuadamente para observar el efecto de la transformación en la media y la varianza

Se graficarán los datos

[2]:
```r
flu <- ts(read.table("https://www.stat.pitt.edu/stoffer/tsa2/data/flu.dat.
 ↪txt"), frequency=12)
length(flu)
ts.plot(flu, main = 'Flu')
```

132



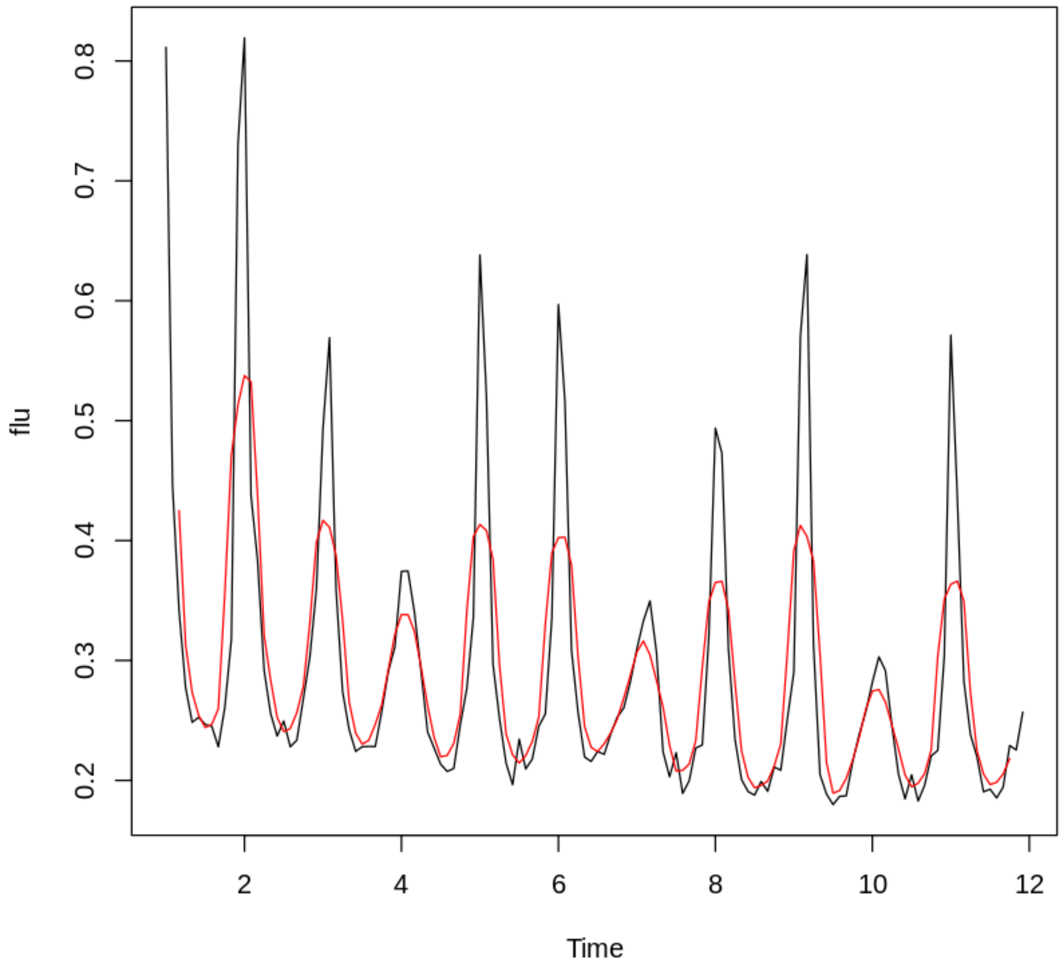
Deseamos estabilizar la varianza, en particular, reducir la dependencia de la variabilidad del tiempo. Para esto, se utilizará la **Media movil**.

[3]:
```
rm_flu <- rollmean(flu, 5)
head(rm_flu)
```

A Time Series: 1 × 6

	Mar	Apr	May	Jun	Jul	Aug
1	0.4249440	0.3131781	0.2733503	0.2540707	0.2441794	0.2466861

[4]:
```
seqplot.ts(flu,rm_flu)
```



[5]:
```
mean(rm_flu)
mean(flu)
```

0.28863929609375

0.291862284090909

[6]:
```
var(rm_flu)[1,1]
var(flu)[1,1]
```

0.00615542112341169

0.0157817324621534

Podemos ver que se cumple el **Resultado 1.6**, es decir, la media se mantiene relativamente constante y la varianza cambia pues se transforma en $\frac{\sigma^2}{2*q+1}$, donde se usó $q = 5$.

2. Se propone un modelo de la forma

$$Z_t = \sum_{j=0}^{\infty} \beta_j t^j + \epsilon_t$$

donde ϵ_t es un ruido blanco con varianza σ^2.

1. Proponga un método que permita truncar la serie infinita que define el modelo, de tal modo que el modelo resultante sea de la forma

$$Z_t = \sum_{j=0}^{p} b_j t^j + u_t,$$

donde $p \in \mathbb{N}$. Es decir, proponga un método para estimar p.

2. Estime p usando los datos de la serie `flu.dat`.
3. Estime los parámetros del modelo: β_j, $j = 0, 1, ..., \hat{p}$, donde $\hat{p}$ es la estimación de p propuesta en B. **Solución.**

Se propone transformar el modelo de la forma

$$Z_t = \sum_{j=0}^{\infty} \beta_j t^j + \epsilon_t \quad \Rightarrow \quad Z_t = \sum_{j=0}^{p} \beta_j t^j + \underbrace{\sum_{j=p+1}^{\infty} \beta_j t^j + \epsilon_t}_{u_t}$$

donde $u_t \approx N(\sum_{j=p+1}^{\infty} \beta_j t^j, \sigma^2)$. Para estimar p, se propone el siguiente método basado en el filtrado de Diferenciación y la Proposición 1.1.

```
FUNCTION obtener_p(datos)
p = 0
for(){
    datos <- diff(datos)
    calcular regresion lineal
    si |pendiente| < tol
        break
    p += 1
return p
```

En particular, utilizando los datos

[7]:
```
diff_datos <- rm_flu
p <- 0
tol <- 10^-5
previous_slope <- Inf
while(TRUE){
    diff_datos <- diff(diff_datos)
    range <- ts(1:length(diff_datos))
    regression <- lm(diff_datos ~ range)
    slope <- abs(regression$coefficients["range"])
    if(slope < tol | slope > previous_slope){
```

```
        break
    }
    previous_slope <- slope
    p <- p + 1
}
p
```

