Instituto Tecnológico y de Estudios Superiores de Monterrey

Actividad 8. Series de tiempo no estacionarias. Tendencia

TC3007C.501 Inteligencia artificial avanzada para la ciencia de datos II

Profesores:

Iván Mauricio Amaya Contreras

Blanca Rosa Ruiz Hernández

Félix Ricardo Botello Urrutia

Edgar Covantes Osuna

Felipe Castillo Rendón

Hugo Terashima Marín

Alumno:

Alberto H Orozco Ramos - A00831719

17 de Noviembre de 2023

Actividad 8: Series de tiempo no estacionarias. Tendencia

Instrucciones

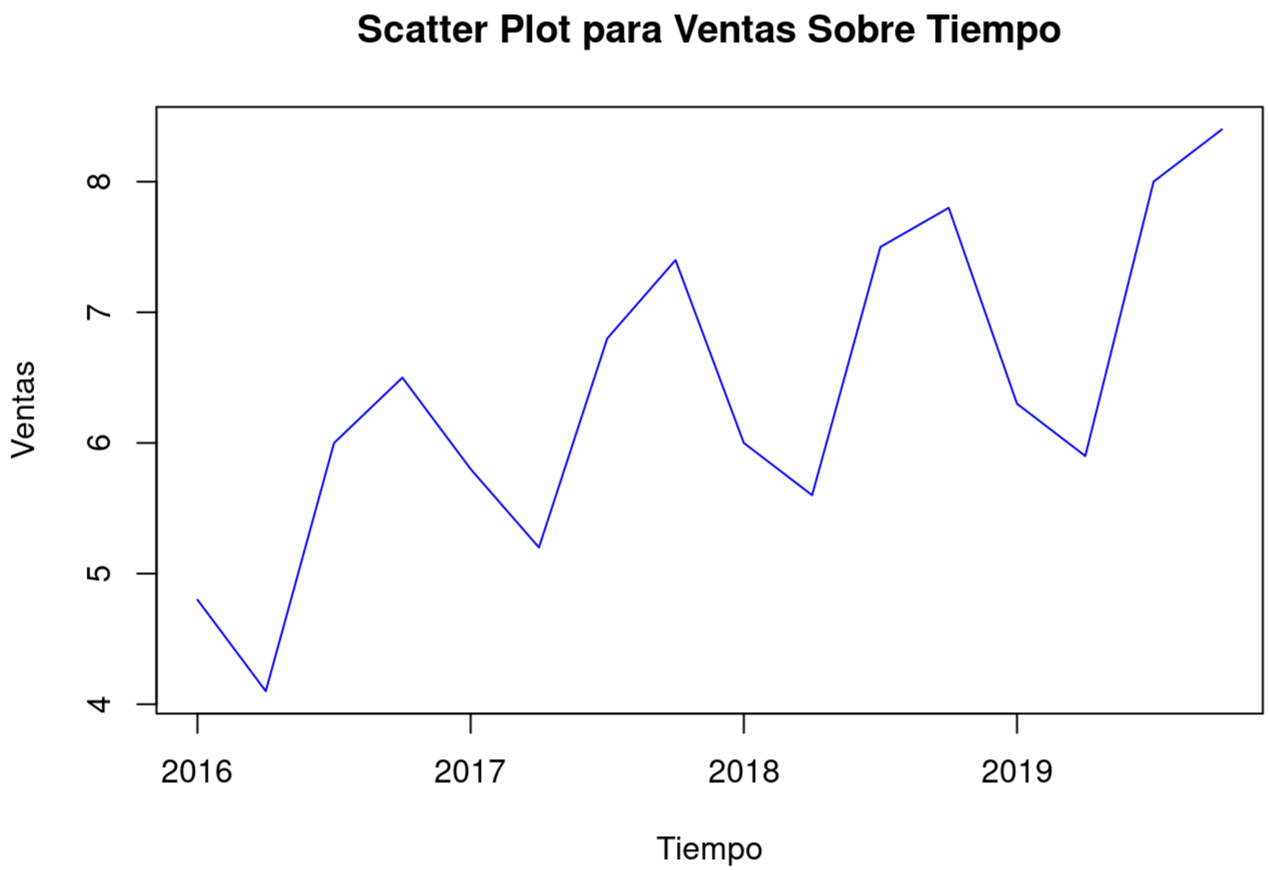
Problema 1

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

Año	Trimestre	Ventas (miles)
1	1	4.8
-	2	4.1
-	3	6.0
-	4	6.5
2	1	5.8
-	2	5.2
-	3	6.8
-	4	7.4
3	1	6.0
-	2	5.6
-	3	7.5
-	4	7.8
4	1	6.3
-	2	5.9
-	3	8.0
-	4	8.4

