

A8-Series de tiempo

Catherine Rojas

2024-11-05

Buscar el mejor método de predicción en una serie de tiempo dada. Utilizar métodos de suavizamiento.

Datos

```
data <- data.frame(  
  Año = rep(1:4, each = 4), # 4 años, cada uno con 4 trimestres  
  Trimestre = rep(1:4, 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)  
)
```

data

##	Año	Trimestre	Ventas
## 1	1	1	4.8
## 2	1	2	4.1
## 3	1	3	6.0
## 4	1	4	6.5
## 5	2	1	5.8
## 6	2	2	5.2
## 7	2	3	6.8
## 8	2	4	7.4
## 9	3	1	6.0
## 10	3	2	5.6
## 11	3	3	7.5
## 12	3	4	7.8
## 13	4	1	6.3
## 14	4	2	5.9
## 15	4	3	8.0
## 16	4	4	8.4

Realiza el análisis de tendencia y estacionalidad:

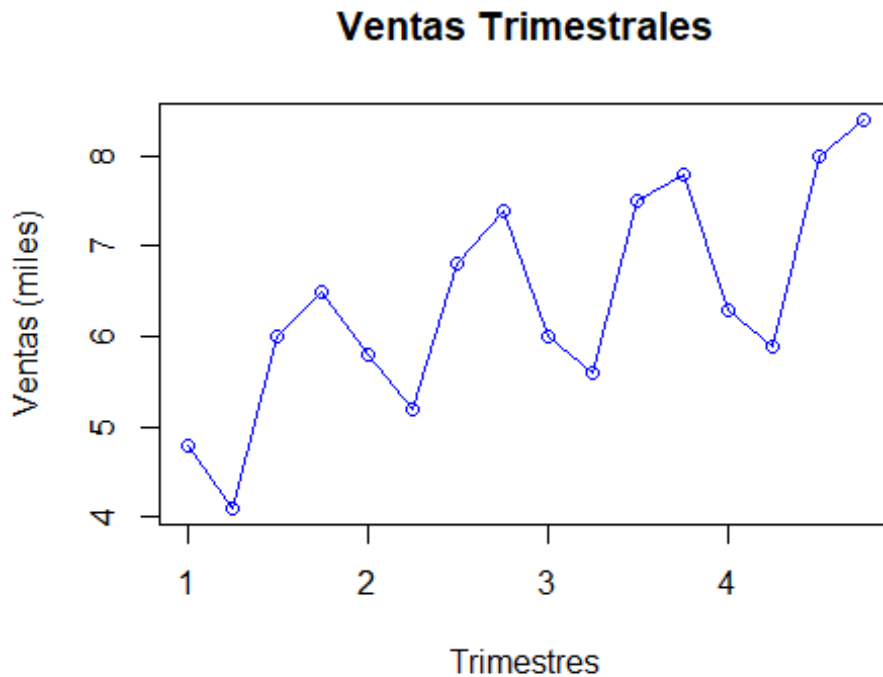
Convertir a serie temporal

```
ventas_ts <- ts(data$Ventas, start = c(1, 1), frequency = 4)
```

Grafica la serie para verificar su tendencia y estacionalidad

Graficar la serie temporal

```
plot(ventas_ts, main = "Ventas Trimestrales", xlab = "Trimestres", ylab =  
"Ventas (miles)",  
col = "blue", type = "o")
```

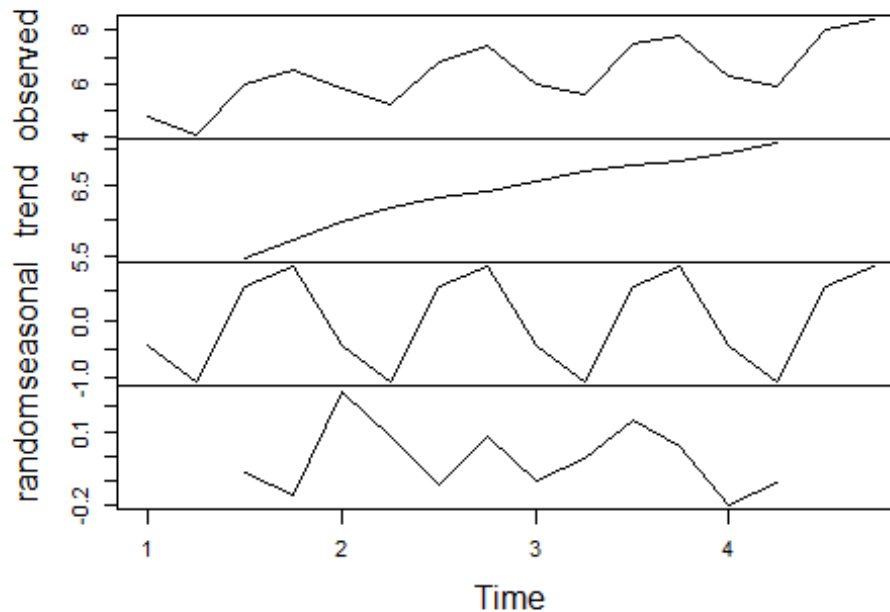


Descomponer la serie temporal para analizar tendencia y estacionalidad
descomposicion <- **decompose**(ventas_ts, type = "additive")

Graficar la descomposición de la serie temporal

```
plot(descomposicion)
```

Decomposition of additive time series



Interpretación

Ventas Trimestrales

- Este gráfico muestra la serie temporal de las ventas trimestrales a lo largo de los 4 años (16 trimestres).

Tendencia general: * Hay una tendencia creciente en las ventas a lo largo del tiempo. Esto se puede observar en el aumento paulatino del valor mínimo y máximo de las ventas en cada trimestre.

Estacionalidad: * Hay un patrón repetitivo en cada ciclo anual (4 trimestres). Las ventas tienden a caer en el segundo trimestre y aumentan hacia el tercero y cuarto trimestre.

Fluctuaciones: * Aunque hay variabilidad entre trimestres, los valores parecen seguir un comportamiento cíclico y regular.

Descomposición de la serie de tiempo

- Este gráfico descompone la serie temporal en tres componentes principales: tendencia, estacionalidad y ruido (componente aleatorio).

Componente Observado (observed): * Representa la serie original que incluye la combinación de tendencia, estacionalidad y ruido.

Tendencia (trend): * Muestra una tendencia ascendente a lo largo del tiempo, lo que confirma el crecimiento general de las ventas trimestre a trimestre.

Estacionalidad (seasonal):

- Hay un patrón estacional claro en cada ciclo de 4 trimestres:
- Durante el segundo trimestre las ventas disminuyen y en el tercer y cuarto trimestre, estas aumentan.

Componente Aleatorio (random): * Representa las variaciones residuales no explicadas por la tendencia ni la estacionalidad. * Estas variaciones son relativamente pequeñas, indicando que la mayoría del comportamiento de la serie es explicado por la tendencia y estacionalidad.

Identifica si es una serie estacionaria

- En la prueba Dickey-Fuller (ADF), el parámetro k representa el número de retardos(lag) que se utilizan para eliminar la autocorrelación en los residuos.

```
# Identifica si es una serie estacionaria
library(tseries)

## Warning: package 'tseries' was built under R version 4.3.3

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

# Realizar la prueba de estacionariedad Dickey-Fuller
adf_test <- adf.test(ventas_ts, k=0)
print(adf_test)

##
## Augmented Dickey-Fuller Test
##
## data:  ventas_ts
## Dickey-Fuller = -3.2388, Lag order = 0, p-value = 0.1004
## alternative hypothesis: stationary

# Verificar si la serie es estacionaria
if (adf_test$p.value < 0.05) {
  print("La serie es estacionaria.")
} else {
  print("La serie no es estacionaria.")
}

## [1] "La serie no es estacionaria."
```