1

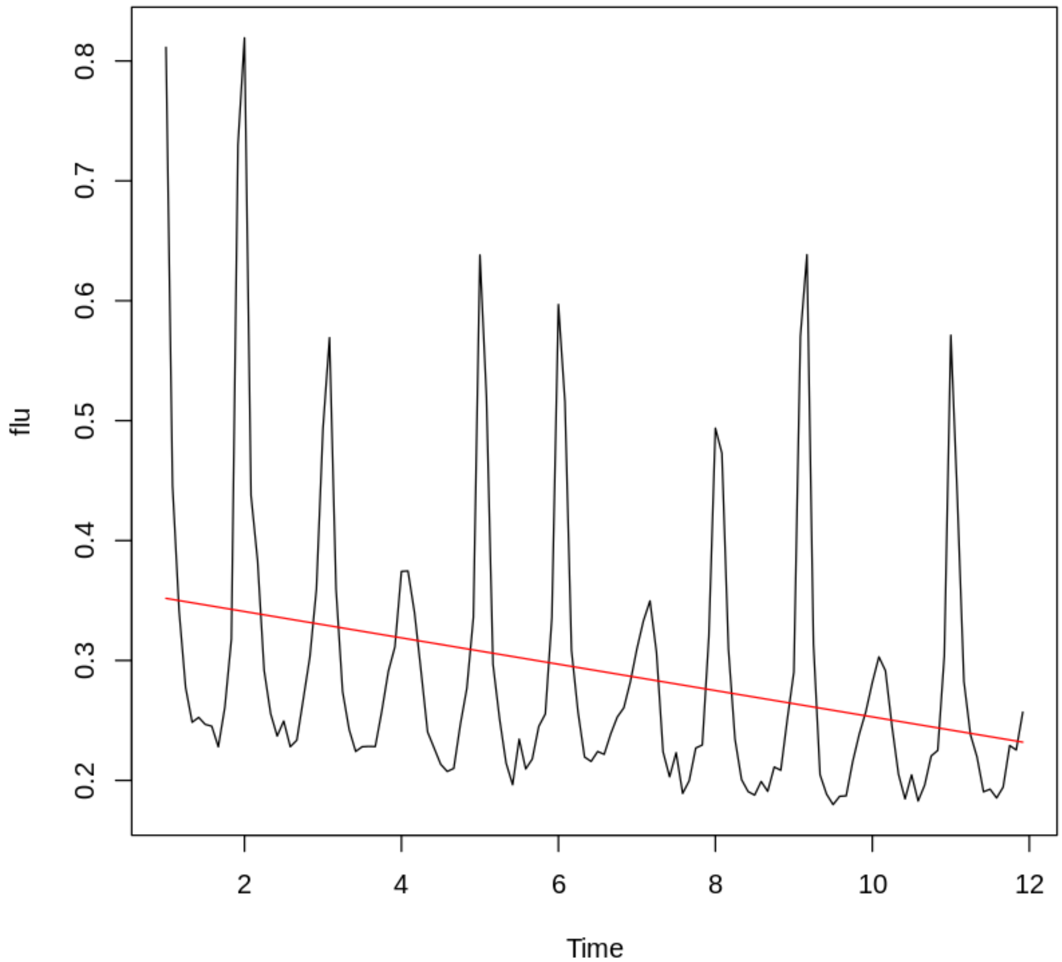
De esta forma, se tiene un modelo de regresión lineal, por lo que se calculará utilizando `lm`.

[8]:
```
q <- ts(1:length(flu), frequency=12)
regression <- lm(flu ~ poly(q,p,raw=TRUE))
regression
```

```
Call:
lm(formula = flu ~ poly(q, p, raw = TRUE))

Coefficients:
           (Intercept)  poly(q, p, raw = TRUE)
             0.3527291              -0.0009153
```

3. Grafique la serie original y la serie ajustada en un mismo gráfico. Hay evidencia para aseverar que el modelo estimado es una buena representación de los patrones de la serie original? Justifique.

[9]:
```
intercept = regression$coefficients["(Intercept)"]
range <- 1:length(flu)
regressiondata <- matrix(1, length(flu), 1)*intercept

if(p == 1){
    regressiondata <- regressiondata +  1:
 ↪length(flu)*regression$coefficients["poly(q, p, raw = TRUE)"]
} else {
    for(i in 1:p){
        regressiondata <- regressiondata +  1:
 ↪length(flu)*regression$coefficients[paste("poly(q, p, raw = TRUE)",i, sep =␣
 ↪"")]
    }
}

regressionts = ts(regressiondata, frequency=12)
seqplot.ts(flu, regressionts)
```



4. Ajuste un modelo de descomposición a la serie `flu.dat`.

```
[10]: flu <- ts(read.table("https://www.stat.pitt.edu/stoffer/tsa2/data/flu.dat.
       ↪txt"), frequency=12)
```

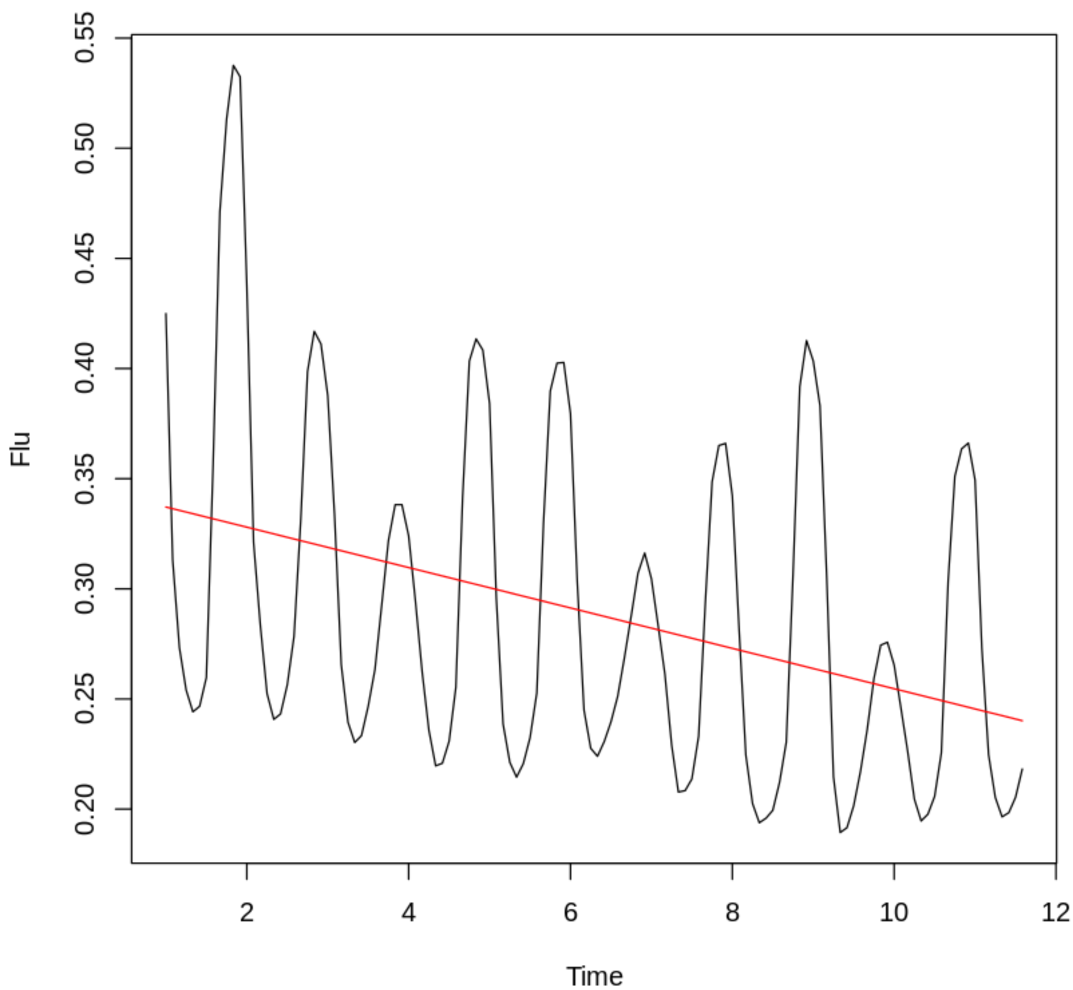
Se estimará la tendencia a través de un modelo lineal.

```
[11]: ma_flu <- ts(rollmean(flu, 5), frequency=12)
      range <- ts(1:length(ma_flu), frequency=12)
      regression <- lm(ma_flu ~ range)
      regression
```

```
Call:
```

```
lm(formula = ma_flu ~ range)

Coefficients:
(Intercept)        range
  0.3379070   -0.0007638
```