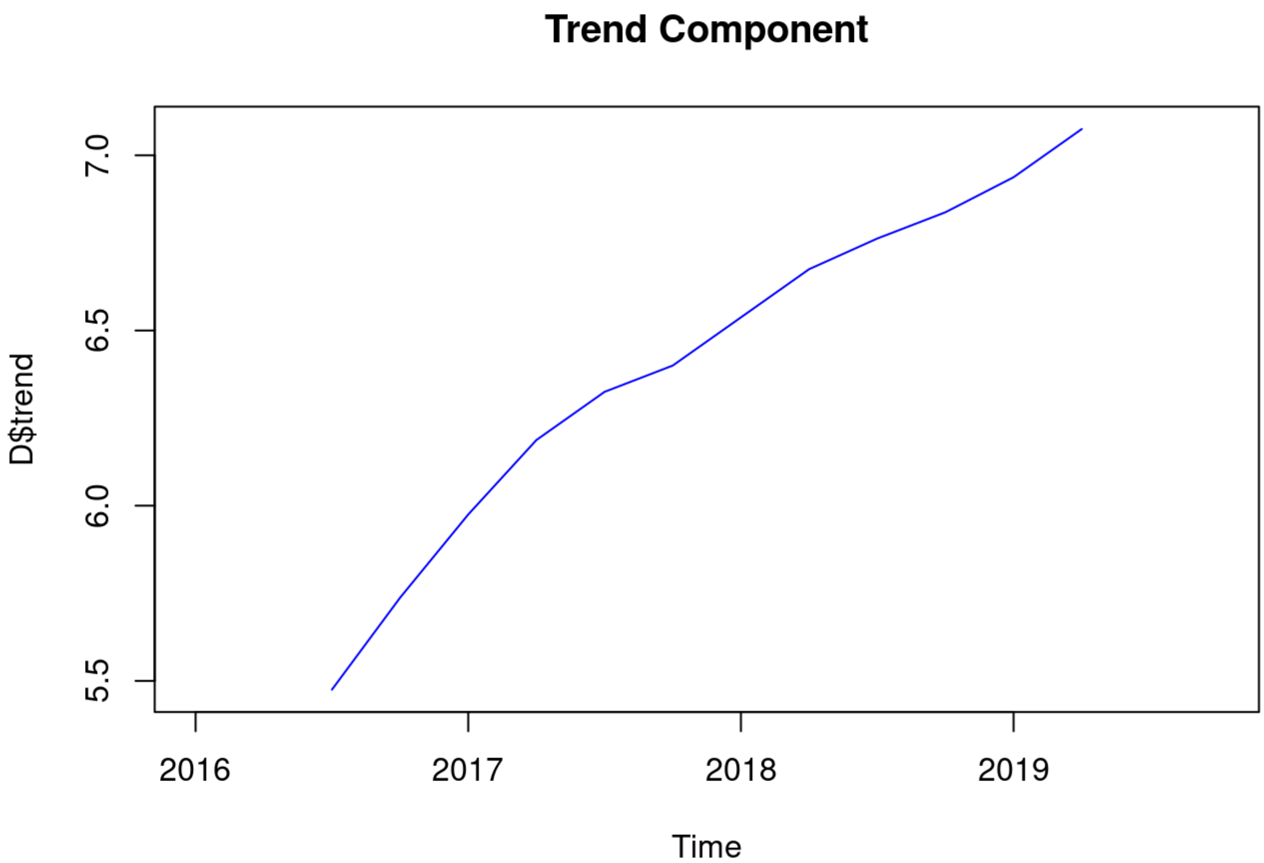
- Realiza el gráfico de dispersión. Observa la tendencia y los ciclos.

```
T = ts(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), freque
ncy = 4, start = c(2016, 1))
plot(T, main = "Scatter Plot para Ventas Sobre Tiempo", xlab = "Tiempo", ylab = "Ventas", col =
"blue", pch = 19)
```

