Series de tiempo no estacionarias

Samuel Méndez Villegas

11/17/2022

Usa los datos de las ventas de televisores para familiarizarte con el análisis de tendencia de una serie de tiempo:

Año 1 2 3 4

Trimestre 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

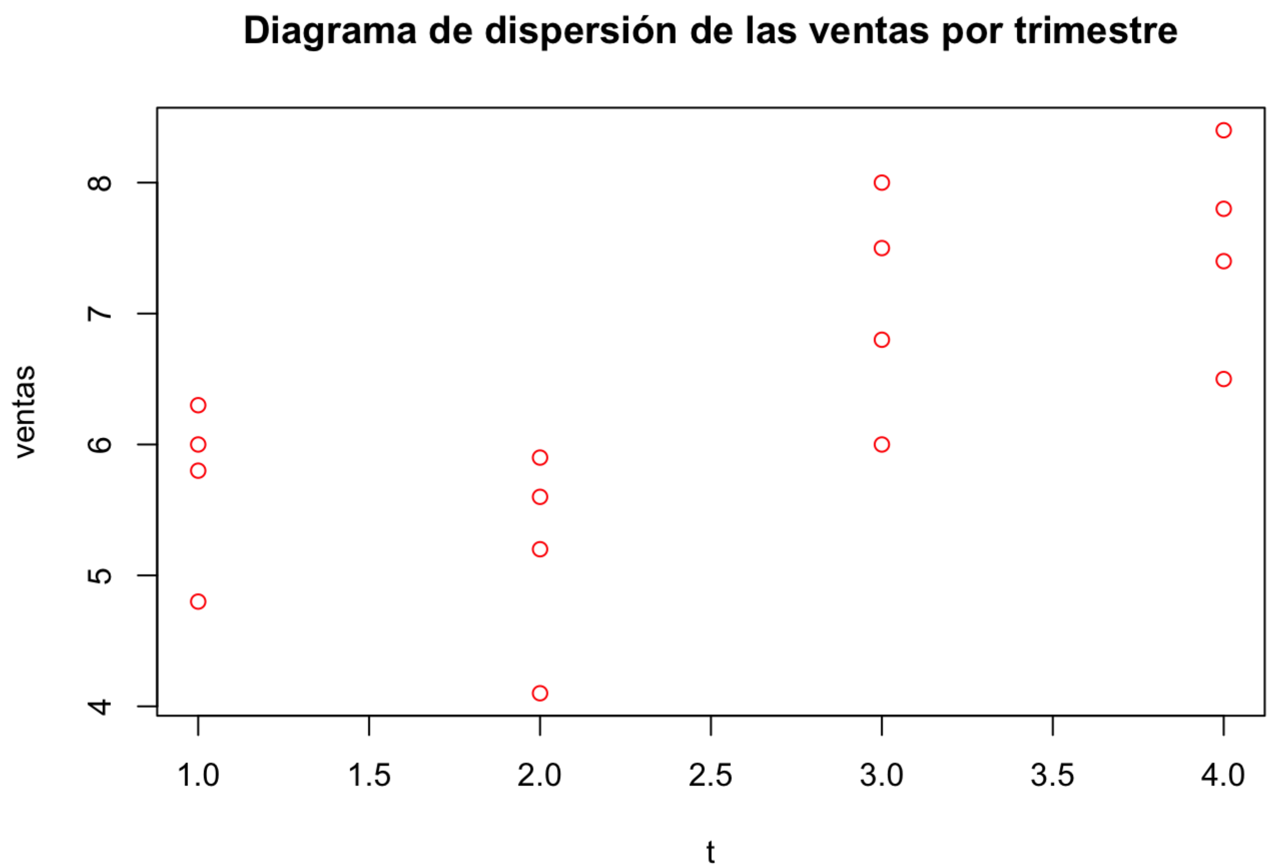
Ventas (miles) 4.8 4.1 6.0 6.5 5.8 5.2 6.8 7.4 6.0 5.6 7.5 7.8 6.3 5.9 8.0 8.4

```
t = c(1,2,3,4,1,2,3,4,1,2,3,4,1,2,3,4)
ventas = c(4.8,4.1,6.0,6.5,5.8,5.2,6.8,7.4,6.0,5.6,7.5,7.8,6.3,5.9,8.0,8.4)
p = NA
c = NA

df = data.frame(t, ventas, p, c)
df
```

```
##    t ventas  p  c
## 1  1    4.8 NA NA
## 2  2    4.1 NA NA
## 3  3    6.0 NA NA
## 4  4    6.5 NA NA
## 5  1    5.8 NA NA
## 6  2    5.2 NA NA
## 7  3    6.8 NA NA
## 8  4    7.4 NA NA
## 9  1    6.0 NA NA
## 10 2    5.6 NA NA
## 11 3    7.5 NA NA
## 12 4    7.8 NA NA
## 13 1    6.3 NA NA
## 14 2    5.9 NA NA
## 15 3    8.0 NA NA
## 16 4    8.4 NA NA
```

Realiza el gráfico de dispersión. Observa la tendencia y los ciclos.

```
plot(t, ventas, main = 'Diagrama de dispersión de las ventas por trimestre', col = 'red'
)
```