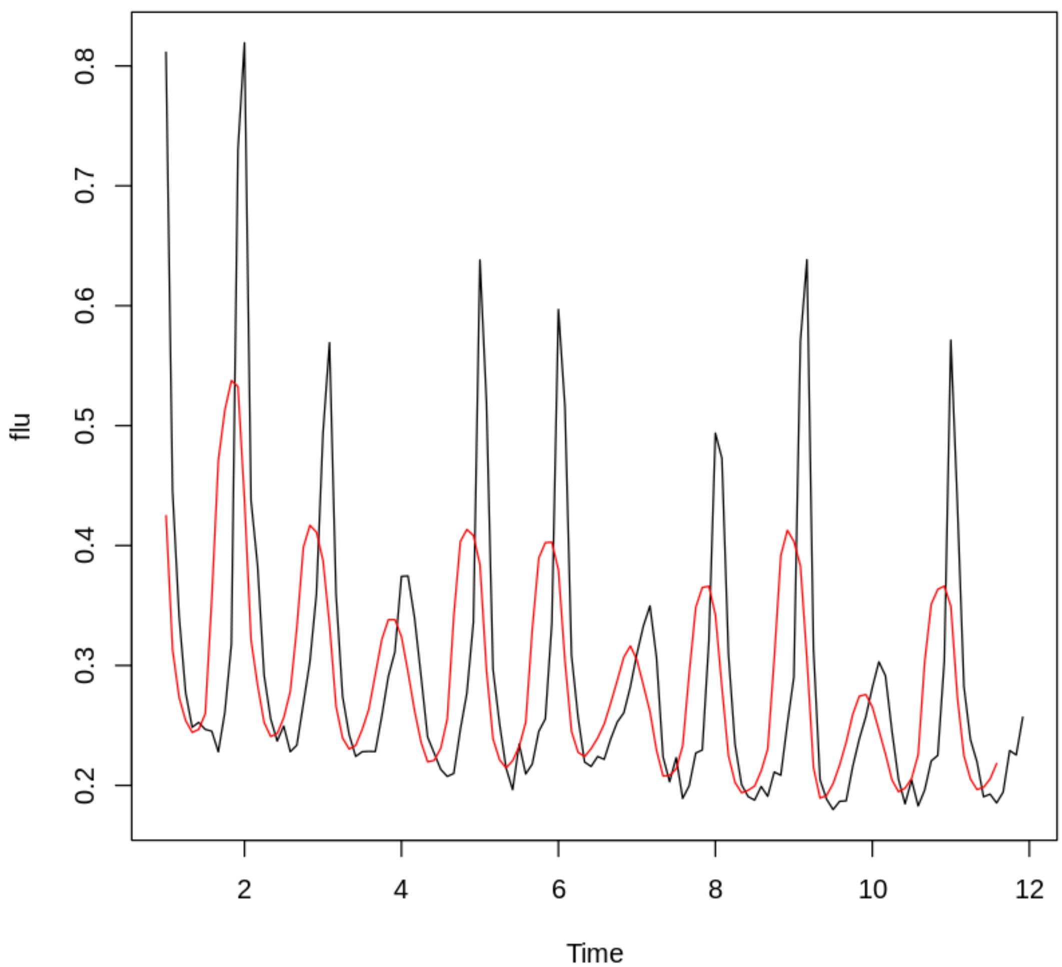
[12]:
```
intercept = regression$coefficients["(Intercept)"]
slope = regression$coefficients["range"]
range <- ts(1:length(ma_flu), frequency=12)
trend <- ts(intercept + slope * range, frequency=12)
seqplot.ts(ma_flu, trend, ylab="Flu")
```



Estimamos entonces el modelo estacional

[13]:
```
seasonal <- ma_flu - trend
```

[14]:
```
linearmodel <- ts(trend + seasonal, frequency=12)
seqplot.ts(flu,linearmodel)
```