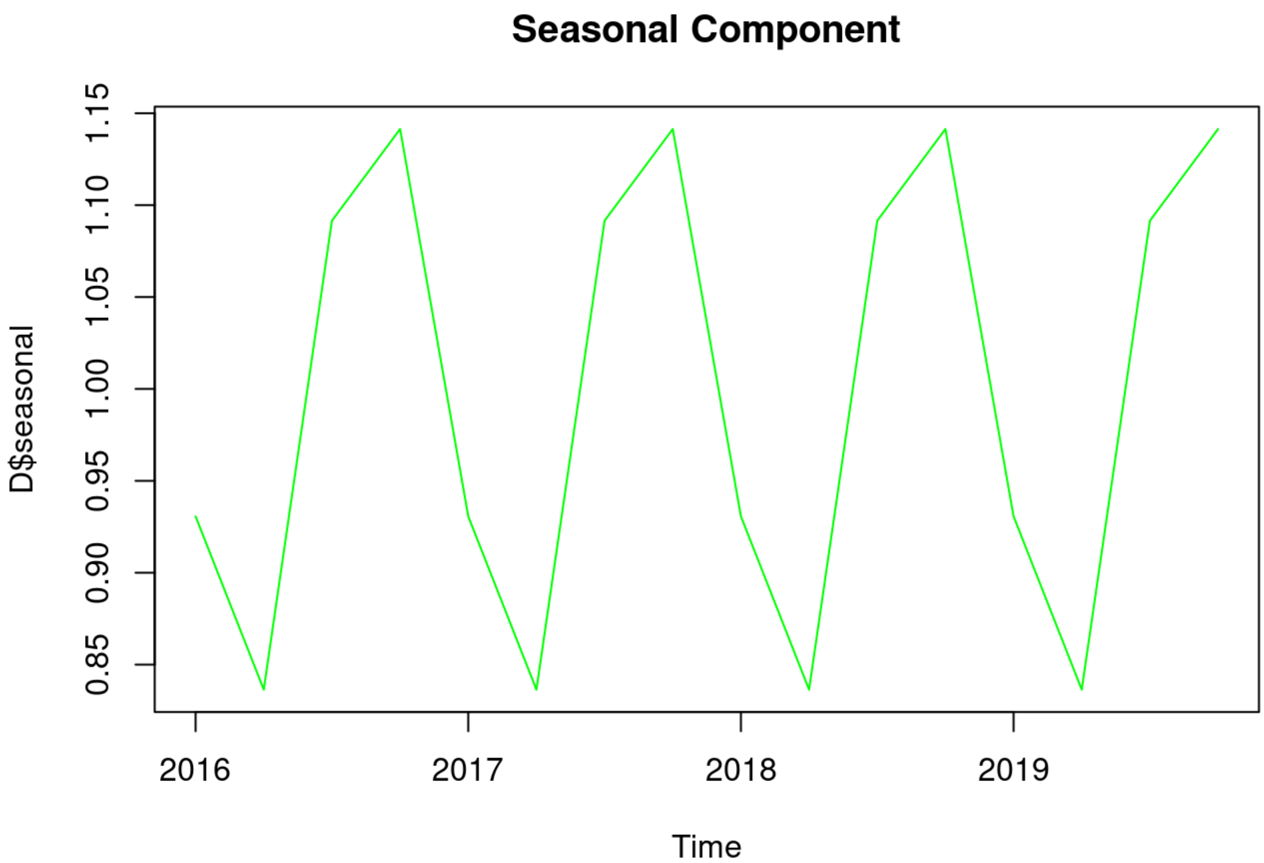


- Realiza el análisis de tendencia y estacionalidad
    - Descompón la serie en sus 3 componentes e interprétalos