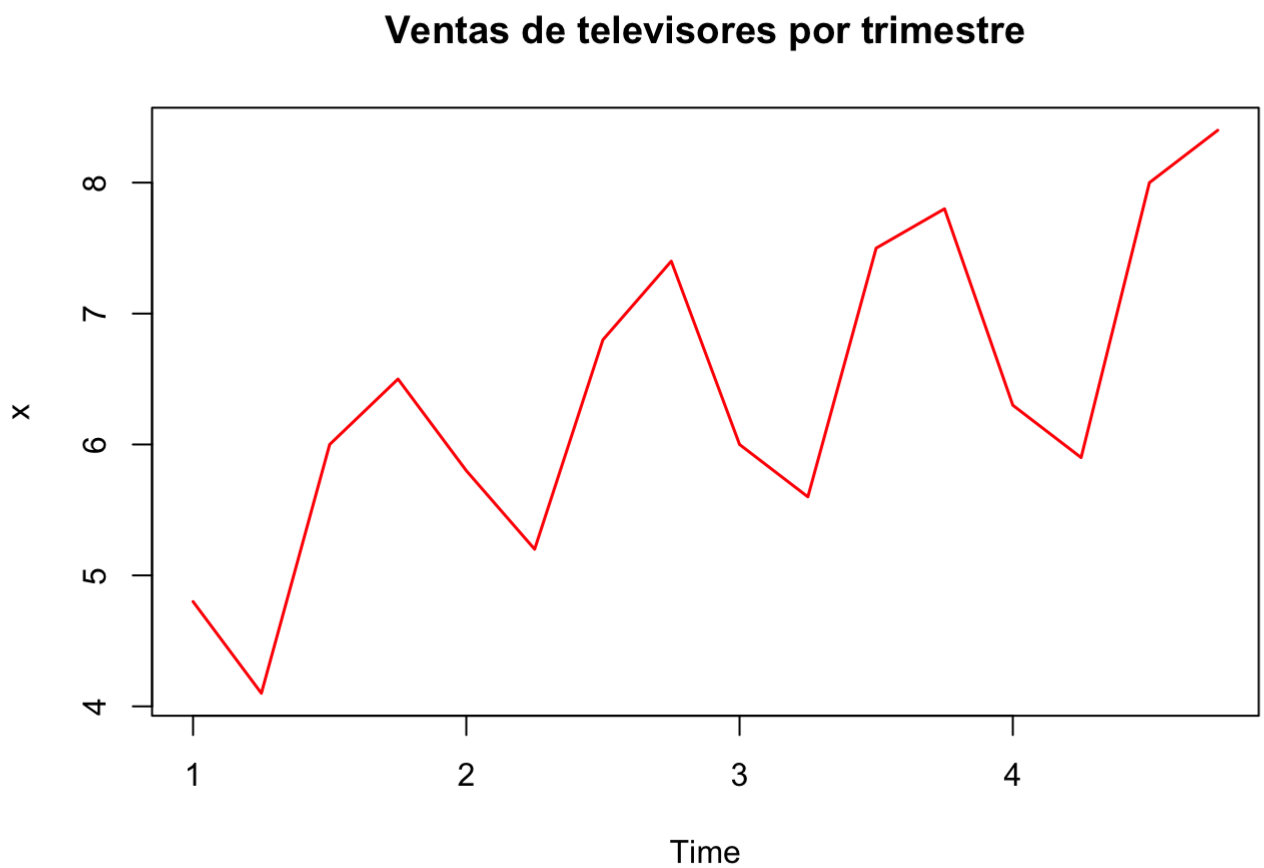


Se observa que dependiendo el trimestre del año, las ventas tienden a comportarse de manera distinta.

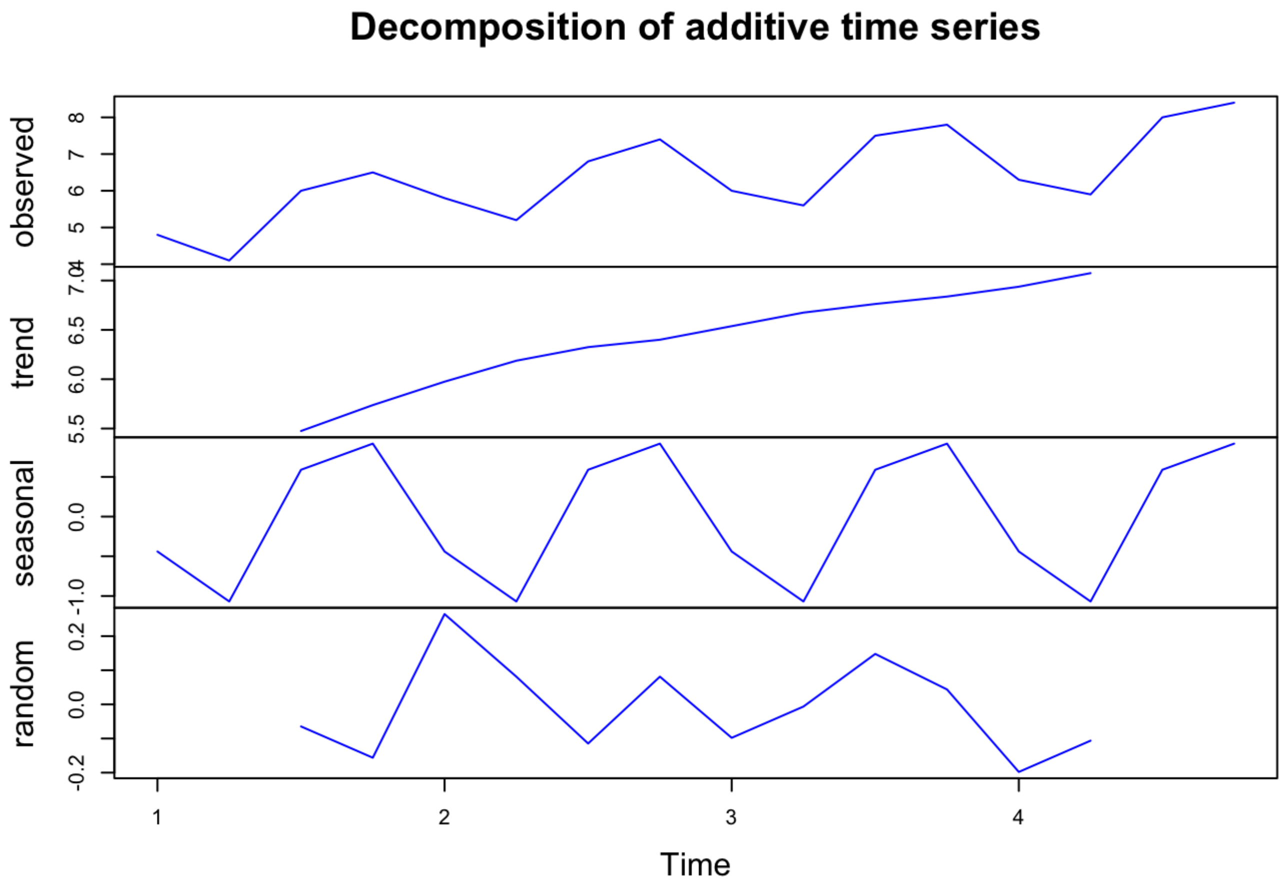
Realiza el análisis de tendencia y estacionalidad

Descompón la serie en sus 3 componentes e interprétalos

```
x = ts(ventas, frequency = 4, start(c(2016,1)))
plot.ts(x,
        col = "red",
        main = 'Ventas de televisores por trimestre',
        lwd = 1.5)
```




```
T = decompose(x)
plot(T, col ="blue")
```



En la serie de tiempo original, es decir la primer gráfica, se observa claramente que no es estacionaria, ya que presenta una tendencia a aumentar las ventas con el paso del tiempo. Sin embargo, se puede ver que hay existe estacionalidad, es decir que se sigue un ciclo que se repite cada año.

Lo mencionado anteriormente se puede observar de mejor forma en la descomposición de la serie de tiempo. La primer gráfica muestra la serie original. La segunda, gráfica muestra la tendencia, que como se mencionó, es una tendencia positiva. La tercer gráfica muestra la estacionariedad, en donde se observa que los años tienen un comportamiento bastante similar.