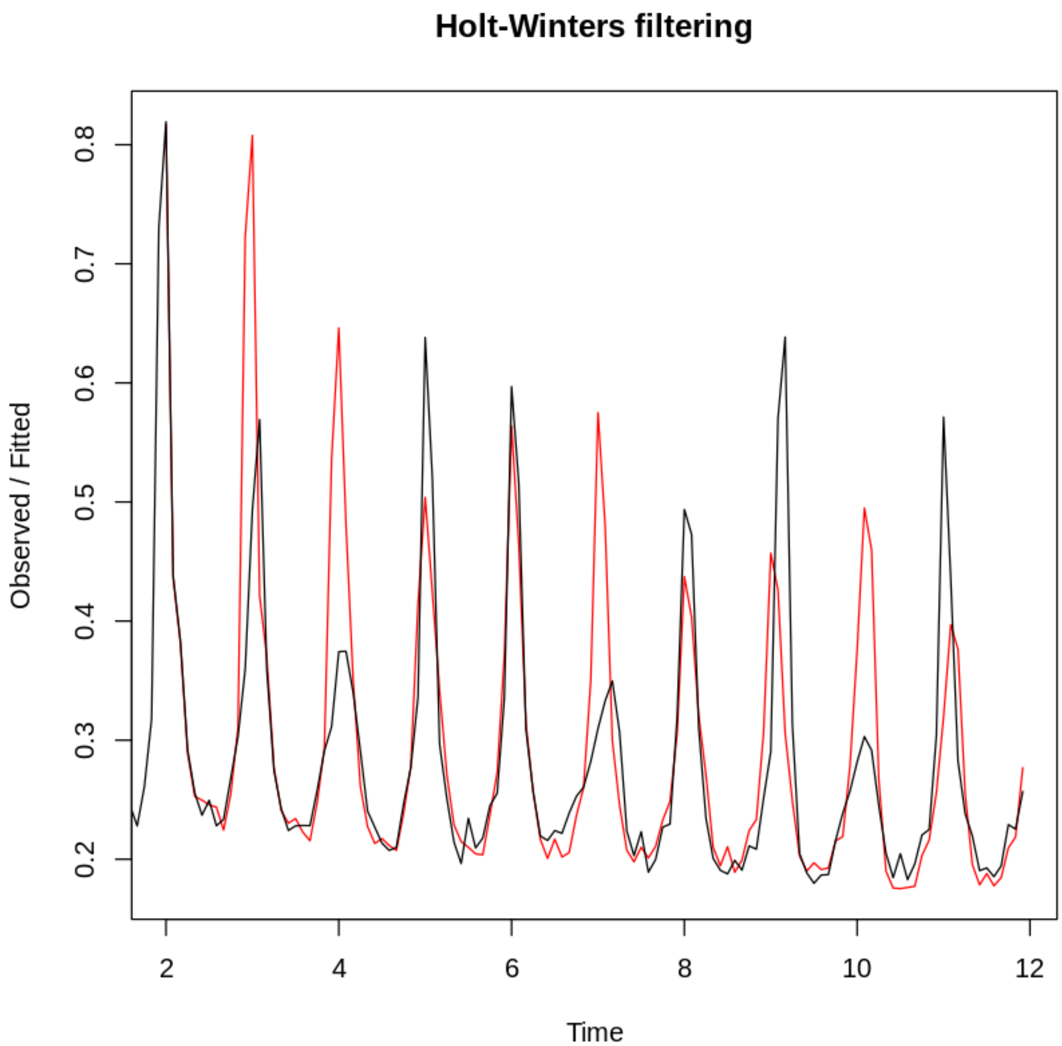


5. Ajuste un modelo de Holt-Winters a la serie `flu.dat`.

[15]:
```
flu <- ts(flu, frequency=12)
hw_flu <- HoltWinters(flu)
plot(hw_flu)
summary(hw_flu)
```

```
                Length Class  Mode
fitted          480    mts    numeric
```

```
x             132    ts     numeric
alpha           1    -none- numeric
beta            1    -none- numeric
gamma           1    -none- numeric
coefficients   14    -none- numeric
seasonal        1    -none- character
SSE             1    -none- numeric
call            2    -none- call
```



6. De los modelos propuestos en la parte 2., 4. y 5., ¿cuál es el más apropiado?

El modelo 5. parece ser más apropiado.

---

3 Problema 2

En este ejercicio es necesario obtener la serie asociada al calentamiento de la tierra descrito en grados centígrados entre los años 1900-1997. Esta serie es presentada en `Shumway and Stoffer (2000), página 5`. Para bajar el archivo `globtemp.dat`, encuentre el sitio web http://www.stat.pitt.edu/stoffer/tsa2/tsa2.html.

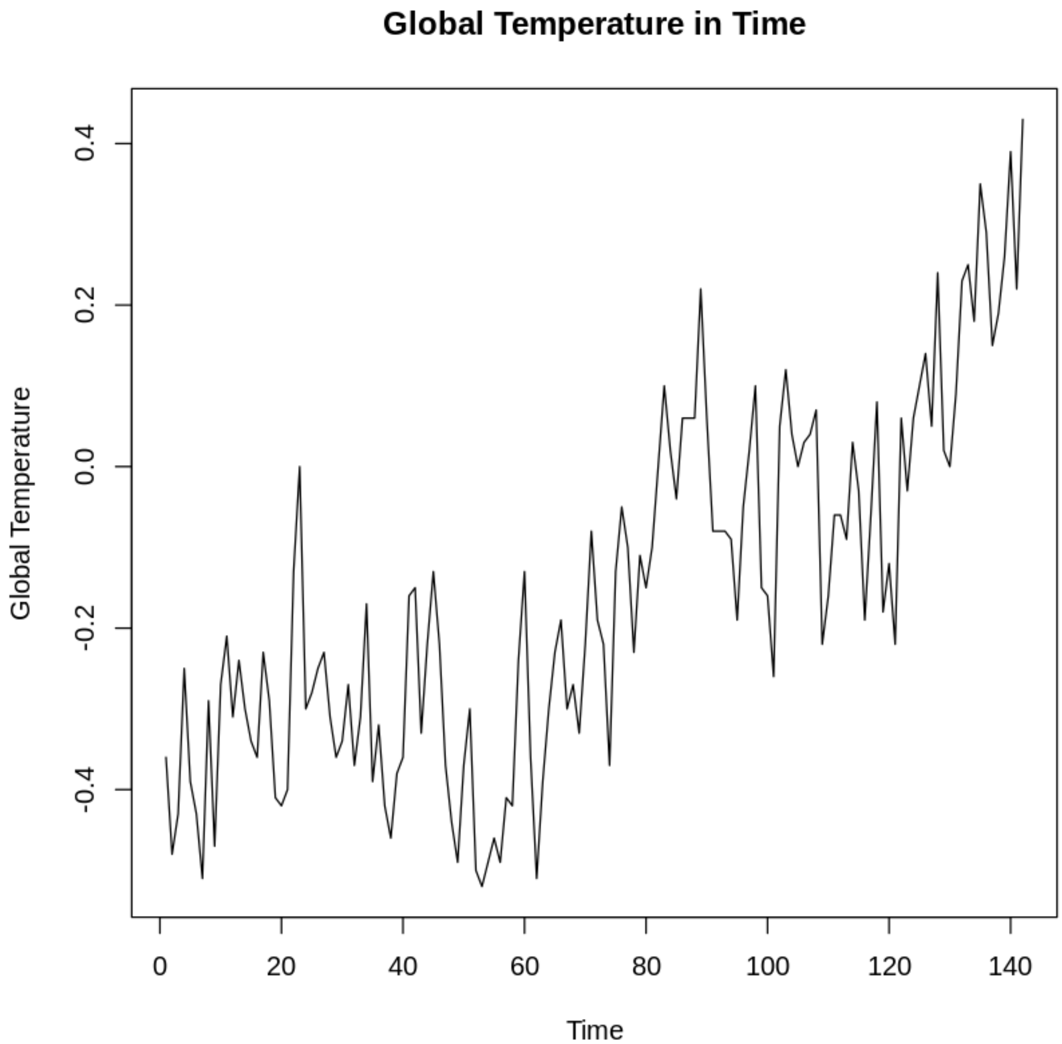
1. Grafique la serie `globtemp.dat` en el tiempo.

[16]:
```
globtemp <- ts(read.table("https://www.stat.pitt.edu/stoffer/tsa2/data/globtemp.
↪dat.txt"))
head(globtemp)
```

A Time Series: 6 × 1

V1
-0.36
-0.48
-0.43
-0.25
-0.39
-0.43

[17]:
```
ts.plot(globtemp, main='Global Temperature in Time', ylab="Global Temperature")
```



2. Use un modelo de suavizamiento exponencial simple para predecir la serie hacia el futuro. Considere un valor apropiado para α.