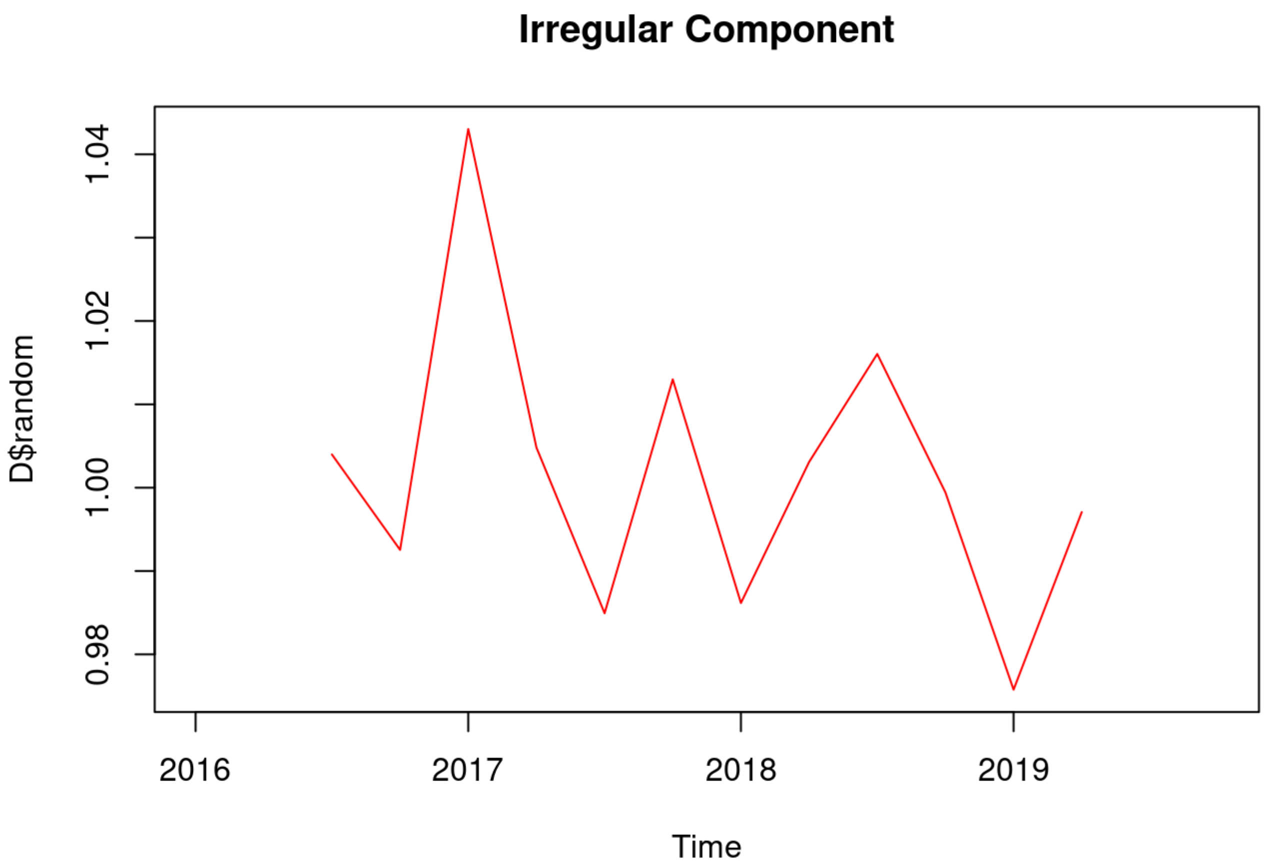
```
D = decompose(T, type = "multiplicative")
plot(D$trend, main = "Trend Component", col = "blue", type = "l")
```




```
plot(D$seasonal, main = "Seasonal Component", col = "green", type = "l")
```




```
plot(D$random, main = "Irregular Component", col = "red", type = "l")
```


Irregular Component
D$random
1.04
1.02
1.00
0.98
2016
2017
2018
2019
Time

```
D
```

```
## $x
##      Qtr1 Qtr2 Qtr3 Qtr4
## 2016  4.8  4.1  6.0  6.5
## 2017  5.8  5.2  6.8  7.4
## 2018  6.0  5.6  7.5  7.8
## 2019  6.3  5.9  8.0  8.4
##
## $seasonal
##           Qtr1      Qtr2      Qtr3      Qtr4
## 2016 0.9306617 0.8363763 1.0915441 1.1414179
## 2017 0.9306617 0.8363763 1.0915441 1.1414179
## 2018 0.9306617 0.8363763 1.0915441 1.1414179
## 2019 0.9306617 0.8363763 1.0915441 1.1414179
##
## $trend
##        Qtr1   Qtr2   Qtr3   Qtr4
## 2016     NA     NA 5.4750 5.7375
## 2017 5.9750 6.1875 6.3250 6.4000
## 2018 6.5375 6.6750 6.7625 6.8375
## 2019 6.9375 7.0750     NA     NA
##
## $random
##           Qtr1      Qtr2      Qtr3      Qtr4
## 2016        NA        NA 1.0039818 0.9925353
## 2017 1.0430335 1.0048157 0.9849340 1.0129944
## 2018 0.9861607 1.0030787 1.0160445 0.9994305
## 2019 0.9757661 0.9970658        NA        NA
##
## $figure
## [1] 0.9306617 0.8363763 1.0915441 1.1414179
##
## $type
## [1] "multiplicative"
##
## attr(,"class")
## [1] "decomposed.ts"
```

1. Componente Estacional:

- Representa el patrón regular y repetitivo en los datos que ocurre en momentos específicos del año (trimestres, en este caso). Cada fila corresponde a un año, y cada columna corresponde a un trimestre. Los valores en cada celda representan el efecto multiplicativo del componente estacional en las ventas totales para ese trimestre. Por ejemplo, un valor de 0.9306617 en la primera celda (Qtr1, 2016) significa que las ventas en el primer trimestre de ese año son aproximadamente el 93.07% del promedio de ventas para todos los trimestres.