Analiza el modelo lineal de la tendencia:

Realiza la regresión lineal de la tendencia (ventas desestacionalizadas vs tiempo)

En primer lugar, se calculan los promedios móviles centrados, los cuales se obtienen a partir de los primeros promedios móviles. En este caso se estarán utilizando el promedio móvil de cuatro trimestres.

```r
## Se calculan los primeros promedios móviles
k = nrow(df) # Número de observaciones del data frame
n = 4 # Número de promedios móviles

promedios_moviles = c()

for(i in 1:(k-n+1)){
  promedios_moviles[i+n] = (ventas[i] + ventas[i+1] + ventas[i+2] + ventas[i+3]) / n
}

## Promedios móviles generados
promedios_moviles
```

```
##  [1]    NA    NA    NA    NA 5.350 5.600 5.875 6.075 6.300 6.350 6.450 6.625
## [13] 6.725 6.800 6.875 7.000 7.150
```

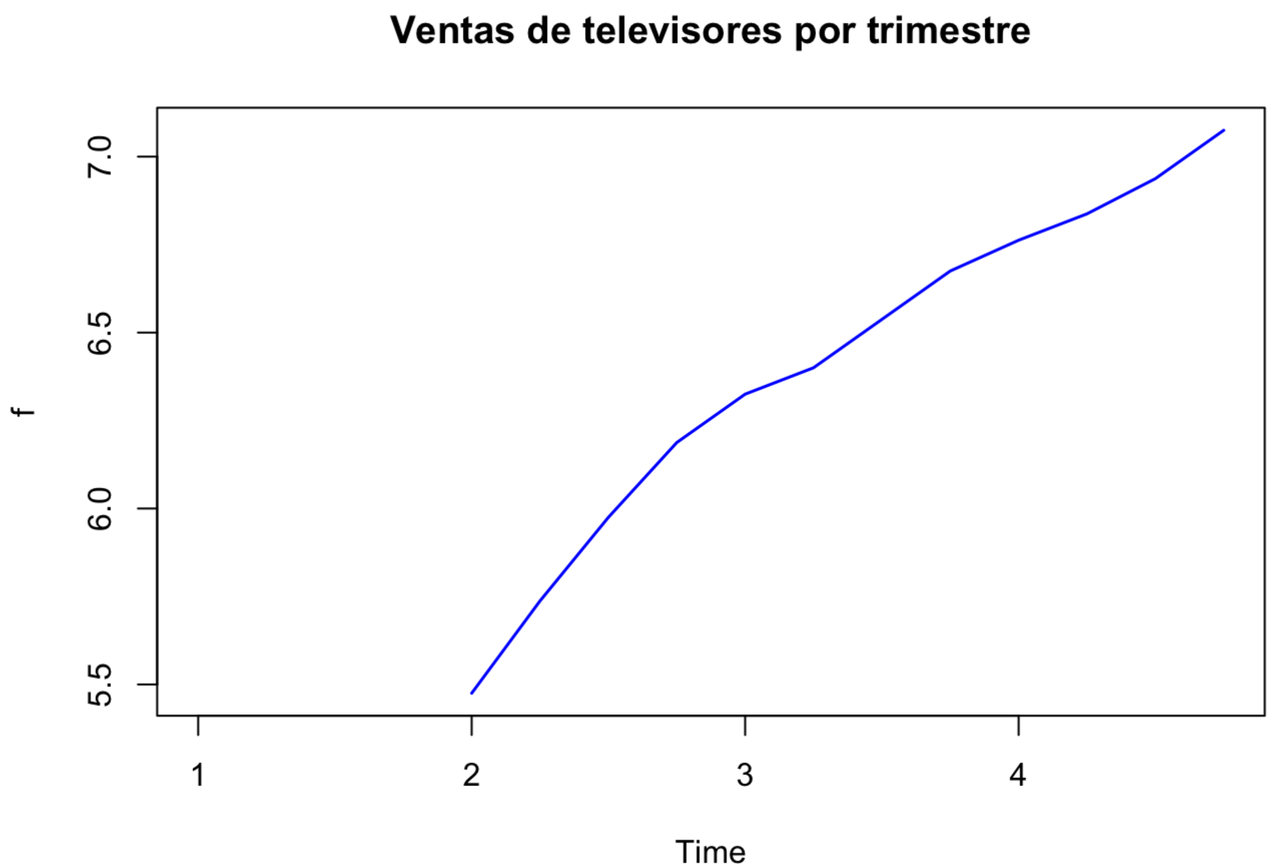
En la columna p del data frame anterior se encuentran los promedios móviles de cuatro trimestres. Para obtener los promedios móviles centrados, se obtiene el promedio de los promedios (cada 2), quedando de la siguiente forma:

```r
promedios_centrados = c()

for(i in 5:16){
  promedios_centrados[i] = (promedios_moviles[i] + promedios_moviles[i+1]) / 2
}

## Se obtienen los promedios moviles centrados
promedios_centrados
```

```
##  [1]     NA     NA     NA     NA 5.4750 5.7375 5.9750 6.1875 6.3250 6.4000
## [11] 6.5375 6.6750 6.7625 6.8375 6.9375 7.0750
```

Realizando la gráfica correspondiente a los promedios móviles centrados, la serie de tiempo se observa de la siguiente manera.

```r
f = ts(promedios_centrados, frequency = 4, start(c(2016,1)))

plot.ts(f,
        col = "blue",
        main = 'Ventas de televisores por trimestre',
        lwd = 1.5)
```



Aqui es en donde se puede observar de mejor forma la tendecia de aumentar la ventas conforme pasa el tiempo.

Una vez obtenida la gráfica anterior, se pasará a obtener los valores estacionales irregulares. Estos se calculan dividiendo el valor de las ventas entre su respectivo promedio móvil centrado.

```
## Cálculo de los valores estacionales irregulares
sub_ventas = ventas[3:14]
val_estacionales_irregulares = sub_ventas/promedios_centrados[5:16]
cat('Los valores estacionales irregulares son:\n\n', val_estacionales_irregulares)
```

```
## Los valores estacionales irregulares son:
##
##  1.09589 1.132898 0.9707113 0.840404 1.075099 1.15625 0.917782 0.8389513 1.109057 1.1
40768 0.9081081 0.8339223
```

Ya obtenidos dichos valores, ahora se agrupan por su similitud y se saca el promedio de cada uno de los grupos para obtener el índice estacional.

```r
trimestre_1 = c(0.9707113, 0.917782, 0.9081081)
trimestre_2 = c(0.840404, 0.8389513, 0.8339223)
trimestre_3 = c(1.09589, 1.075099, 1.109057)
trimestre_4 = c(1.132898, 1.15625, 1.140768)

## índices estacionales
cat('Índice estacional del primer trimestre:', mean(trimestre_1))
```

```
## Índice estacional del primer trimestre: 0.9322005
```

```r
cat('\nÍndice estacional del segundo trimestre:', mean(trimestre_2))
```

```
##
## Índice estacional del segundo trimestre: 0.8377592
```

```r
cat('\nÍndice estacional del tercer trimestre:', mean(trimestre_3))
```

```
##
## Índice estacional del tercer trimestre: 1.093349
```

```r
cat('\nÍndice estacional del cuarto trimestre:', mean(trimestre_4))
```

```
##
## Índice estacional del cuarto trimestre: 1.143305
```