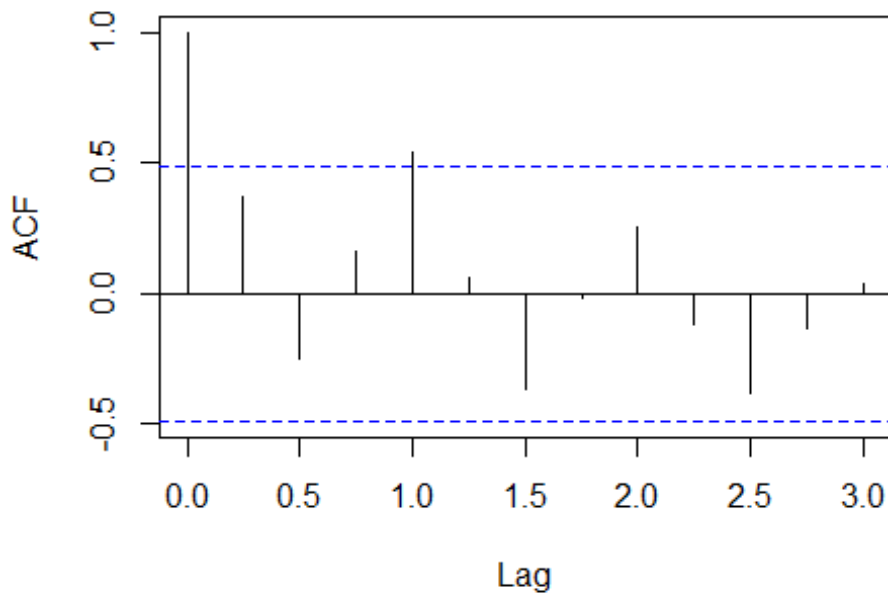
Interpretación

- Dado que $p\text{-value} = 0.1004 > 0.05$, no podemos rechazar la hipótesis nula. Esto significa que la serie no es estacionaria.

Analiza su gráfico de autocorrelación

```
# Generar el gráfico de autocorrelación
acf(ventas_ts, main = "Gráfico de Autocorrelación (ACF)")
```

Gráfico de Autocorrelación (ACF)



Interpretación

- El gráfico de Autocorrelación (ACF) muestra la correlación de la serie temporal con sus propios retardos (lag), lo que ayuda a identificar patrones de dependencia temporal, como tendencia o estacionalidad.

Correlación significativa en el lag 0:

- En el lag 0, la autocorrelación es igual a 1 (correlación perfecta).

Disminución lenta de la autocorrelación:

- Los valores de ACF disminuyen lentamente con los lags iniciales, indicando la presencia de tendencia en la serie. Esto es característico de series no estacionarias.

Valores significativos:

- Las líneas verticales que superan los límites azules (calculados con base en un intervalo de confianza del 95%) indican lags donde la autocorrelación es estadísticamente significativa:
- Hay correlación significativa en los primeros lags, particularmente en el lag 1.

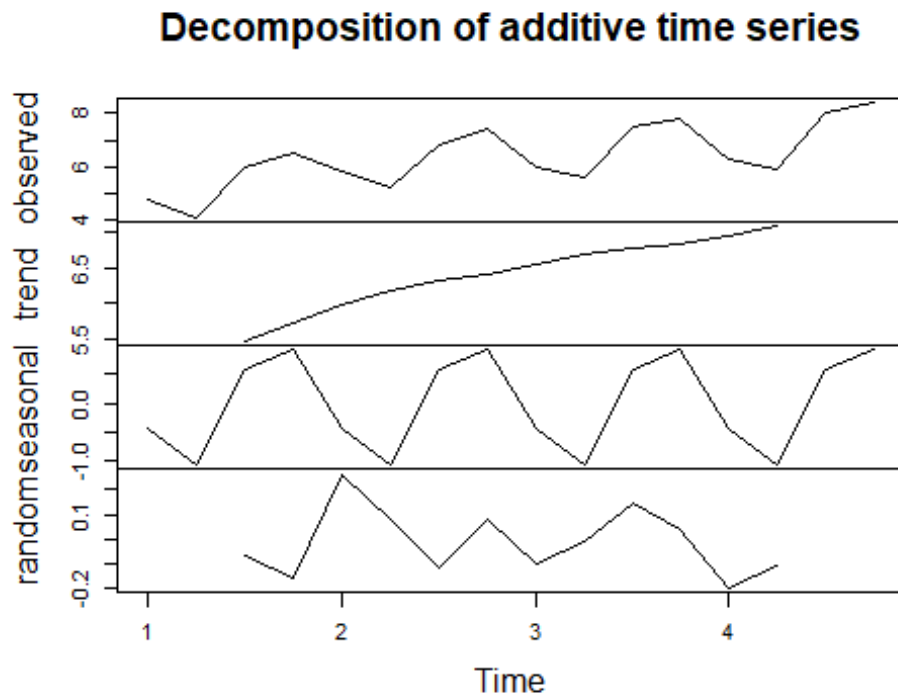
Estacionalidad:

No se observa un patrón claro de repeticiones periódicas en las barras de ACF (picos espaciados de manera regular). Esto indica que la estacionalidad no es prominente en esta serie.

Identifica si el modelo puede ser sumativo o multiplicativo (puedes probar con ambos para ver con cuál es mejor el modelo)

```
# Descomposición aditiva
```

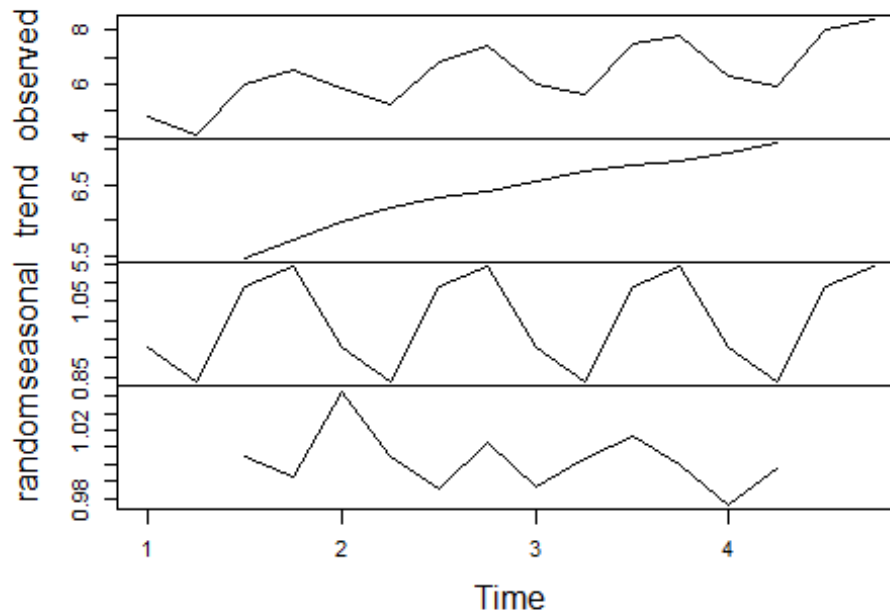
```
descom_aditiva <- decompose(ventas_ts, type = "additive")  
plot(descom_aditiva)
```



```
# Descomposición multiplicativa
```

```
descom_multiplicativa <- decompose(ventas_ts, type = "multiplicative")  
plot(descom_multiplicativa)
```

Decomposition of multiplicative time series



Interpretación de Gráficos de Descomposición de la Serie Temporal

Gráfico de Descomposición Aditiva

En la descomposición aditiva, la serie temporal se descompone como:

Serie Observada = Tendencia + Estacionalidad + Componente Aleatorio

1. **Componente Observado (observed):**
 - Se observa una tendencia creciente y un patrón estacional repetitivo.
2. **Tendencia (trend):**
 - Aumento gradual en las ventas a lo largo del tiempo.
 - Tendencia positiva en la serie temporal.
3. **Estacionalidad (seasonal):**
 - Muestra un patrón estacional repetitivo en cada ciclo anual.
 - La amplitud de las fluctuaciones estacionales es constante.
4. **Componente Aleatorio (random):**
 - Incluye las variaciones residuales no explicadas por la tendencia ni la estacionalidad.
 - Las fluctuaciones son pequeñas, lo que sugiere que la mayor parte de la variabilidad se explica por la tendencia y la estacionalidad.