2. Componente de Tendencia:

- Representa la dirección a largo plazo o el movimiento en los datos. Los valores en cada celda representan la tendencia para ese trimestre en el año respectivo. Los valores "NA" indican que no se pudo estimar la tendencia para esos trimestres en esos años. Por ejemplo, en 2016, la tendencia comienza a estimarse desde el tercer trimestre y aumenta en los años siguientes.

3. Componente Irregular (Aleatorio):

- El componente irregular, también conocido como componente aleatorio o residual, captura la variación no explicada en los datos después de tener en cuenta los componentes de tendencia y estacionales. Los valores en cada celda representan las fluctuaciones aleatorias alrededor de las tendencias y patrones estacionales estimados. Los valores "NA" indican que no se pudo estimar el componente irregular para esos trimestres en esos años.

- Analiza el modelo lineal de la tendencia:
  - Realiza la regresión lineal de la tendencia (ventas desestacionalizadas vs tiempo)
  - Dibuja la recta junto con las ventas desestacionalizadas.
    - Analiza la pertinencia del modelo lineal:
    - Significancia de $\beta 1$
    - Variabilidad explicada por el modelo (c.d)
    - Análisis de los residuos
    - Prueba de normalidad

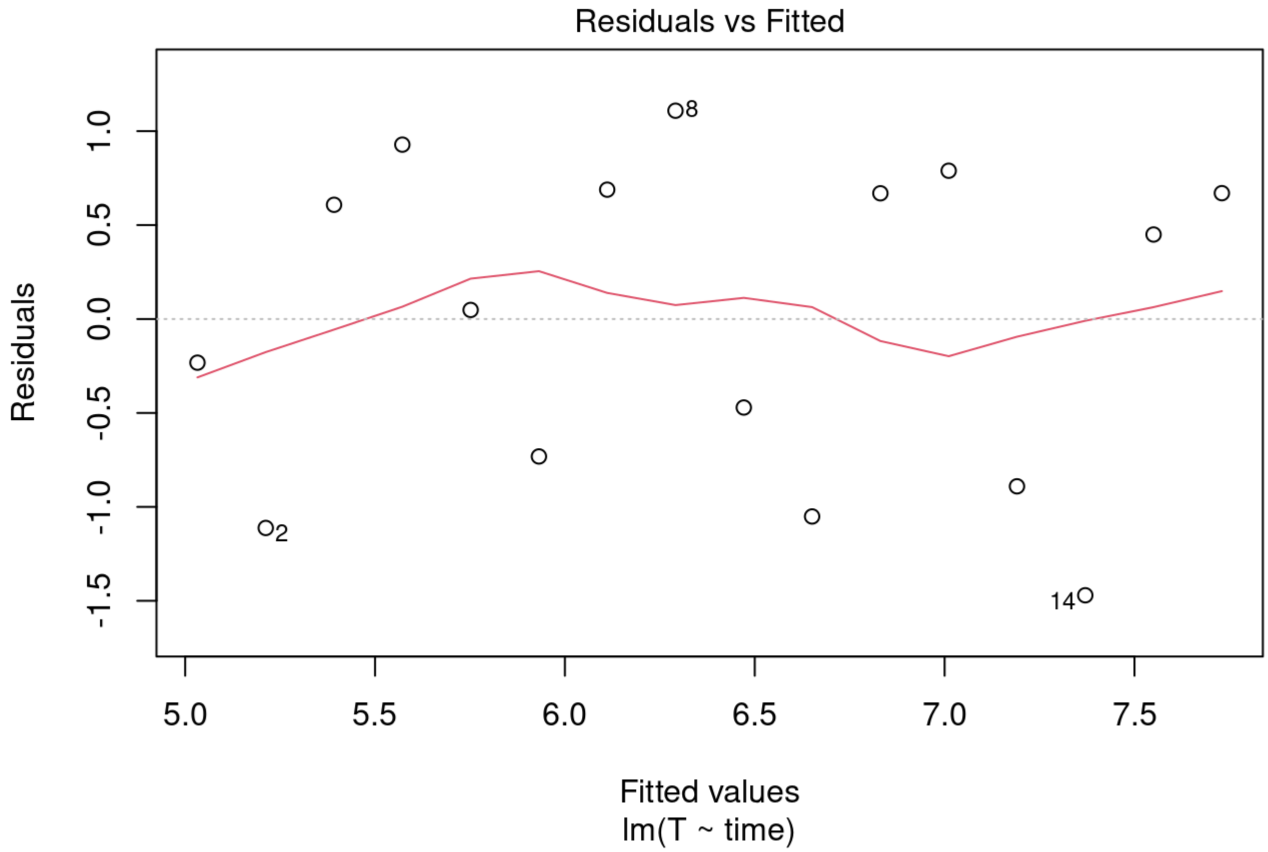
```
time <- 1:length(T)
model <- lm(T ~ time)

summary(model)
```