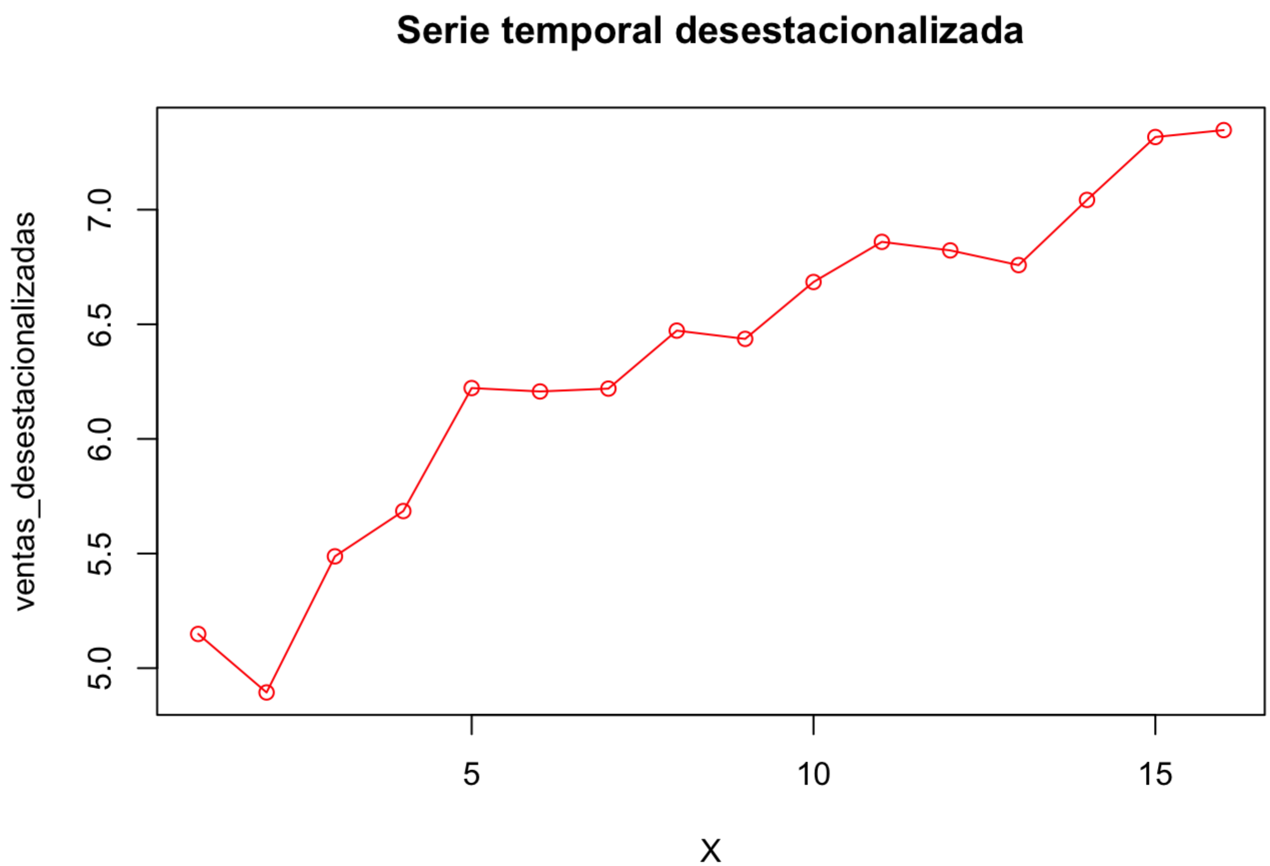
Con los índices estacionales, finalmente se puede obtener las ventas desestacionalizadas dividiendo las ventas entre su ínidice correspondiente a su trimestre.

```r
indices = c(0.9322005, 0.8377592, 1.093349, 1.143305, 0.9322005, 0.8377592, 1.093349, 1.
143305, 0.9322005, 0.8377592, 1.093349, 1.143305, 0.9322005, 0.8377592, 1.093349, 1.1433
05)


ventas_desestacionalizadas = ventas/indices
ventas_desestacionalizadas
```

```
##  [1] 5.149107 4.894008 5.487726 5.685272 6.221837 6.207034 6.219423 6.472464
##  [9] 6.436384 6.684498 6.859658 6.822327 6.758203 7.042596 7.316968 7.347121
```

Con estos nuevos valores, se puede graficar la serie de tiempo desestacionalizada que se ve de la siguiente forma:

```r
X = 1:16
plot(X, ventas_desestacionalizadas, type = 'o' , col = 'red', main = 'Serie temporal des
estacionalizada')
```



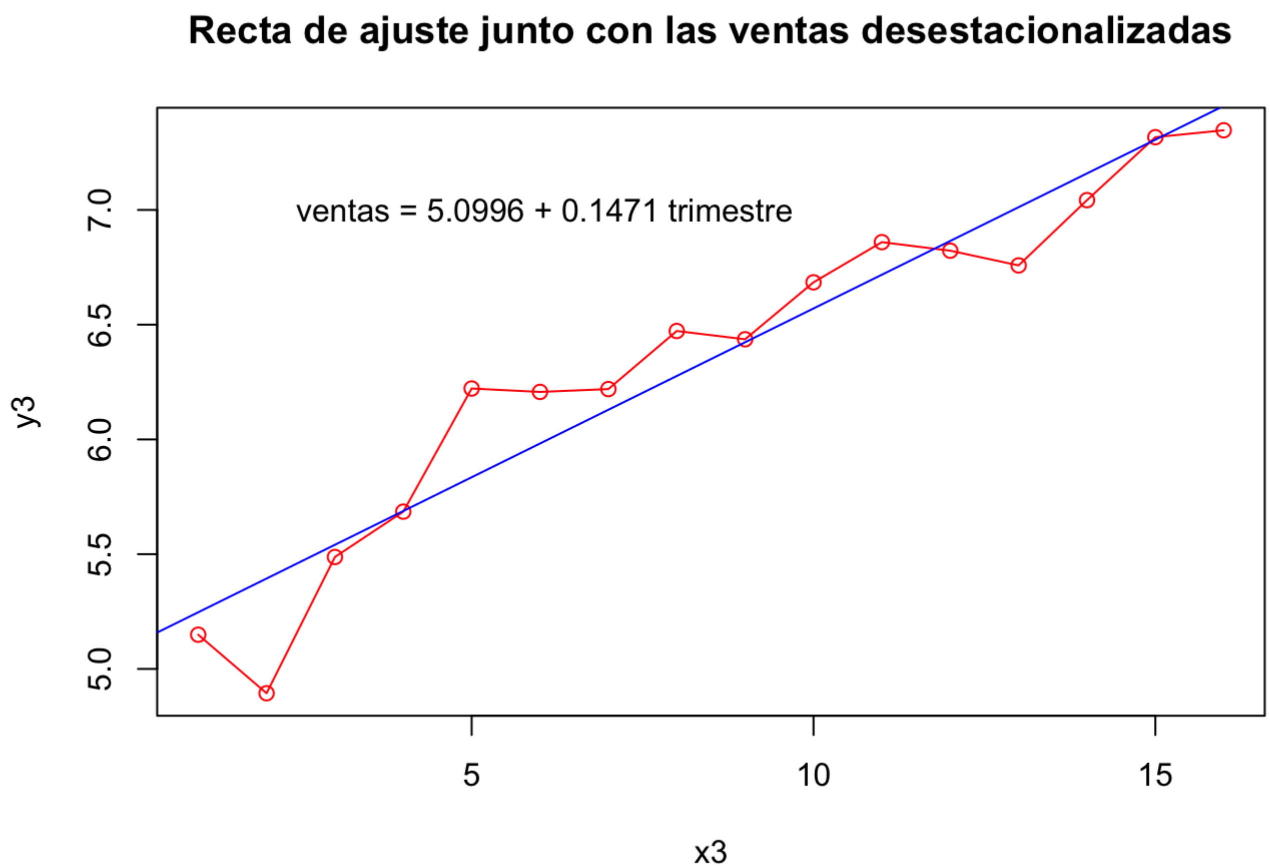
Como se observa, es un suavizamiento diferente al que se tenía originalmente con promedios móviles. Una vez hecho lo anterior, ahora si se obtendrá el modelo de regresión lineal que se obtienen a través de las ventas desestacionalizadas.

```
x3 = 1:16
y3 = ventas_desestacionalizadas
N3 = lm(y3~x3)
N3
```

```
##
## Call:
## lm(formula = y3 ~ x3)
##
## Coefficients:
## (Intercept)           x3
##      5.0996       0.1471
```