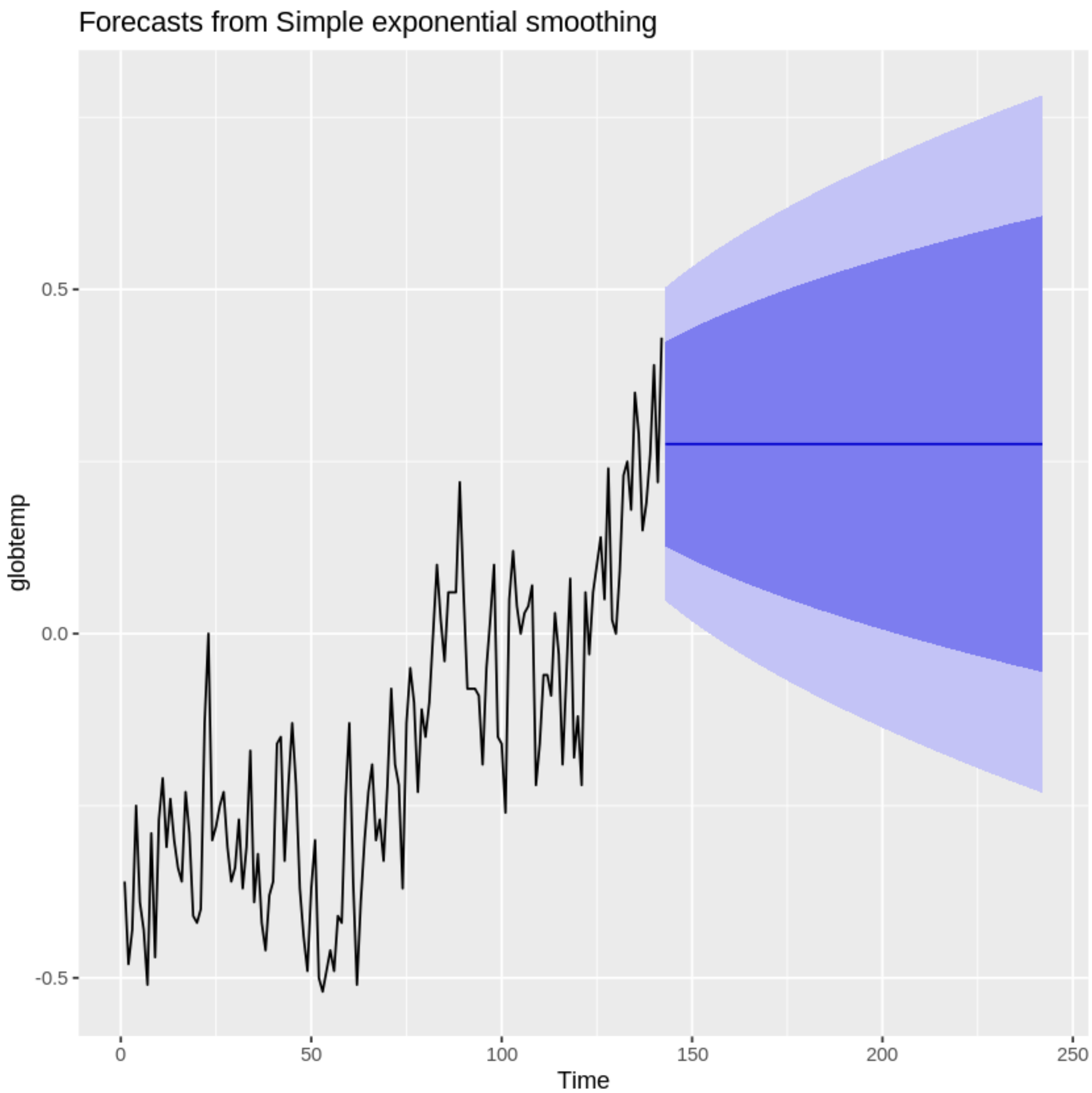
Debemos recordar que para utilizar un modelo de suavizamiento exponencial simple, se deben qui

[18]:
```
length(globtemp)
```

142

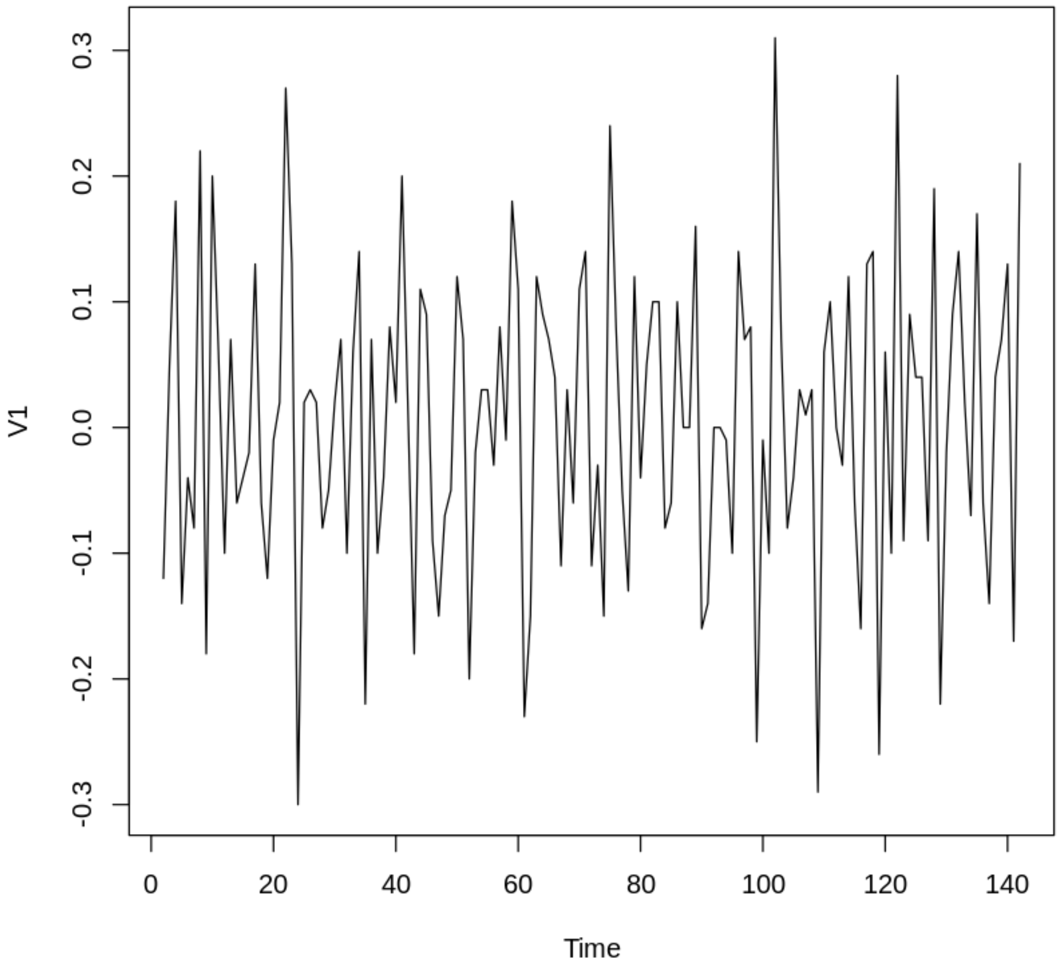
[19]:
```
globtemp.train <- window(globtemp, end = 100)
globtemp.test <- window(globtemp, start = 101)
```