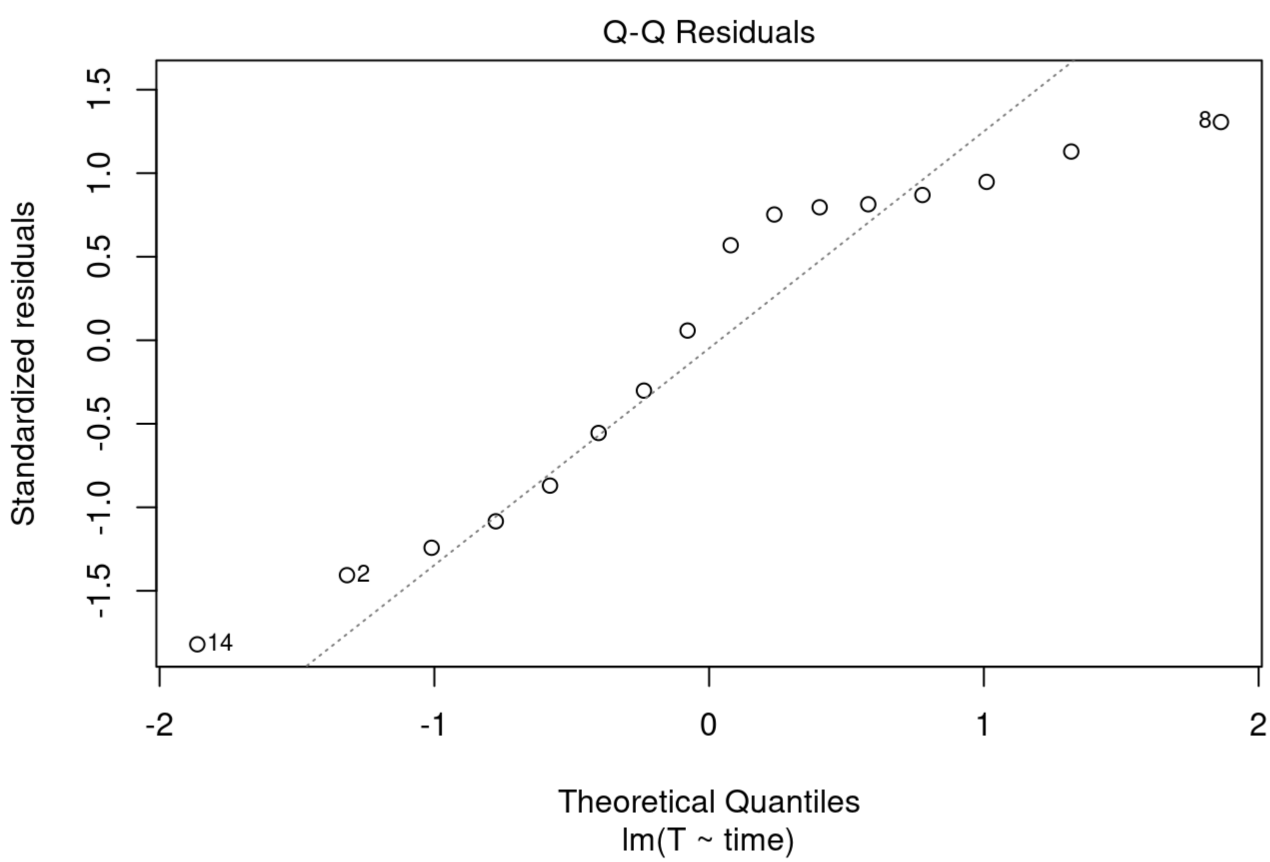
```
##
## Call:
## lm(formula = T ~ time)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4704 -0.7714  0.2490  0.6745  1.1087
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.85250    0.45987  10.552 4.78e-08 ***
## time         0.17985    0.04756   3.782  0.00202 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8769 on 14 degrees of freedom
## Multiple R-squared:  0.5053, Adjusted R-squared:   0.47
## F-statistic:  14.3 on 1 and 14 DF,  p-value: 0.002023
```

```
summary(model)$r.squared
```

```
## [1] 0.5053215
```

```
plot(model, which = 1:2)
```