Dibuja la recta junto con las ventas desestacionalizadas.

```
plot(x3, y3, type = "o", col="red", main = "Recta de ajuste junto con las ventas desesta
cionalizadas")
abline(N3, col = "blue")
text(6, 7, " ventas = 5.0996 + 0.1471 trimestre")
```



Claramente se observa que la recta de predicción no es la mejor estimación para la serie de tiempo, sin embargo se aproxima bastante a ella, por lo que los siguientes pasos serán analizar justament la pertinencia del modelo generado.

Significancia de β_1

Una vez construido el modelo, es importante verificar la significancia de β_1 dado a que este no puede ser igual a 0, ya que como se multiplica con la variable independiente , si es 0, el producto de igual forma daría 0 inidcando así independencia entre las variables.

Para verificar justamente que el valor del estimador β_1 sea significativo para el modelo, realizamos una prueba de hipótesis de la siguiente forma:

- $H_0 : \beta_1 = 0$
- $H_1 : \beta_1 \neq 0$

En este caso, la regla de decisión sería que el valor del estadístico de prueba $|t*|$ sea mayor que $|t_0|$ y que el valor p sea menor que 0.05, dado a que es el nivel de significancia para la prueba.

```
# Para ambos modelos
alpha = 0.05
t0 = qt(alpha/2, 14)
cat('t0=', t0)
```

```
## t0= -2.144787
```

Por lo tanto, el valor de $|t*|$ debe de ser mayor que $|-2.144787|$.

```
# Resumen del modelo
summary(N3)
```

```
##
## Call:
## lm(formula = y3 ~ x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49988 -0.09991  0.00369  0.12051  0.38653
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.09961    0.11153   45.73  < 2e-16 ***
## x3           0.14714    0.01153   12.76 4.25e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2127 on 14 degrees of freedom
## Multiple R-squared:  0.9208, Adjusted R-squared:  0.9151
## F-statistic: 162.7 on 1 and 14 DF,  p-value: 4.248e-09
```

Según el resumen anterior, el valor de $|t*| = 12.70$, mientras que el $p = 4.51e-09$ dejando asi claro que se rechaza la hipótesis nula.

Analiza la pertinencia del modelo lineal:

De igual forma, del resumen anterior se puede visualizar que el modelo tiene un coeficiente de determinación de 0.92, lo cual inidica que es bastante adecuado para predecir lo valores de la variable predictora.

Análisis de los residuos

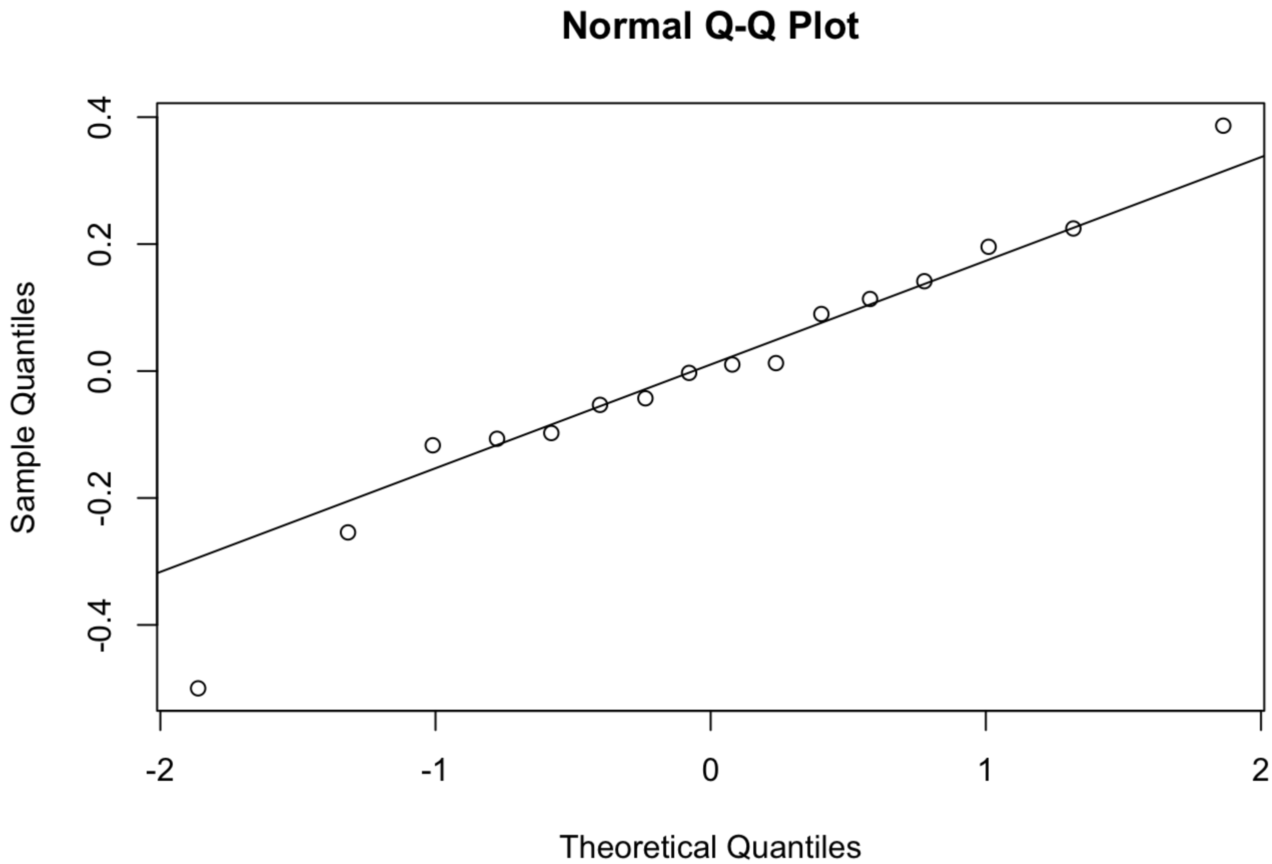
Prueba de normalidad

Para ver cómo se distribuyen los residuos, se puede plantear una prueba de hipótesis, en donde: H_0 : los datos provienen de una población normal H_1 : los datos no provienen de una población normal