Gráfico de Descomposición Multiplicativa

En la descomposición multiplicativa, la serie temporal se descompone como:

Serie Observada = Tendencia × Estacionalidad × Componente Aleatorio

1. **Componente Observado (observed):**
 - Muestra la serie original con tendencia creciente y patrón estacional.
2. **Tendencia (trend):**
 - Muestra un crecimiento gradual en las ventas a lo largo del tiempo.
3. **Estacionalidad (seasonal):**
 - Las fluctuaciones son proporcionales a los valores de la serie.
4. **Componente Aleatorio (random):**
 - Contiene las variaciones residuales. Los valores parecen proporcionalmente pequeños, lo que indica que el modelo multiplicativo captura bien la estructura de la serie.

```
# Comparar residuos para evaluar el modelo
```

```
# Aditivo
```

```
residuos_aditivo <- descom_aditiva$random  
cat("Residuos Aditivos:\n")
```

```
## Residuos Aditivos:
```

```
print(summary(residuos_aditivo))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's  
## -0.19792 -0.10833 -0.03542 -0.01042  0.08125  0.26458      4
```

```
# Multiplicativo
```

```
residuos_multiplicativo <- descom_multiplicativa$random  
cat("Residuos Multiplicativos:\n")
```

```
## Residuos Multiplicativos:
```

```
print(summary(residuos_multiplicativo))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's  
##  0.9758  0.9909  1.0013  1.0017  1.0069  1.0430      4
```

Interpretación

- Los residuos representan las variaciones de la serie que no fueron explicadas por los componentes de tendencia y estacionalidad.

Modelo Aditivo:

- Los residuos son pequeños y están bien distribuidos alrededor de 0.
- Indica que este modelo explica bien las variaciones de la serie temporal.

Modelo Multiplicativo:

- Los residuos están cercanos a 1, con fluctuaciones proporcionales.


```

# Calcular varianza de los residuos
var_aditivo <- var(residuos_aditivo, na.rm = TRUE)
var_multiplicativo <- var(residuos_multiplicativo, na.rm = TRUE)

cat("\nVarianza de los residuos - Aditivo:", var_aditivo, "\n")

##
## Varianza de los residuos - Aditivo: 0.01886364

cat("Varianza de los residuos - Multiplicativo:", var_multiplicativo,
"\n")

## Varianza de los residuos - Multiplicativo: 0.000305848

# Elegir mejor modelo
if (var_aditivo < var_multiplicativo) {
  cat("El modelo aditivo es mejor.\n")
} else {
  cat("El modelo multiplicativo es mejor.\n")
}

## El modelo multiplicativo es mejor.

```

Calcula los índices estacionales y grafica la serie desestacionalizada

```

# Extraer índices estacionales (aditivo)
indices_estacionales_aditiva <- descom_aditiva$seasonal
cat("Índices estacionales (aditiva):\n")

## Índices estacionales (aditiva):

print(indices_estacionales_aditiva)

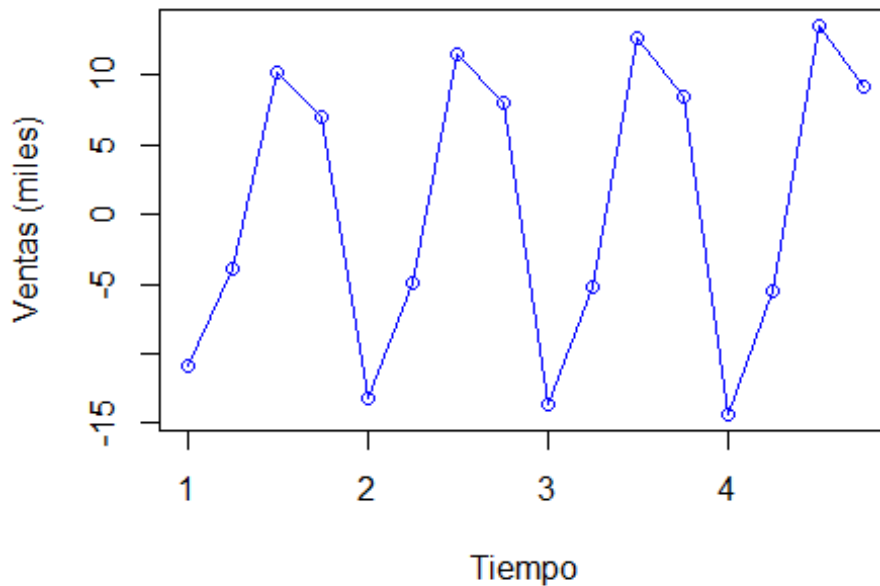
##           Qtr1           Qtr2           Qtr3           Qtr4
## 1 -0.4395833 -1.0687500  0.5895833  0.9187500
## 2 -0.4395833 -1.0687500  0.5895833  0.9187500
## 3 -0.4395833 -1.0687500  0.5895833  0.9187500
## 4 -0.4395833 -1.0687500  0.5895833  0.9187500

# Serie desestacionalizada (aditivo)
serie_desestacionalizada_aditiva <- ventas_ts /
indices_estacionales_aditiva

# Graficar la serie desestacionalizada (aditivo)
plot(serie_desestacionalizada_aditiva, main = "Serie Desestacionalizada
(Aditiva)",
      xlab = "Tiempo", ylab = "Ventas (miles)", col = "blue", type = "o")

```

Serie Desestacionalizada (Aditiva)



Interpretación

- Los índices estacionales indican que Qtr2 tiene un impacto negativo significativo en las ventas, mientras que Qtr4 refleja el mayor incremento estacional.
- Las ventas muestran oscilaciones alrededor de una tendencia creciente o decreciente, dependiendo del momento de la serie temporal.

```
# Extraer índices estacionales (multiplicativo)
indices_estacionales_multiplicativos <- descom_multiplicativa$seasonal
cat("Índices estacionales (multiplicativos):\n")

## Índices estacionales (multiplicativos):

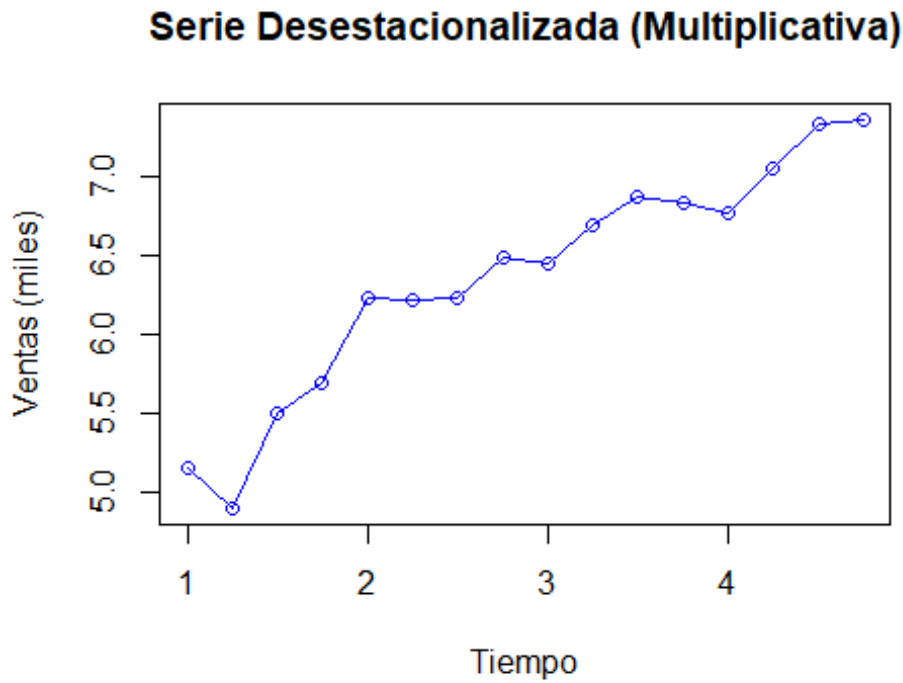
print(indices_estacionales_multiplicativos)

##          Qtr1          Qtr2          Qtr3          Qtr4
## 1 0.9306617 0.8363763 1.0915441 1.1414179
## 2 0.9306617 0.8363763 1.0915441 1.1414179
## 3 0.9306617 0.8363763 1.0915441 1.1414179
## 4 0.9306617 0.8363763 1.0915441 1.1414179

# Serie desestacionalizada (multiplicativa)
serie_desestacionalizada_multiplicativa <- ventas_ts /
indices_estacionales_multiplicativos

# Graficar la serie desestacionalizada (multiplicativa)
```

```
plot(serie_desestacionalizada_multiplicativa, main = "Serie
Desestacionalizada (Multiplicativa)",
     xlab = "Tiempo", ylab = "Ventas (miles)", col = "blue", type = "o")
```



Interpretación

- El Trimestre 2 (Qtr2) tiene el menor impacto (reducción), mientras que el Trimestre 4 (Qtr4) tiene el mayor impacto positivo.
- La serie desestacionalizada muestra una tendencia subyacente clara, lo que facilita el análisis de tendencias a largo plazo y eventos aleatorios.