Residuals vs Fitted
Residuals
1.0
0.5
0.0
-0.5
-1.0
-1.5
8
2
14
5.0
5.5
6.0
6.5
7.0
7.5
Fitted values
lm(T ~ time)


Q-Q Residuals
Standardized residuals
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
8
2
14
-2
-1
0
1
2
Theoretical Quantiles
lm(T ~ time)

```
shapiro.test(model$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.90452, p-value = 0.09498
```

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

```
predictions <- predict(model)
errors <- T - predictions
CME <- mean(errors^2, na.rm = TRUE)
EPAM <- mean(abs(errors / T) * 100, na.rm = TRUE)
cat("CME:", CME, "\n")
```

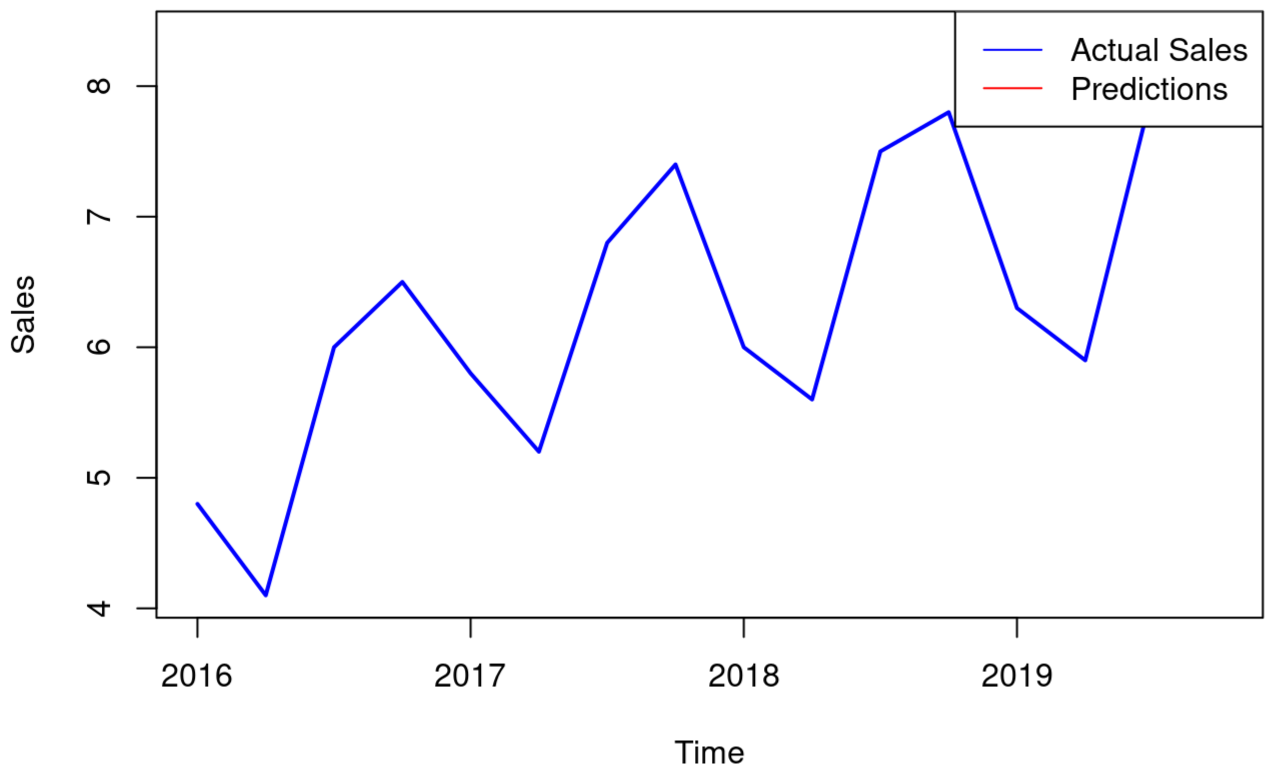
```
## CME: 0.672898
```

```
cat("EPAM:", EPAM)
```

```
## EPAM: 12.16897
```

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

```
plot(T, col = "blue", type = "l", lwd = 2, ylab = "Sales")
lines(predictions, col = "red", lwd = 2)
legend("topright", legend = c("Actual Sales", "Predictions"), col = c("blue", "red"), lty = 1:1)
```



- Concluye sobre el modelo: de acuerdo al análisis de verificación de los supuestos, ¿es el mejor modelo que puedes obtener?