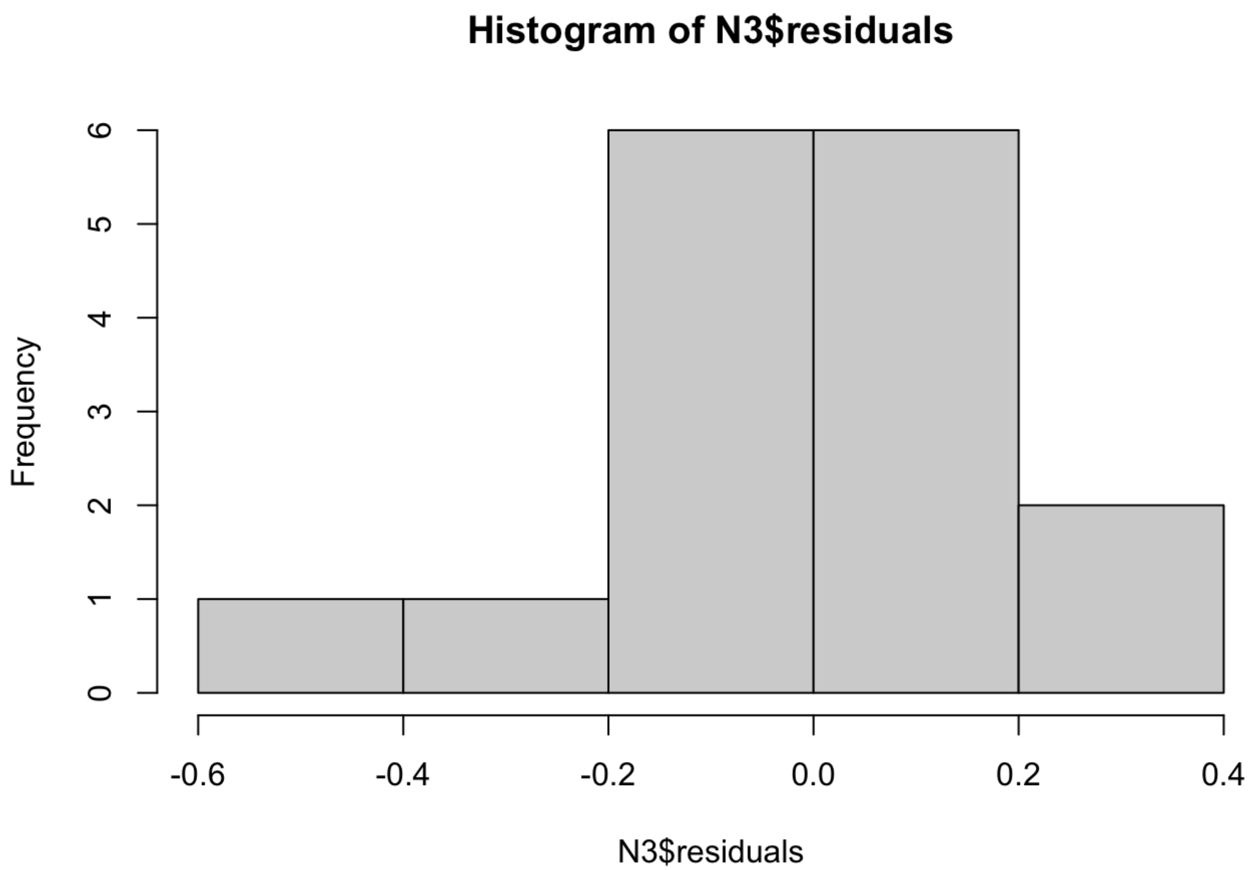
```
# Prueba de normalidad
shapiro.test(N3$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  N3$residuals
## W = 0.96379, p-value = 0.7307
```

```
#Gráficas auxiliares:
qqnorm(N3$residuals)
qqline(N3$residuals)
```




```
hist(N3$residuals)
```



Se observa en el Q-Q plot que los residuos están en una línea recta, siendo así la forma ideal de los residuos cuando pertenecen a una población normal.

Media de los residuos igual a 0

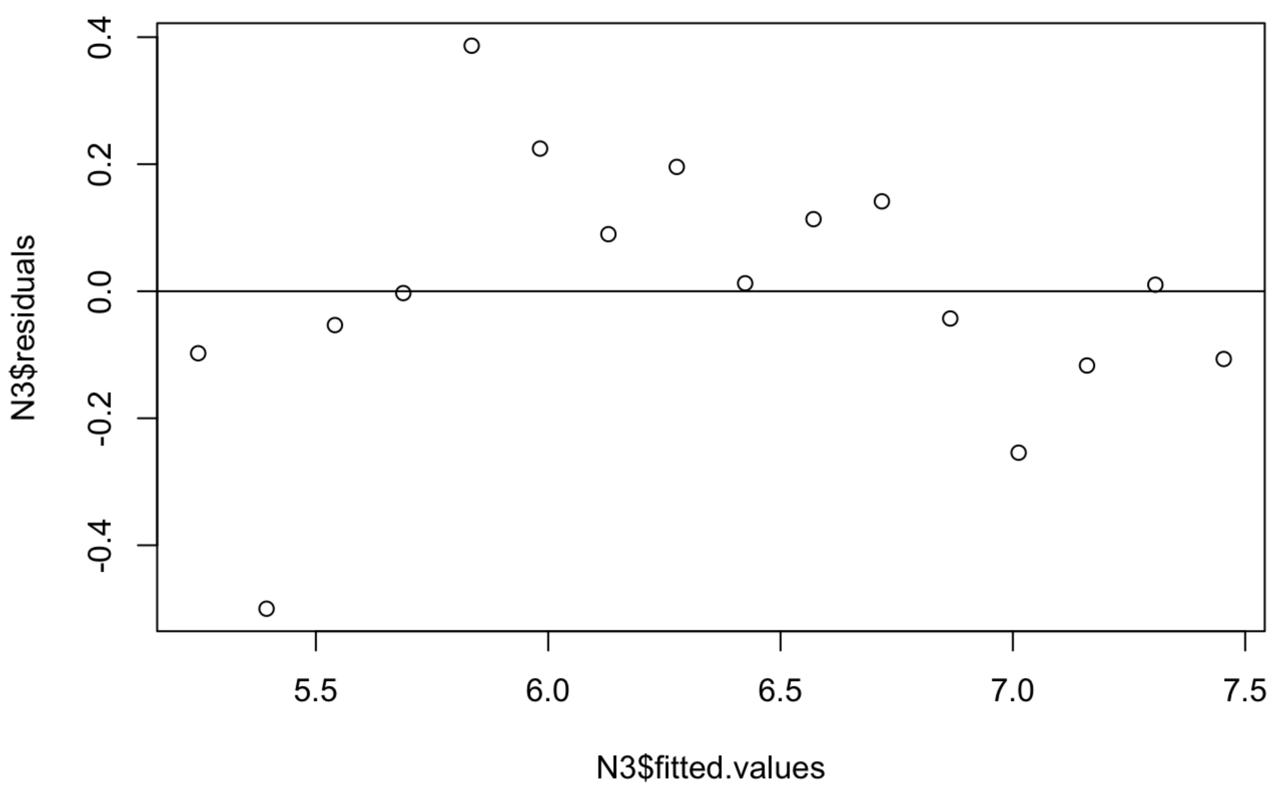
Para la suma de los resiudos, se puede aplicar de igual forma una prueba de hipótesis, o obtener un intervalo de confianza.

```
t.test(N3$residuals)
```

```
##
##  One Sample t-test
##
## data:  N3$residuals
## t = 1.3931e-16, df = 15, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.1094818  0.1094818
## sample estimates:
##    mean of x
## 7.155734e-18
```