Analiza el modelo lineal de la tendencia

1. Descomposición aditiva

Librerías necesarias

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

Extraer la tendencia de la serie descompuesta (descomposición aditiva)

```
descom_aditiva <- decompose(ventas_ts, type = "additive")
```

Ventas desestacionalizadas (observado menos estacionalidad)

```
ventas_desest <- ventas_ts - descom_aditiva$seasonal
```

```

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

# Crear el dataframe para la regresión lineal
df_tendencia <- data.frame(
  Tiempo = 1:length(ventas_ts),
  Ventas_Desest = as.numeric(ventas_desest)
)

# Ajustar el modelo de regresión lineal de la tendencia
modelo_lineal <- lm(Ventas_Desest ~ Tiempo, data = df_tendencia)

# Analiza la significancia del modelo lineal, global e individual
# Resumen del modelo
summary(modelo_lineal)

##
## Call:
## lm(formula = Ventas_Desest ~ Tiempo, data = df_tendencia)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2992 -0.1486 -0.0037  0.1005  0.3698
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.13917    0.10172   50.52 < 2e-16 ***
## Tiempo        0.14613    0.01052   13.89 1.4e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.194 on 14 degrees of freedom
## Multiple R-squared:  0.9324, Adjusted R-squared:  0.9275
## F-statistic: 193 on 1 and 14 DF, p-value: 1.399e-09

```

Ecuación del Modelo

$$\text{Ventas Desestacionalizadas} = 5.1391 + 0.1461 \times \text{Tiempo}$$

Interpretación de la Ecuación

1. **Intercepto (beta_0 = 5.1391):**
 - Representa el nivel estimado de las ventas desestacionalizadas en el tiempo inicial (t = 0).
2. **Pendiente (beta_1 = 0.1461):**
 - Indica que las ventas desestacionalizadas aumentan en promedio **0.1461 unidades** por cada incremento de una unidad de tiempo.

Bondad del Ajuste

1. **Coefficiente de Determinación ($R^2 = 0.9324$):**

- El 93.24% de la variabilidad en las ventas desestacionalizadas es explicada por el tiempo.
- Este valor indica que el modelo lineal tiene un ajuste muy bueno, ya que explica la mayor parte de la variación en las ventas.

2. **Estadística (F):**

- ($F = 193$) con un p-value extremadamente bajo.
- Esto indica que el modelo global es altamente significativo, confirmando que el tiempo es una variable relevante para explicar las ventas desestacionalizadas.

Significancia Individual (Coeficientes)

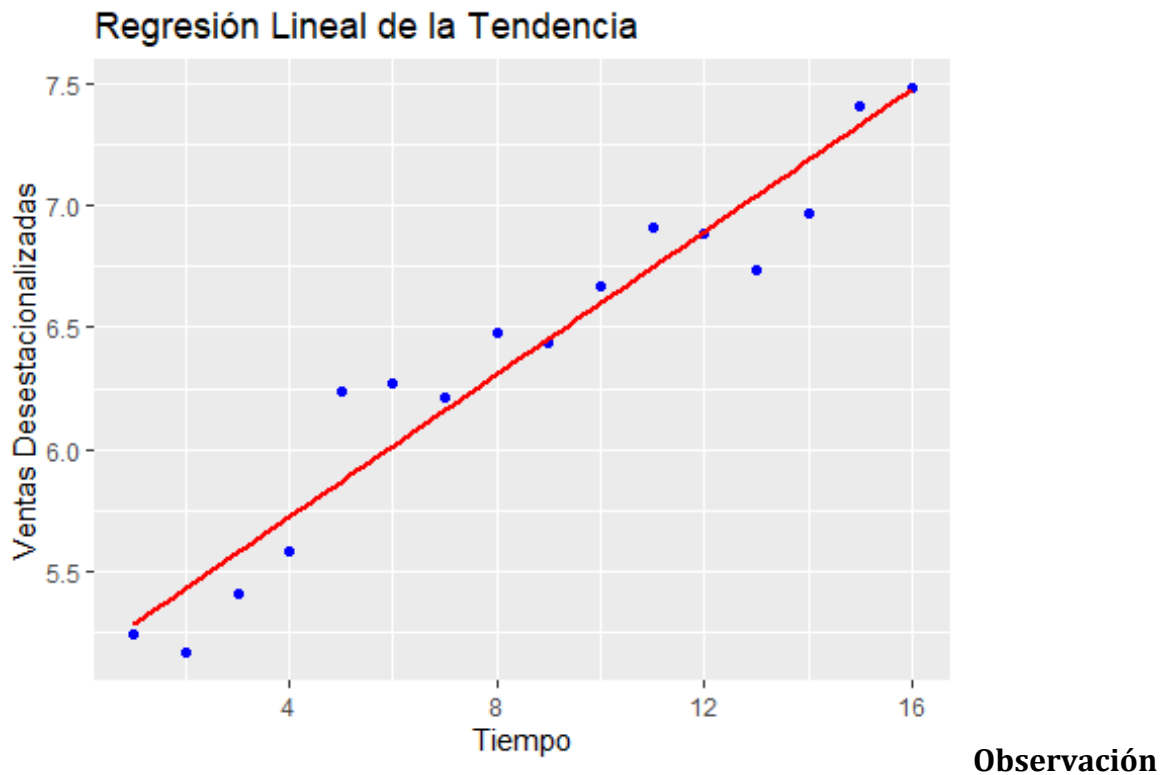
- **Valor estimado:** (5.1391).
- **(p)-value:** ($< 2 \times 10^{-16}$), extremadamente significativo.
- Esto confirma que el intercepto contribuye de manera significativa al modelo.

Tiempo (beta_1): - **Valor estimado:** (0.1461). - **(p)-value:** (1.4×10^{-9}), altamente significativo. - Indica que existe una relación lineal fuerte y significativa entre el tiempo y las ventas desestacionalizadas.

Significancia global: - El modelo lineal es estadísticamente significativo para explicar las ventas desestacionalizadas con base en el tiempo, con un R^2 alto y un p-value muy bajo. - **Impacto del tiempo:** El tiempo es una variable relevante y significativa para predecir las ventas desestacionalizadas, mostrando un crecimiento constante promedio en las ventas de **0.1461 unidades por unidad de tiempo**.

```
# Análisis gráfico de la tendencia ajustada
ggplot(df_tendencia, aes(x = Tiempo, y = Ventas_Desest)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("Regresión Lineal de la Tendencia") +
  xlab("Tiempo") +
  ylab("Ventas Desestacionalizadas")

## `geom_smooth()` using formula = 'y ~ x'
```



- La gráfica muestra una buena correspondencia entre los datos (puntos azules) y la línea de tendencia ajustada (línea roja). El modelo lineal captura claramente el aumento en las ventas desestacionalizadas a lo largo del tiempo.

```
# Haz el análisis de residuos
```

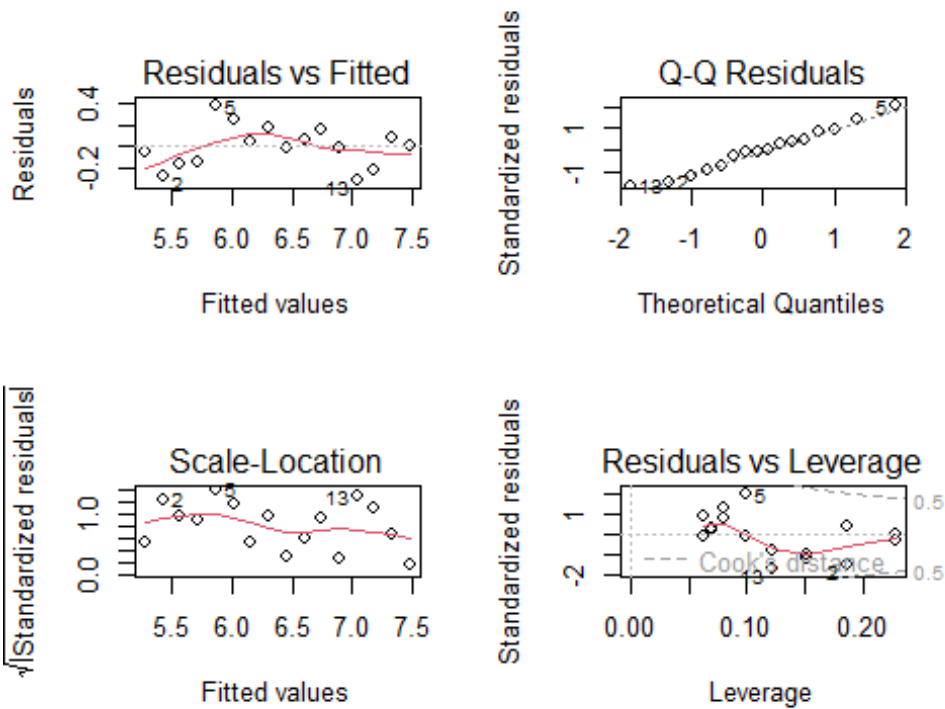
```
# Análisis de residuos
```

```
residuos <- residuals(modelo_lineal)
```

```
# Gráfica de residuos
```

```
par(mfrow = c(2, 2))
```

```
plot(modelo_lineal)
```



```
# Análisis estadístico de los residuos
```

```
shapiro.test(residuos) # Normalidad
```

```
##
```

```
## Shapiro-Wilk normality test
```