Los residuos (errores) parecen estar razonablemente centrados alrededor de cero, lo que indica que el modelo está capturando parte de la variación en los datos. Tanto el intercepto como el coeficiente de tiempo son estadísticamente significativos (valores p bajos), lo que sugiere que existe una relación lineal significativa entre el tiempo y las ventas. El modelo explica una parte sustancial de la varianza de los datos (R cuadrado ajustado = 0,47) y los coeficientes son estadísticamente significativos, sin embargo, es crucial comprobar las suposiciones. En cuanto a la prueba de Shapiro-Wilk para la normalidad de los residuos arroja un valor p de 0,09498 (> 0,05). Si bien no es significativo, sugiere una ligera desviación de la normalidad.

Es posible que los supuestos del modelo, especialmente los relacionados con la normalidad de los residuos, no se cumplan por completo.

- Propón un posible mejor modelo para la tendencia de los datos.

Dado el posible problema con la normalidad en los residuos, es posible que desee explorar transformaciones o modelos no lineales para abordar este problema. No sería mal enfoque probar una regresión polinómica u otros modelos de regresión no lineal. Alternativamente, podría experimentar transformando la variable dependiente (ventas) utilizando métodos como la transformación de Box-Cox para inducir normalidad en los residuos.

Sugerencias para mejorar:

- Experimentar con diferentes formas funcionales para el componente de tendencia.
- Considerar explorar modelos de series temporales como ARIMA o SARIMA, especialmente si hay evidencia de estacionalidad en los datos.

- Realiza el pronóstico para el siguiente año.

```
future_time <- seq(length(T) + 1, length(T) + 4, by = 1)
future_predictions <- predict(model, newdata = data.frame(time = future_time))
cat(future_predictions)
```

```
## 7.91 8.089853 8.269706 8.449559
```

Problema 2

- 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	2615

```
sales <- ts(c(1690, 1800, 1850, 940, 900, 1100, 2625, 2900, 2930, 2500, 2360, 2615), frequency = 4, start = c(1, 1))
```

a. Encuentre los promedios móviles de cuatro trimestres y los promedios móviles centrados

```
# Calculamos los promedios móviles simples de 4 trimestres
four_quarter_ma <- stats::filter(sales, rep(1, 4)/4, sides = 1)

# Calculamos los promedios móviles centrados de 4 trimestres
centered_four_quarter_ma <- stats::filter(sales, rep(1, 4)/4, sides = 2)

# Desplegamos los resultados
cbind(sales, four_quarter_ma, centered_four_quarter_ma)
```

```
##      sales four_quarter_ma centered_four_quarter_ma
## 1 Q1  1690              NA                       NA
## 1 Q2  1800              NA                  1570.00
## 1 Q3  1850              NA                  1372.50
## 1 Q4   940         1570.00                  1197.50
## 2 Q1   900         1372.50                  1391.25
## 2 Q2  1100         1197.50                  1881.25
## 2 Q3  2625         1391.25                  2388.75
## 2 Q4  2900         1881.25                  2738.75
## 3 Q1  2930         2388.75                  2672.50
## 3 Q2  2500         2738.75                  2601.25
## 3 Q3  2360         2672.50                       NA
## 3 Q4  2615         2601.25                       NA
```

b. Calcule los índices estacionales de los cuatro trimestres

```
# Descomposición de las series de tiempo
decomposition <- decompose(sales, type = "multiplicative")

if (!is.matrix(decomposition$seasonal)) {
  decomposition$seasonal <- matrix(decomposition$seasonal, nrow = length(decomposition$seasona
l), ncol = 1)
}

# Calcular los índices estacionarios
seasonal_indices <- decomposition$seasonal / rowMeans(decomposition$seasonal)

# Desplegar los índices estacionarios
seasonal_indices
```

```
##       [,1]
##  [1,]    1
##  [2,]    1
##  [3,]    1
##  [4,]    1
##  [5,]    1
##  [6,]    1
##  [7,]    1
##  [8,]    1
##  [9,]    1
## [10,]    1
## [11,]    1
## [12,]    1
```

c. ¿Cuándo obtiene la editorial el mayor índice estacional? ¿Parece razonable este resultado? ¿Por qué?

Los índices estacionales de 1 sugieren que el patrón de ventas en cada trimestre es similar y no existe una variación significativa atribuida a la estacionalidad. La editorial no parece tener un trimestre específico en el que experimente consistentemente un mayor o menor impacto estacional. Este resultado podría ser razonable si el tipo de producto o servicio ofrecido por el editorial no muestra fuertes patrones estacionales.