Homocedasticidad

```
plot(N3$fitted.values, N3$residuals)
abline(h=0)
```



Se observa que los residuos siguen un patrón como tal, o más bien no muestran una estructura evidente. Por lo tanto el modelo es adecuado ya que se muestra independencia.

Calcula el CME y el EPAM (promedio de los errores porcentuales) de la predicción de la serie de tiempo.

Ahora se pasará a calcular el CME y el EPAM con ayuda de las predicciones arrojadas por el modelo lineal construido.

```
trimestres = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16)
predicciones = (trimestres  * 0.1471 + 5.0966) * indices
e = ventas - predicciones

CME2 = mean(e^2, na.rm = TRUE)
cat("El error cuadrático medio de los pronósticos es de:", CME2)
```

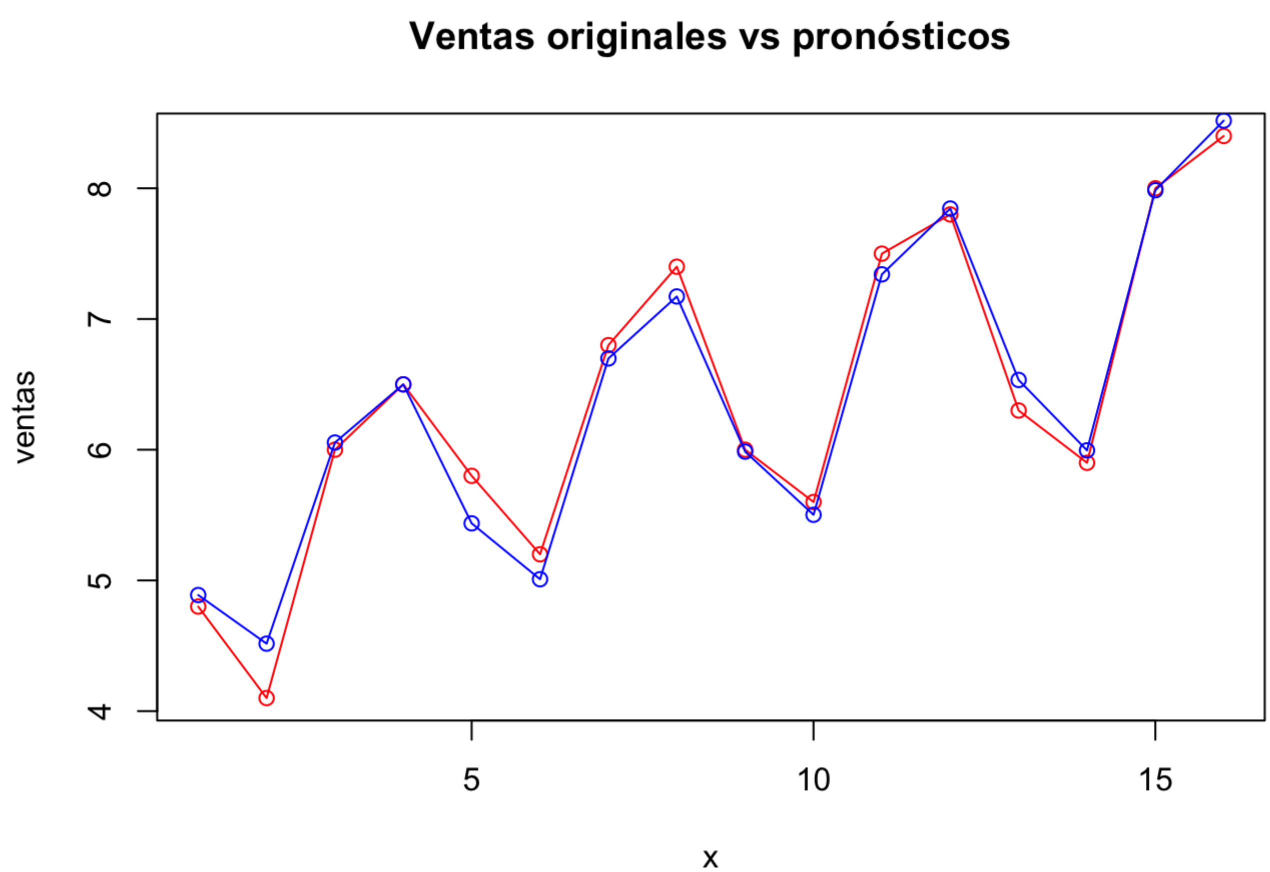
```
## El error cuadrático medio de los pronósticos es de: 0.03308245
```

```
e_porcentual = abs(e/ventas)
cat("El promedio de los errores porcentuales es:", mean(e_porcentual))
```

```
## El promedio de los errores porcentuales es: 0.02438393
```

Dibuja el gráfico de los valores de las ventas y las predicciones vs el tiempo

```
x = 1:16

plot(x, ventas, type= 'o', col = 'red', main = "Ventas originales vs pronósticos")
lines(x, predicciones, type = 'o', col = 'blue')
```



Concluye sobre el modelo

En conclusión, el modelo lineal generado es bastante bueno, ya que además de superar con éxito todos los supuestos, tiene un alto valor de coeficiente de detemrinación. Claramente el modelo se aplicó a las ventas desestabilizadas por lo tanto al trabajar con series de tiempo siempre es buena idea aplicar métodos de suavizamiento y posteriormente los modelos correspondientes.

De igual forma, con los valores de CME y EPAM, se observa que son bastante bajos, lo cual indica que los pronósticos se están realizando de manera stisfactoria.