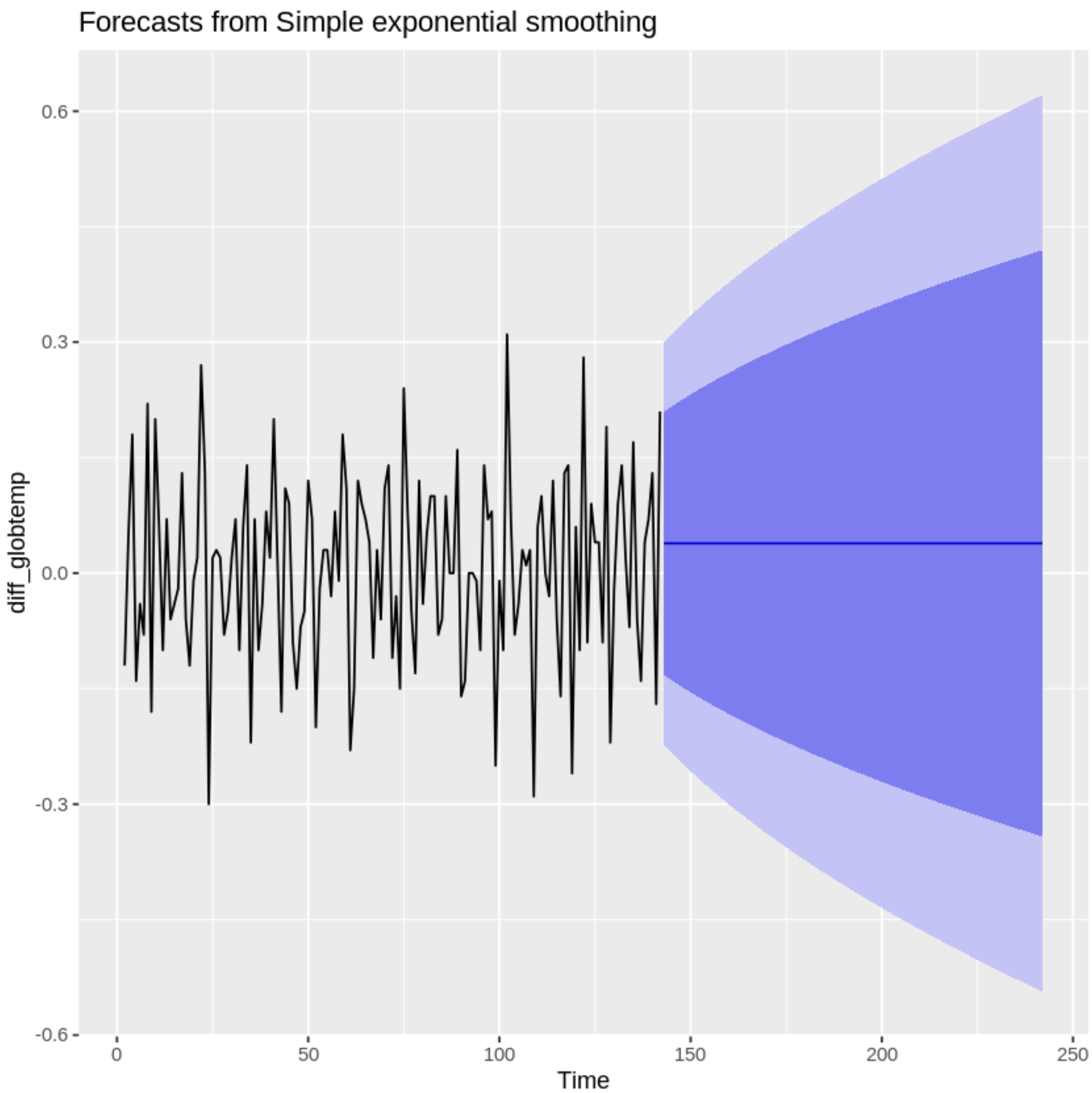
[20]:
```
ses_globtemp <- ses(globtemp, alpha = .2, h = 100)
autoplot(ses_globtemp)
```



Observamos que no está capturando la tendencia correctamente. Esto es porque el Suavizamiento E

[21]:
```
diff_globtemp <- diff(globtemp)
plot(diff_globtemp)
ses_diff_globtemp <- ses(diff_globtemp, alpha = .2, h = 100)
autoplot(ses_diff_globtemp)
```


0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
V1
0
20
40
60
80
100
120
140
Time

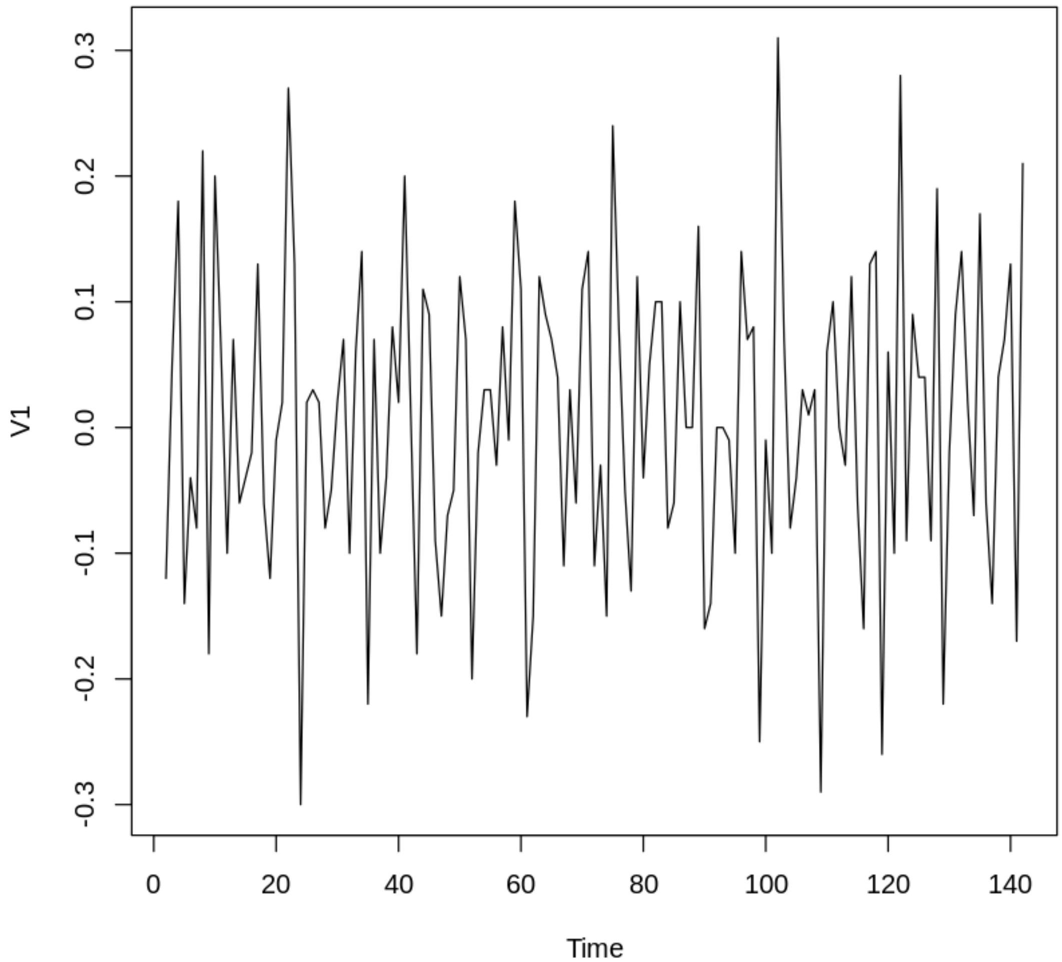


3. Describa las bondades y limitaciones del modelo usado en los puntos anteriores.

```
Entre las bondades que uno puede encontrar, se tiene que le da mayor peso al pasado reciente y

El mayor problema del modelo es que no son precisos al momento de utilizar tendencias, y es po
```