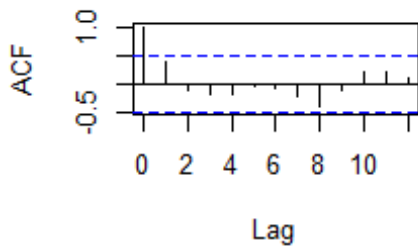
```
##
```

```
## data: residuos
```

```
## W = 0.97816, p-value = 0.9473
```

```
acf(residuos, main = "Autocorrelación de Residuos") # Autocorrelación
```

Autocorrelación de Residuos



Resumen del

Análisis de Residuos

Gráficos de Diagnóstico

- 1. Residuals vs Fitted:**
 - Los residuos están distribuidos alrededor de 0 sin patrones claros.
 - Hay una ligera curva que indica una posible no linealidad leve, aunque no parece crítica.
- 2. Q-Q Plot:**
 - Los puntos están alineados con la línea diagonal, lo que sugiere que los residuos siguen una distribución normal.
- 3. Scale-Location:**
 - La varianza de los residuos parece constante, no se detecta heteroscedasticidad.
- 4. Residuals vs Leverage:**
 - No hay observaciones influyentes según la distancia de Cook, indicando que ningún punto tiene un impacto excesivo en el modelo.

Prueba de Shapiro-Wilk - Resultado: ($W = 0.97816$), ($p\text{-value} = 0.9473$). - No se rechaza la hipótesis nula; los residuos siguen una distribución normal.

Autocorrelación de Residuos (ACF)

- Las barras del gráfico de ACF están dentro de los límites para la mayoría de los retardos, indicando ausencia de autocorrelación significativa.

- Un pequeño valor significativo en el primer lag no parece crítico.

Conclusión 1. Supuestos del Modelo Lineal: - Los residuos cumplen con los supuestos de normalidad ($p\text{-value} = 0.9473$), independencia y homoscedasticidad.

2. Validez del Modelo:

- El modelo ajustado es adecuado para explicar la relación lineal entre el tiempo y las ventas desestacionalizadas.

2. Descomposición multiplicativa

```
# Extraer la tendencia de la serie descompuesta (usaremos la
descomposición multiplicativa)
descom_multiplicativa <- decompose(ventas_ts, type = "multiplicative")

# Ventas desestacionalizadas (observado dividido por estacionalidad)
ventas_desest_1 <- ventas_ts / descom_multiplicativa$seasonal

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

# Crear el dataframe para la regresión lineal
df_tendencia_1 <- data.frame(
  Tiempo = 1:length(ventas_ts),
  Ventas_Desest = as.numeric(ventas_desest_1)
)

# Ajustar el modelo de regresión lineal de la tendencia
modelo_lineal_1 <- lm(Ventas_Desest ~ Tiempo, data = df_tendencia_1)

# Resumen del modelo
summary(modelo_lineal_1)

##
## Call:
## lm(formula = Ventas_Desest ~ Tiempo, data = df_tendencia_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5007 -0.1001  0.0037  0.1207  0.3872
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.10804    0.11171   45.73  < 2e-16 ***
## Tiempo       0.14738    0.01155   12.76 4.25e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.213 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
```

Ecuación del modelo

$$\text{Ventas Desestacionalizadas} = 5.10804 + 0.1473 \times \text{Tiempo}$$

Interpretación de la Ecuación

1. Intercepto (beta_0):

- Valor estimado: (5.10804), lo que indica que el nivel estimado de ventas desestacionalizadas en el tiempo inicial (t = 0) es aproximadamente 5.10.

2. Pendiente (beta_1):

- Valor estimado: (0.1473), lo que significa que las ventas desestacionalizadas aumentan en promedio **0.1473 unidades por unidad de tiempo**.

Significancia del modelo global

La **significancia global del modelo** se evalúa utilizando la estadística (F): - (**F - Statistic**): **162.7** con un \$ p-value: 4.248×10^{-9} \$. - El p-value es menor a 0.05, lo que indica que el modelo es **globalmente significativo**. - Esto significa que existe evidencia suficiente para concluir que el tiempo tiene una relación lineal significativa con las ventas desestacionalizadas.

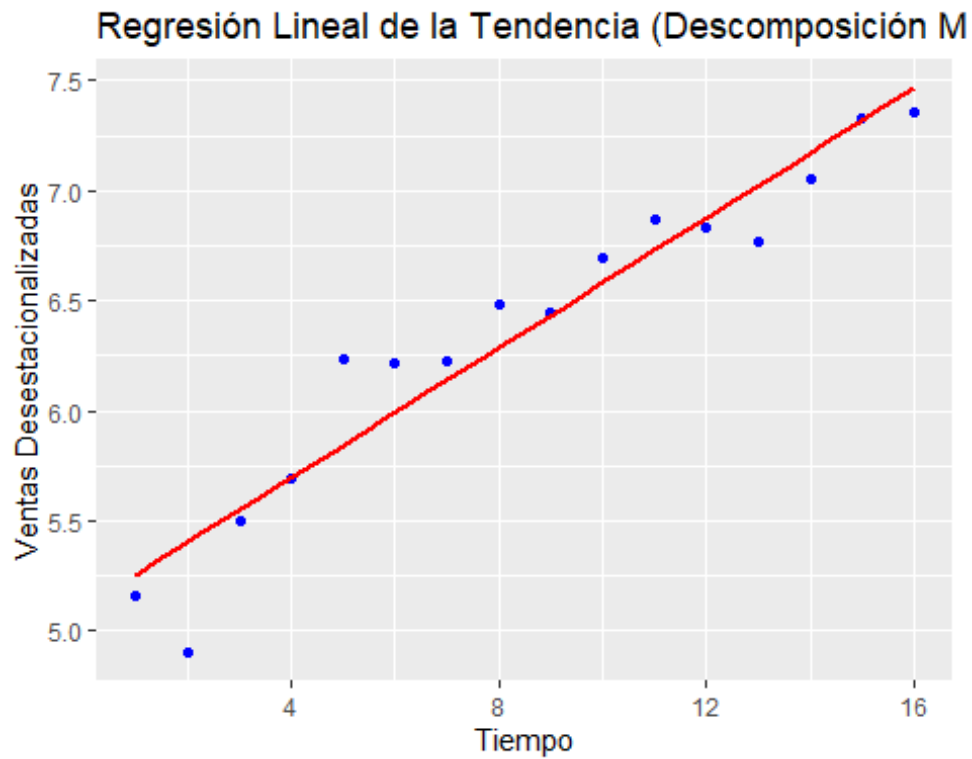
Análisis de significancia individual Intercepto (beta_0): - **t-value: 45.73**, *p* – *value*: ($< 2 \times 10^{-16}$). - El intercepto es estadísticamente significativo (*p* < 0.05), lo que significa que su contribución al modelo es relevante.

Pendiente (beta_1): - **t-value: 12.76**, \$ p-value(4.25×10^{-9})\$ - La pendiente es altamente significativa (\$ p < 0.05\$), lo que indica que el tiempo tiene un impacto positivo y significativo sobre las ventas desestacionalizadas.

Bondad del Ajuste - Con un R^2 ajustado del 91.51 y errores residuales pequeños, el modelo lineal ajusta bien los datos y explica adecuadamente la tendencia de las ventas desestacionalizadas.