Realiza el pronóstico para el siguiente año.

```
## Se realizan las predicciones correspondientes al siguiente año con ayuda del modelo
f = function(x) {5.0996 + 0.1471*x}
# Los ídices estacionales son:
a1 = 0.9322005
a2 = 0.8377592
a3 = 1.093349
a4 = 1.143305;

cat("Pronóstico del primer trimestre del quinto año:", f(17)*a1*1000)
```

```
## Pronóstico del primer trimestre del quinto año: 7085.003
```

```
cat("\nPronóstico del segundo trimestre del quinto año:", f(18)*a2*1000)
```

```
##
## Pronóstico del segundo trimestre del quinto año: 6490.456
```

```
cat("\nPronóstico del tercer trimestre del quinto año:", f(19)*a3*1000)
```

```
##
## Pronóstico del tercer trimestre del quinto año: 8631.444
```

```
cat("\nPronóstico del cuarto trimestre del quinto año:", f(20)*a4*1000)
```

```
##
## Pronóstico del cuarto trimestre del quinto año: 9194.001
```

Los resultados anteriores son las ventas en miles de pesos que se estiman para los siguientes cuatro trimestres. Tdos esto fue utilizando el modelo de regresión lineal construido tomando como base las ventas desestacionalizadas.

Realiza el problema de “Un problemilla más” sobre las ventas trimestraless de libros de texto universitarios.

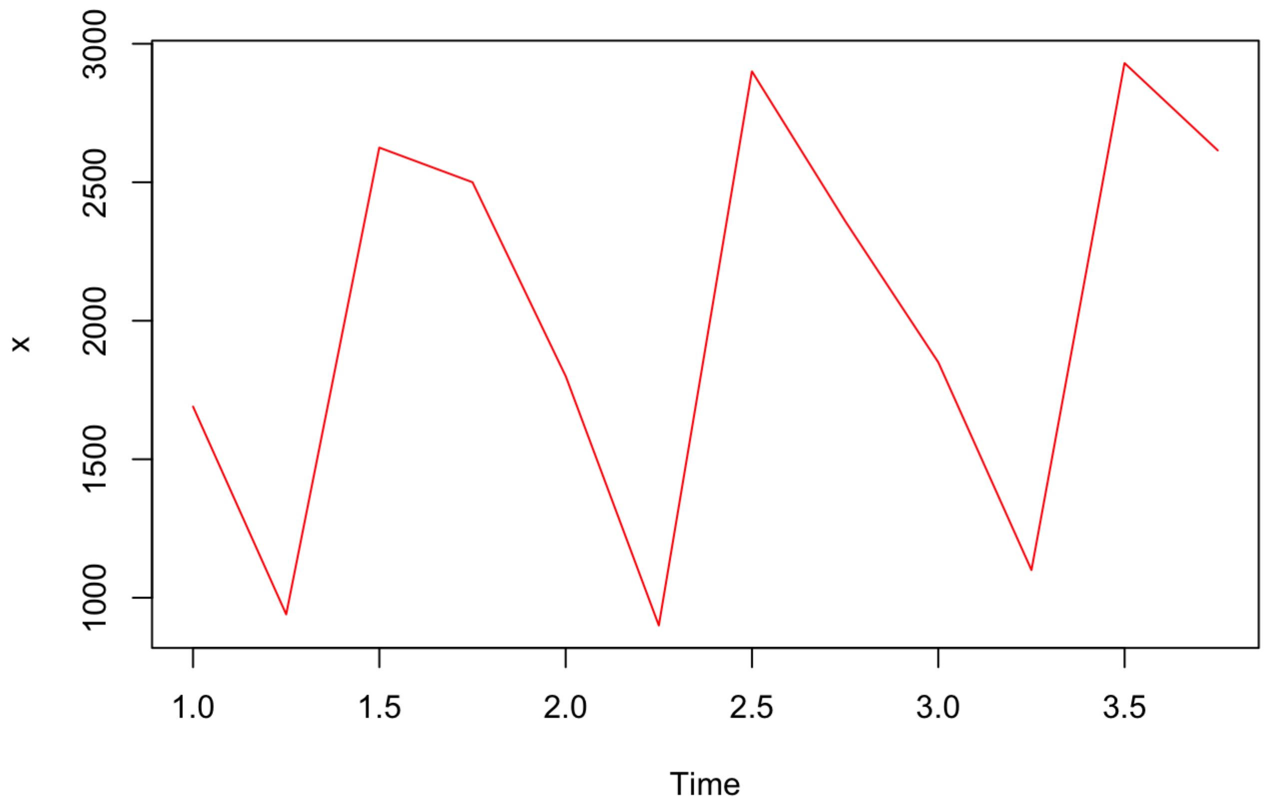
A continuación, se presentan los datos correspondientes a los últimos tres años de ventas trimestrales (número de ejemplares vendidos) de un libro de texto universitario.

Trimestre Año 1 Año 2 Año 3 1 1690 1800 1850 2 940 900 1100 3 2625 2900 2930 4 2500 2360 261

Encuentre los promedios móviles de cuatro trimestres y los promedios móviles

centrados

```
ventas = c(1690, 940, 2625, 2500, 1800, 900, 2900, 2360, 1850, 1100, 2930, 2615)
x= ts(ventas, frequency = 4, start(c(2016,1)))
plot.ts(x, col = "red")
```



Para esto, realizamos el mismo procedimiento que en el problema anterior.

```
k = 12 # Número de observaciones del data frame
n = 4 # Número de promedios móviles

moviles = c()

for(i in 1:(k-n+1)){
  moviles[i] = (ventas[i] + ventas[i+1] + ventas[i+2] + ventas[i+3]) / n
}

## Se obtienen los promedios móviles
moviles
```

```
## [1] 1938.75 1966.25 1956.25 2025.00 1990.00 2002.50 2052.50 2060.00 2123.75
```

```
centrados = c()

for(i in 1:8){
  centrados[i] = (moviles[i] + moviles[i+1]) / 2
}

## Se obtienen los promedios moviles centrados
centrados
```

```
## [1] 1952.500 1961.250 1990.625 2007.500 1996.250 2027.500 2056.250 2091.875
```

Calcule los índices estacionales de los cuatro trimestres

```
## Se obtienen los valores estacionales irregulares

sub_ventas = ventas[3:10]
irregulares = sub_ventas/centrados
irregulares
```

```
## [1] 1.3444302 1.2746973 0.9042386 0.4483188 1.4527239 1.1639951 0.8996960
## [8] 0.5258440
```

```
## Finalmente se obtienen los índices estacionarios
q1 = (irregulares[3] + irregulares[7])/2
q2 = (irregulares[4] + irregulares[8])/2
q3 = (irregulares[1] + irregulares[5])/2
q4 = (irregulares[2] + irregulares[6])/2

q1
```

```
## [1] 0.9019673
```

```
q2
```

```
## [1] 0.4870814
```

```
q3
```

```
## [1] 1.398577
```

```
q4
```

```
## [1] 1.219346
```

¿Cuándo obtiene la editorial el mayor índice estacional? ¿Parece razonable este

resultado? ¿Por qué?

Se puede observar que el mayor índice estacional de la editorial es en el tercer trimestre. Esto se puede observar de igual forma en la gráfica de la serie de tiempo correspondiente, en donde se ve un pico en el grafo cada vez este trimestre.

Este resultado parece bastante razonable, ya que como se están analizando ventas trimestrales de un libro de texto univesrsitario, se sabe que la mayoría de ciclos escolares comienzan en el mes de agosto, el cual pertenece al tercer trimestre del año. Por lo tanto son en estas fechas en donde más ejemplares se venden.