4. A partir de la serie original obtenga una serie sin tendencia.

[22]:
```
diff_globtemp = diff(globtemp)
plot(diff_globtemp)
```



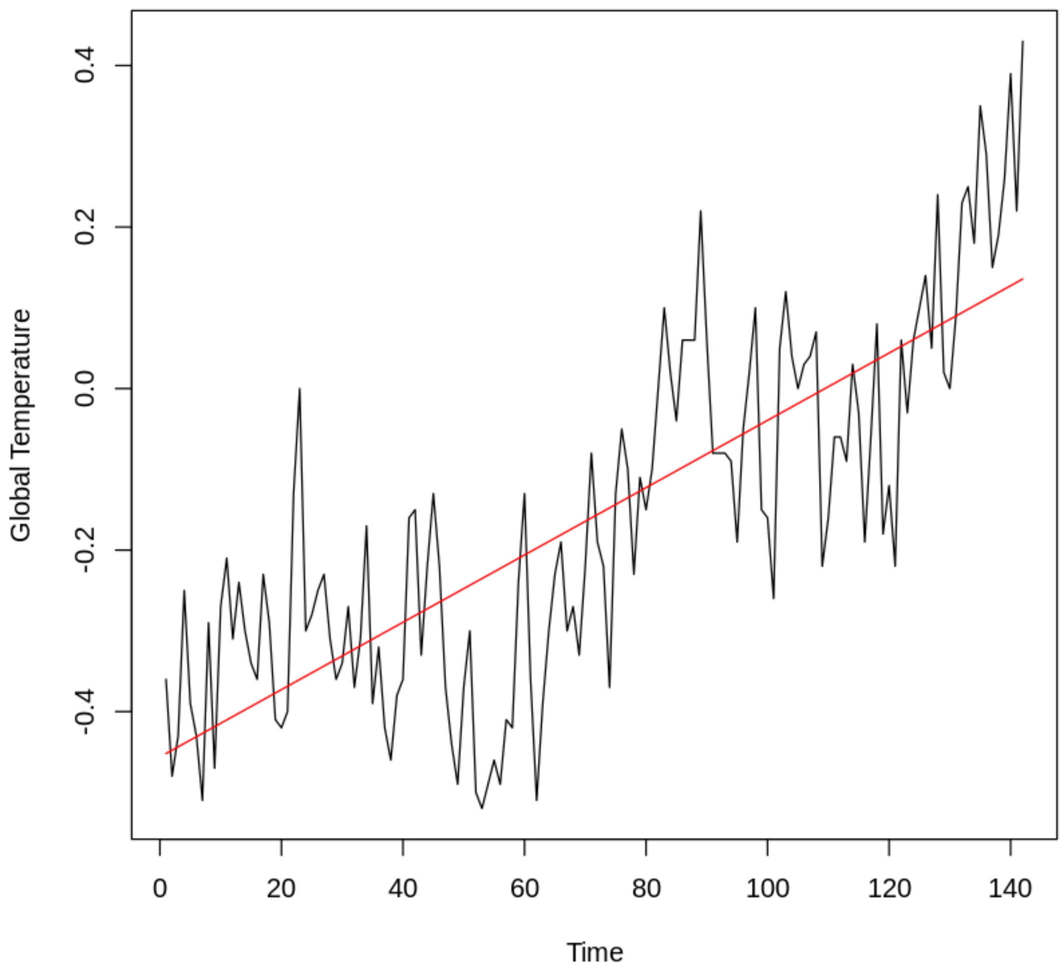
5. Estime un modelo de regresión de la forma

$$Z_t = \beta_0 + \beta_1 t + \epsilon_t,$$

para la serie **globtemp.dat**, donde ϵ_t es una colección de variables aleatorias no correlacionadas en media cero y varianza δ^2. ¿El modelo ajustado luce similar a la serie original?

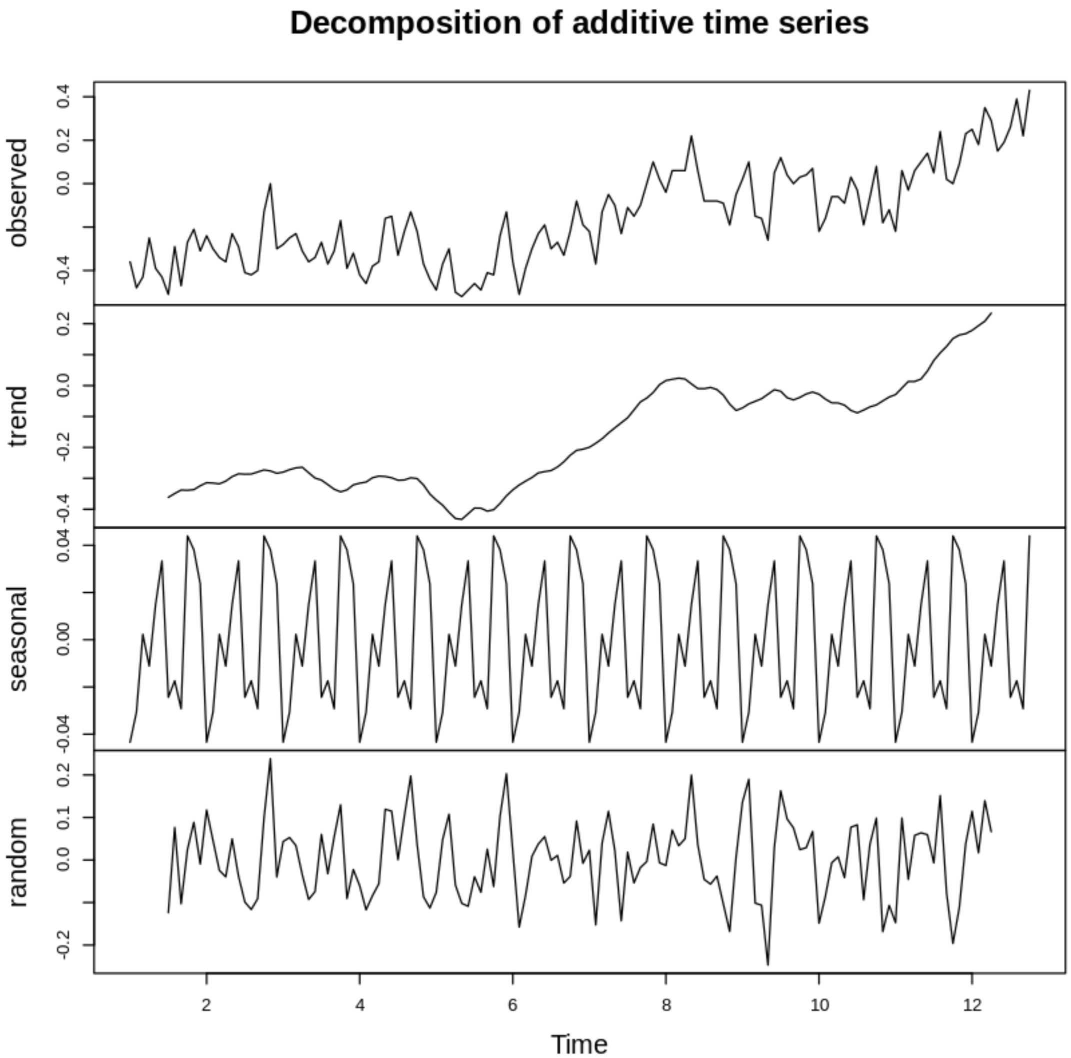
[23]:
```
range <- ts(1:length(globtemp))
regression <- lm(globtemp ~ range)

intercept = regression$coefficients["(Intercept)"]
slope = regression$coefficients["range"]
trend <- intercept + slope * range
```

```
seqplot.ts(globtemp, trend, ylab="Global Temperature")
```