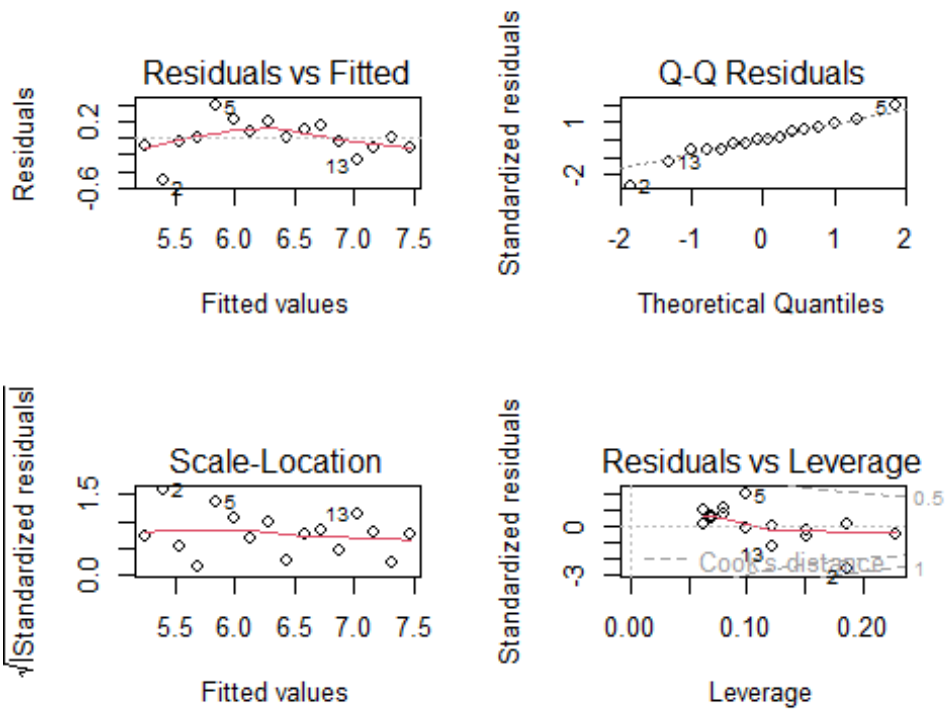
```
# Análisis gráfico de la tendencia ajustada
library(ggplot2)
ggplot(df_tendencia_1, aes(x = Tiempo, y = Ventas_Desest)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("Regresión Lineal de la Tendencia (Descomposición
Multiplicativa)") +
  xlab("Tiempo") +
  ylab("Ventas Desestacionalizadas")

## `geom_smooth()` using formula = 'y ~ x'
```



Observaciones - La línea ajustada captura adecuadamente el aumento de las ventas desestacionalizadas a lo largo del tiempo. - Los datos parecen ajustarse bien al modelo.

```
# Análisis de residuos  
residuos_1 <- residuals(modelo_lineal_1)  
  
# Gráfica de residuos  
par(mfrow = c(2, 2))  
plot(modelo_lineal_1)
```

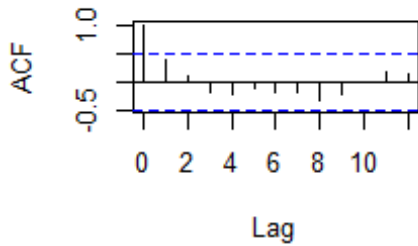


```
# Análisis estadístico de los residuos
shapiro.test(residuos_1) # Normalidad

##
##  Shapiro-Wilk normality test
##
## data:  residuos_1
## W = 0.96379, p-value = 0.7307

acf(residuos_1, main = "Autocorrelación de Residuos") # Autocorrelación
```

Autocorrelación de Residuos



Resumen del

Análisis de Residuos

Gráficos de diagnóstico del modelo

- 1. Residuals vs Fitted:**
 - Los residuos están distribuidos alrededor de 0, pero muestran una leve curvatura, lo que podría indicar una ligera no linealidad o que el modelo no captura completamente la tendencia.
- 2. Q-Q Plot**
 - Los puntos siguen mayormente la línea diagonal, lo que sugiere que los residuos se distribuyen aproximadamente de manera normal.
- 3. Scale-Location**
 - Los residuos estandarizados están dispersos de manera uniforme a lo largo de los valores ajustados.
 - Esto indica que no hay evidencia clara de heteroscedasticidad.
- 4. Residuals vs Leverage**
 - No se observan puntos con valores de leverage o distancia de Cook lo suficientemente altos como para considerarse influyentes.
 - Esto indica que ninguna observación tiene un impacto significativo en el modelo.

Prueba de normalidad de Shapiro-Wilk - Resultados - $W = 0.97816$, $p\text{-value} = 0.9473$.

- Dado que el $p > 0.05$, no se rechaza la hipótesis nula de que los residuos siguen una distribución normal. - Esto confirma que el supuesto de normalidad de los residuos es válido.

Autocorrelación de residuos (ACF) - El gráfico ACF muestra que todas las barras (excepto el lag 0) están dentro de los límites azules, lo que indica que no hay autocorrelación significativa entre los residuos. - Esto respalda el supuesto de independencia de los residuos.

Conclusión 1. Normalidad de residuos: - La prueba de Shapiro-Wilk y el Q-Q Plot confirman que los residuos siguen una distribución normal.

2. Independencia de los residuos:

- El análisis ACF muestra que no hay autocorrelación significativa en los residuos.

3. Homoscedasticidad:

- El gráfico Scale-Location indica que la varianza de los residuos es constante a lo largo de los valores ajustados.

4. Influencia de puntos:

- No se detectaron observaciones influyentes significativas en el gráfico Residuals vs Leverage.

Calcula el CME y el EPAM de la predicción de la serie de tiempo

Aditiva

Predicciones del modelo basado en la tendencia ajustada

```
predicciones <- predict(modelo_lineal, newdata = df_tendencia)
```

Valores originales desestacionalizados

```
valores_reales <- df_tendencia$Ventas_Desest
```

Calcular el CME (Error Cuadrático Medio)

```
cme <- mean((valores_reales - predicciones)^2)
print(paste("CME (Error Cuadrático Medio):", cme))
```

```
## [1] "CME (Error Cuadrático Medio): 0.0329191687091504"
```

Calcular el EPAM (Error Porcentual Absoluto Medio)

```
epam <- mean(abs((valores_reales - predicciones) / valores_reales)) * 100
print(paste("EPAM (Error Porcentual Absoluto Medio) (%) :", epam))
```

```
## [1] "EPAM (Error Porcentual Absoluto Medio) (%): 2.3413189805655"
```

Multiplicativa

Predicciones del modelo basado en la tendencia ajustada

```
predicciones_1 <- predict(modelo_lineal_1, newdata = df_tendencia_1)
```

Valores originales desestacionalizados

```
valores_reales_1 <- df_tendencia_1$Ventas_Desest
```

```

# Calcular el CME (Error Cuadrático Medio)
cme_1 <- mean((valores_reales_1 - predicciones_1)^2)
print(paste("CME (Error Cuadrático Medio):", cme_1))

## [1] "CME (Error Cuadrático Medio): 0.0397064045453044"

## [1] "CME (Error Cuadrático Medio): 0.0397064045453044"
# Calcular el EPAM (Error Porcentual Absoluto Medio)
epam_1 <- mean(abs((valores_reales_1 - predicciones_1) /
valores_reales_1)) * 100
print(paste("EPAM (Error Porcentual Absoluto Medio) (%):", epam_1))

## [1] "EPAM (Error Porcentual Absoluto Medio) (%): 2.4395328407222"

```

Comparación entre los modelos

Modelo aditivo:

Tiene un menor CME y un menor EPAM, lo que sugiere que este modelo es más adecuado para capturar la tendencia en las ventas desestacionalizadas.

Modelo multiplicativo:

Aunque el error es ligeramente mayor, este modelo sigue siendo bastante preciso (con un EPAM de solo 2.43%), lo que lo hace una alternativa razonable si se considera que las fluctuaciones estacionales podrían depender de los niveles de las ventas.

Explora un mejor modelo, por ejemplo un modelo cuadrático:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

* Para ello transforma la variable ventas (recuerda que la regresión no lineal es una regresión lineal con una transformación).

```

# Aditivo

# Transformar el dataframe para incluir el término cuadrático
df_tendencia$Tiempo2 <- df_tendencia$Tiempo^2

# Ajustar el modelo cuadrático
modelo_cuadratico_aditivo <- lm(Ventas_Desest ~ Tiempo + Tiempo2, data =
df_tendencia)

# Resumen del modelo cuadrático aditivo
summary(modelo_cuadratico_aditivo)

##
## Call:
## lm(formula = Ventas_Desest ~ Tiempo + Tiempo2, data = df_tendencia)
##
## Residuals:

```

```
##      Min      1Q   Median      3Q      Max
## -0.30333 -0.13440 -0.01928  0.11368  0.33301
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.930833   0.155679  31.673 1.08e-13 ***
## Tiempo      0.215572   0.042149   5.115 0.000199 ***
## Tiempo2     -0.004085   0.002410  -1.695 0.113918
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1822 on 13 degrees of freedom
## Multiple R-squared:  0.9446, Adjusted R-squared:  0.9361
## F-statistic: 110.8 on 2 and 13 DF,  p-value: 6.805e-09

# Predicciones con el modelo cuadrático
predicciones_aditivo_cuadratico <- predict(modelo_cuadratico_aditivo,
newdata = df_tendencia)

# Calcular CME y EPAM para el modelo cuadrático aditivo
cme_cuadratico_aditivo <- mean((df_tendencia$Ventas_Desest -
predicciones_aditivo_cuadratico)^2)
epam_cuadratico_aditivo <- mean(abs((df_tendencia$Ventas_Desest -
predicciones_aditivo_cuadratico) / df_tendencia$Ventas_Desest)) * 100

# Mostrar resultados
print(paste("CME Aditivo Cuadrático:", cme_cuadratico_aditivo))

## [1] "CME Aditivo Cuadrático: 0.0269619247004358"

print(paste("EPAM Aditivo Cuadrático (%):", epam_cuadratico_aditivo))

## [1] "EPAM Aditivo Cuadrático (%): 2.22986795193133"
```

Ecuación del Modelo

$$\text{Ventas Desestacionalizadas} = 4.9308 + 0.2156 \cdot \text{Tiempo} - 0.0041 \cdot \text{Tiempo}^2$$

Significancia Global - ($F = 110.8$), ($p\text{-value} = 6.805 \times 10^{-9}$): - El modelo es globalmente significativo ($p < 0.05$).

Significancia Individual 1. **Intercepto (beta_0)**: - ($p < 0.001$), altamente significativo. 2. **Término Lineal (beta_1)**: - ($p < 0.001$), altamente significativo. 3. **Término Cuadrático (beta_2)**: - ($p = 0.1139$), no significativo ($p > 0.05$). - El término cuadrático no mejora significativamente el modelo.

Calidad del Ajuste - $R^2 = 0.9446$, $\$R^2 = 0.9361$: - El modelo explica el 93.61% de la variabilidad en las ventas desestacionalizadas. - **Error estándar residual**: 0.1822 (valor bajo).

Errores (CME y EPAM) - CME: 0.0269 (menor que el modelo lineal aditivo). - **EPAM:** 2.23% (ligeramente menor que el modelo lineal aditivo).

Conclusión: - Aunque el modelo cuadrático mejora ligeramente los errores con respecto al modelo lineal aditivo, el término cuadrático no es significativo, lo que sugiere que el modelo lineal podría ser suficiente.

```
# Multiplicativo

# Transformar el dataframe para incluir el término cuadrático
df_tendencia_1$Tiempo2 <- df_tendencia_1$Tiempo^2

# Ajustar el modelo cuadrático
modelo_cuadratico_multiplicativo <- lm(Ventas_Desest ~ Tiempo + Tiempo2,
data = df_tendencia_1)

# Resumen del modelo cuadrático multiplicativo
summary(modelo_cuadratico_multiplicativo)

##
## Call:
## lm(formula = Ventas_Desest ~ Tiempo + Tiempo2, data = df_tendencia_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36986 -0.07058 -0.00100  0.11345  0.33110
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.790283   0.152429   31.426 1.20e-13 ***
## Tiempo       0.253302   0.041269    6.138 3.56e-05 ***
## Tiempo2      -0.006231   0.002360   -2.640  0.0204  *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1784 on 13 degrees of freedom
## Multiple R-squared:  0.9484, Adjusted R-squared:  0.9405
## F-statistic: 119.6 on 2 and 13 DF, p-value: 4.268e-09

# Predicciones con el modelo cuadrático
predicciones_multiplicativo_cuadratico <-
predict(modelo_cuadratico_multiplicativo, newdata = df_tendencia_1)

# Calcular CME y EPAM para el modelo cuadrático multiplicativo
cme_cuadratico_multiplicativo <- mean((df_tendencia_1$Ventas_Desest -
predicciones_multiplicativo_cuadratico)^2)
epam_cuadratico_multiplicativo <- mean(abs((df_tendencia_1$Ventas_Desest
- predicciones_multiplicativo_cuadratico) /
df_tendencia_1$Ventas_Desest)) * 100
```

```
# Mostrar resultados
print(paste("CME Multiplicativo Cuadrático:",
cme_cuadratico_multiplicativo))

## [1] "CME Multiplicativo Cuadrático: 0.0258476659392304"
```

Ecuación del Modelo

$$\text{Ventas Desestacionalizadas} = 4.7903 + 0.2533 \cdot \text{Tiempo} - 0.0062 \cdot \text{Tiempo}^2$$

Significancia Global - ($F = 119.6$), ($p\text{-value} = 4.268 \times 10^{-9}$): - El modelo es globalmente significativo ($p < 0.05$).

Significancia Individual 1. **Intercepto (beta_0)**: - ($p < 0.001$), altamente significativo. 2. **Término Lineal (beta_1)**: - ($p < 0.001$), altamente significativo. 3. **Término Cuadrático (beta_2)**: - ($p = 0.0204$), significativo ($p < 0.05$). - El término cuadrático mejora el ajuste del modelo.

Calidad del Ajuste - ($R^2 = 0.9484$),

$$(R^2 \text{ ajustado} = 0.9405)$$

: - El modelo explica el 94.05% de la variabilidad en las ventas desestacionalizadas. -

Error estándar residual: 0.1784 (valor bajo).

Errores (CME y EPAM) - **CME**: 0.0258 (ligeramente menor que el modelo lineal multiplicativo). - **EPAM**: 1.95% (notablemente menor que el modelo lineal multiplicativo).

Conclusión: - El modelo cuadrático mejora significativamente los errores y el ajuste con respecto al modelo lineal multiplicativo. - El término cuadrático es significativo y añade valor al modelo.

Concluye sobre el mejor modelo

```
# Crear la tabla comparativa
tabla_comparacion <- data.frame(
  Modelo = c("Lineal Aditivo", "Lineal Multiplicativo", "Cuadrático
Aditivo", "Cuadrático Multiplicativo"),
  R2_Ajustado = c(
    summary(modelo_lineal)$adj.r.squared,
    summary(modelo_lineal_1)$adj.r.squared,
    summary(modelo_cuadratico_aditivo)$adj.r.squared,
    summary(modelo_cuadratico_multiplicativo)$adj.r.squared
  ),
  CME = c(
    mean(residuals(modelo_lineal)^2),
    mean(residuals(modelo_lineal_1)^2),
    cme_cuadratico_aditivo,
    cme_cuadratico_multiplicativo
  ),
```

```

EPAM_Porcentaje = c(
  mean(abs(residuals(modelo_lineal) / df_tendencia$Ventas_Desest)) *
100,
  mean(abs(residuals(modelo_lineal_1) / df_tendencia_1$Ventas_Desest))
* 100,
  epam_cuadratico_aditivo,
  epam_cuadratico_multiplicativo
),
Termino_Cuadratico_Significativo = c("No", "No",
  ifelse(summary(modelo_cuadratico_aditivo)$coefficients["Tiempo2",
"Pr(>|t|)"] < 0.05, "Sí", "No"),

ifelse(summary(modelo_cuadratico_multiplicativo)$coefficients["Tiempo2",
"Pr(>|t|)"] < 0.05, "Sí", "No")
)
)

# Mostrar la tabla
tabla_comparacion

##           Modelo R2_Ajustado      CME EPAM_Porcentaje
## 1      Lineal Aditivo    0.9275276 0.03291917      2.341319
## 2      Lineal Multiplicativo 0.9151333 0.03970640      2.439533
## 3      Cuadrático Aditivo    0.9360767 0.02696192      2.229868
## 4 Cuadrático Multiplicativo 0.9405047 0.02584767      1.949349
## Termino_Cuadratico_Significativo
## 1                               No
## 2                               No
## 3                               No
## 4                               Sí

```

- El modelo Cuadrático Multiplicativo es el más adecuado porque combina:
- Una excelente capacidad explicativa (R^2 ajustado alto), Errores bajos (CME y EPAM) y una mejora significativa debido al término cuadrático.
- Este modelo es ideal para capturar tanto la tendencia como la relación no lineal en los datos desestacionalizados.