6. Descomponga la serie **globtemp.dat** en tres partes: una tendencia, una parte estacional y una componente residual. Describa que observa.

[24]:
```
dec_globtemp <- ts(globtemp, frequency=12)
dec <- decompose(dec_globtemp)
plot(dec)
```




```
Podemos observar que los datos poseen una componente estacional notable, asi como la tendencia
```

---

4 Problema 3

Al analizar cierta serie de tiempo trimestral se usó un método ingenuo obteniéndose:

1. Ecuación de tendencia $T(t) = 84.65 + 4.71t$.
2. Serie de residuos $W(t) = Y(t) - Z(t)$. | Trimestre | 1997 | 1998 | 1999 | 2000 | 2001 | | | | | |:———:|:——:|:——:|:——:|:——:|:—-:|:—:|:—:|:—:|:—:| | 1 | - | 20 | 15 | 13.75 | 8.75 | | | | | | 2 | - | -6.25 | -7.50 | -7.50 | 2.50 | | | | | | 3 | -11.25 | -11.25 | -21.50 | -13.75 | - | | | | | | 4 | 0.95 | 1.05 | 1.11 | 1.05 | - | | | | |

En base a los resultados de 1. y 2., dado que $t = 1$ corresponde al primer trimestre de 1997, prediga los valores de la serie en cada uno de los trimestres de 2002.

Solución.

Primero obtendremos los valores de los e_i, con $i = 1, ..., 4$. Note que e_i corresponde al promedio de los $W(t)$ de cada trimestre. Calculando:

- $e_1 = (20 + 15 + 13.75 + 8.75)/4 = 14.38$
- $e_2 = (-6.25 - 7.50 - 7.50 + 2.50)/4 = -4.69$
- $e_3 = (-11.25 - 11.25 - 21.50 - 13.75)/4 = -14.44$
- $e_4 = (0.95 + 1.05 + 1.11 + 1.05)/4 = 1.04$

Calculando el promedio de los e_i, para $i = 1, ..., 4$

$$\overline{e} = (14.38 - 4.69 - 14.44 + 1.04)/4 = -0.93$$

Por otra parte la estimacion de la componente estacional, está dada por:

- $\hat{E_1} = e_1 - \overline{e} = 15.31$
- $\hat{E_2} = e_2 - \overline{e} = -3.76$
- $\hat{E_3} = e_3 - \overline{e} = -13.51$
- $\hat{E_4} = e_4 - \overline{e} = 1.97$

Note que la tendencia viene dada por $T_t = 84.65 + 4.71 * t$, y dado que nos piden predecir los valores de Z para el año 2002, entonces tendremos que calcular los trimestres 21,22,23,24. En efecto:

- $\hat{T_{21}} = 183.56$
- $\hat{T_{22}} = 188.27$
- $\hat{T_{23}} = 192.98$
- $\hat{T_{24}} = 197.69$

Finalmente , calculando lo pedido (los valores de la serie para cada trimestre del año 2002).

- $\hat{Z_{21}} = \hat{T_{21}} + \hat{E_1} = 198.87$
- $\hat{Z_{22}} = \hat{T_{22}} + \hat{E_2} = 184.51$
- $\hat{Z_{23}} = \hat{T_{23}} + \hat{E_3} = 179.47$
- $\hat{Z_{24}} = \hat{T_{24}} + \hat{E_4} = 199.66$

---

5 Problema 4

Considere una serie de tiempo $\{Z_t : t \in T\}$ descrita por la ecuación:

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + S_t + \epsilon_t$$

donde β_0, β_1 y β_2 son parámetros desconocidos del modelo, S_t es un efecto estacional conocido y ϵ_t es un ruido aleatorio con media y varianza constante. Determine las ecuaciones que permiten estimar los parámetros β_0, β_1 y β_2.

Solución.

Considere una serie de tiempo $Z_t : t \in T$ descrita por la ecuacion:

$$Z_t = \beta_0 + \beta_1 * t + \beta_2 * t^2 + S_t + \epsilon_t.$$

Note que la serie puede ser escrita de la siguiente forma:

$$Z_t = T_t + S_t + \epsilon_t,$$

con T_t, un modelo cuadratico de la estimacion de la tendencia. Por otra parte, notemos que S_t corresponde a la componente estacional y en este caso la eliminaremos debido a que se considera poco interesante y dificulta nuestro analisis. Por lo que obtenemos la siguiente serie:

$$Z_t = \beta_0 + \beta_1 * t + \beta_2 * t^2 + \epsilon_t, t \in T$$

la cual puede ser resuelta a traves de una regresion polinomial.

Por tanto el modelo se puede escribir como un sistema de ecuaciones lineales

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ . \\ . \\ . \\ z_T \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & t_T & t_T^2 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ . \\ . \\ . \\ \epsilon_T \end{bmatrix}$$

y por ende de forma matricial nos quedaria de la forma:

$$Z = X\beta + \epsilon$$

Ahora podemos obtener el vector de coeficientes de regresion polinomial estimados , usando la estimacion de minimos cuadrados. Obteniendo asi:

$$\hat{\beta} = (X^T * X)^{-1} * X^T * Z$$

la unica solucion a traves de minimos cuadrados, si

$$(X^T * X)$$

es invertible. Por tanto lo unico que nos queda vertificar es lo ,mencionado anteriormente. En efecto:

$$(X^T * X) = \begin{bmatrix} 1 & 1 & 1 & . & . & . & 1 \\ t_1 & t_2 & t_3 & . & . & . & t_T \\ t_1^2 & t_2^2 & t_3^2 & . & . & . & t_T^2 \end{bmatrix} \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & t_T & t_T^2 \end{bmatrix} = \begin{bmatrix} T & \sum_{i=1}^T t_i & \sum_{i=1}^T t_i^2 \\ \sum_{i=1}^T t_i & \sum_{i=1}^T t_i^2 & \sum_{i=1}^T t_i^3 \\ \sum_{i=1}^T t_i^2 & \sum_{i=1}^T t_i^3 & \sum_{i=1}^T t_i^4 \end{bmatrix}$$

.

Como $t_i >= 0$, y los $t_{i+1} > t_i$, entonces se puede verificar claramente que el $det(X^t * X) \neq 0$ y por tanto es invertible.