Realiza el pronóstico para el siguiente año y gráfícalo junto con los pronósticos previos y los datos originales.

```

# Crear un nuevo DataFrame para el pronóstico
nuevo_tiempo <- (max(df_tendencia_1$Tiempo) +
1):(max(df_tendencia_1$Tiempo) + 4)
nuevo_tiempo2 <- nuevo_tiempo^2

df_pronostico <- data.frame(
  Tiempo = nuevo_tiempo,

```

```

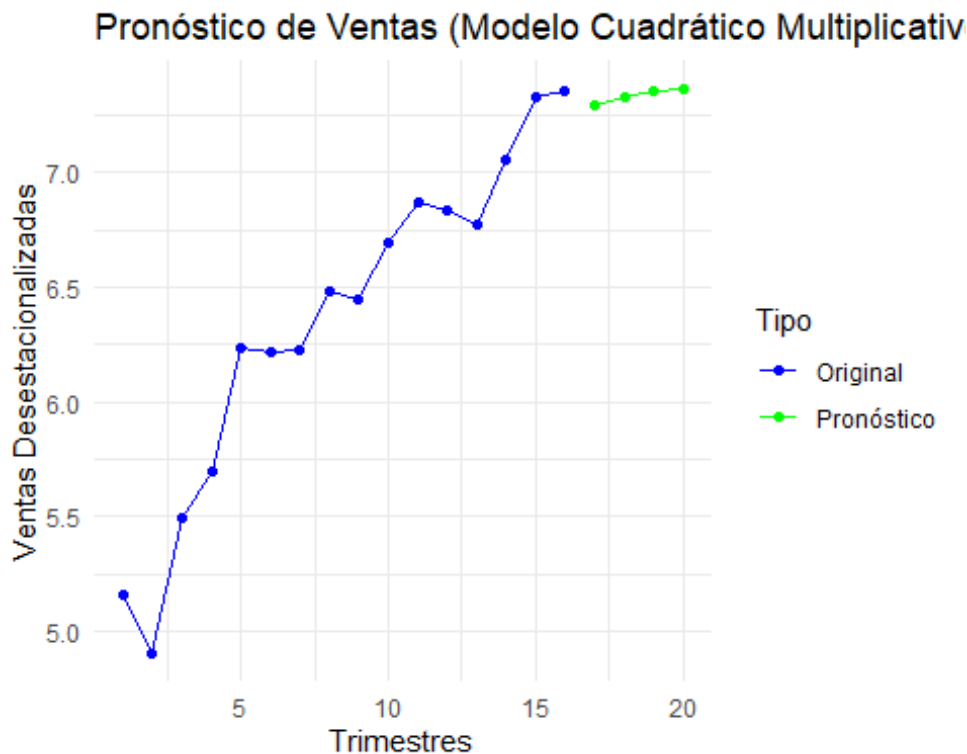
    Tiempo2 = nuevo_tiempo2
  )

  # Realizar el pronóstico
  nuevo_pronostico <- predict(modelo_cuadratico_multiplicativo, newdata =
df_pronostico)

  # Crear un DataFrame combinado con datos originales y pronósticos
  df_todos <- rbind(
    data.frame(Tiempo = df_tendencia_1$Tiempo, Ventas =
df_tendencia_1$Ventas_Desest, Tipo = "Original"),
    data.frame(Tiempo = nuevo_tiempo, Ventas = nuevo_pronostico, Tipo =
"Pronóstico")
  )

  # Graficar los datos originales, el ajuste previo y el pronóstico
  ggplot(df_todos, aes(x = Tiempo, y = Ventas, color = Tipo)) +
    geom_line() +
    geom_point() +
    ggtitle("Pronóstico de Ventas (Modelo Cuadrático Multiplicativo)") +
    xlab("Trimestres") +
    ylab("Ventas Desestacionalizadas") +
    scale_color_manual(values = c("Original" = "blue", "Pronóstico" =
"green")) +
    theme_minimal()

```



```

# Extender las variables de tiempo para el próximo año (4 trimestres
adicionales)
tiempo_extendido <- c(df_tendencia_1$Tiempo,
(length(df_tendencia_1$Tiempo) + 1):(length(df_tendencia_1$Tiempo) + 4))
tiempo_cuadrado_extendido <- tiempo_extendido^2

# Realizar el pronóstico de la tendencia para el siguiente año
pred_tendencia_cuadratica_extendida <- predict(
  modelo_cuadratico_multiplicativo,
  newdata = data.frame(Tiempo = tiempo_extendido, Tiempo2 =
tiempo_cuadrado_extendido)
)

# Repetir Los índices estacionales para el próximo año
indices_estacionales_extendidos <-
rep(descom_multiplicativa$seasonal[1:4], length.out =
length(tiempo_extendido))

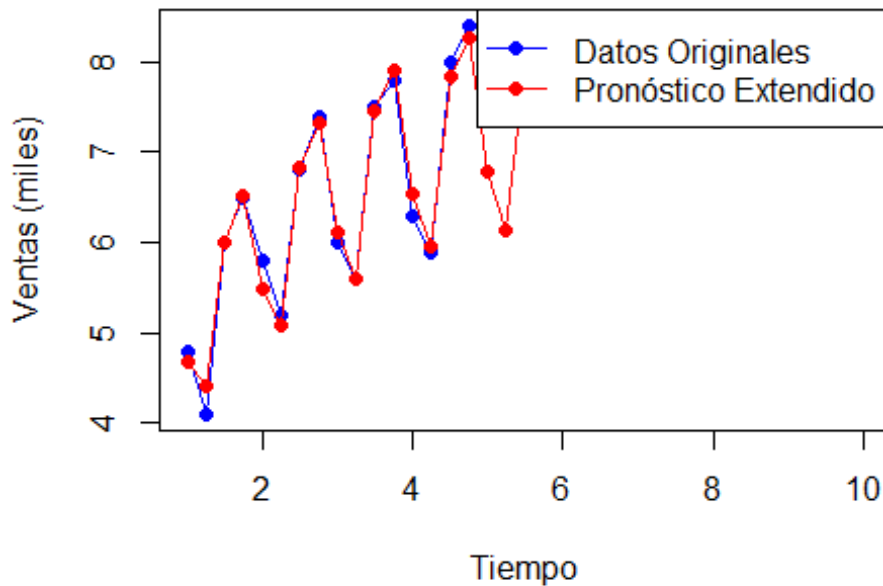
# Aplicar Los índices estacionales para obtener el pronóstico en su forma
original
predicciones_extendidas <- pred_tendencia_cuadratica_extendida *
indices_estacionales_extendidos

# Crear un objeto de serie temporal para el pronóstico extendido
Ventas_pronosticadas <- ts(predicciones_extendidas, start =
start(ventas_ts), frequency = 4)

# Graficar la serie original y el pronóstico del mejor modelo (modelo
cuadrático multiplicativo)
plot(ventas_ts, main = "Pronóstico de Ventas para el Próximo Año", ylab =
"Ventas (miles)", xlab = "Tiempo",
col = "blue", type = "o", pch = 16, xlim = c(1, 10)) # Limitar el
eje x hasta 10
lines(Ventas_pronosticadas, col = "red", type = "o", pch = 16)
legend("topright", legend = c("Datos Originales", "Pronóstico
Extendido"), col = c("blue", "red"), lty=1,pch=16)

```

Pronóstico de Ventas para el Próximo Año



Conclusión * El pronóstico realizado con el modelo cuadrático multiplicativo muestra una tendencia continua y coherente con los datos históricos. * El modelo captura la tendencia creciente en las ventas desestacionalizadas y prevé un crecimiento moderado para los próximos 4 trimestres. * Entre los modelos evaluados, este es el mejor, debido a su R^2 ajustado más alto (0.9405), así como sus menores errores (CME y EPAM).