M5_A8

Sofia Cantu

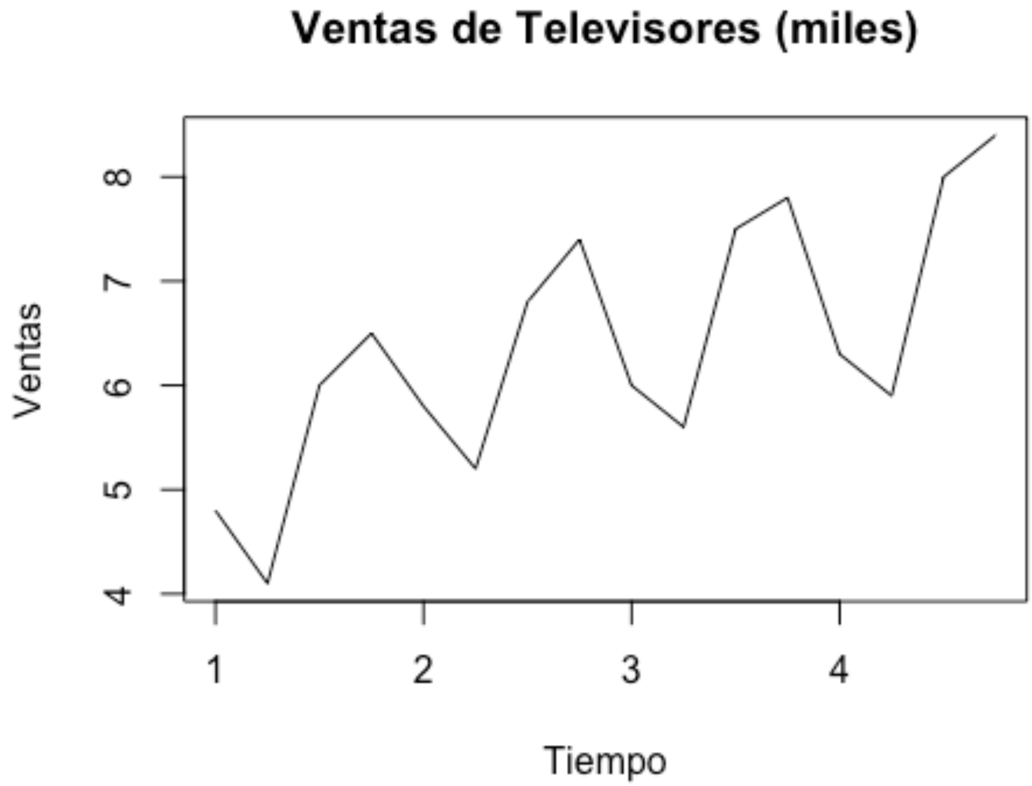
2024-11-12

Series de tiempo

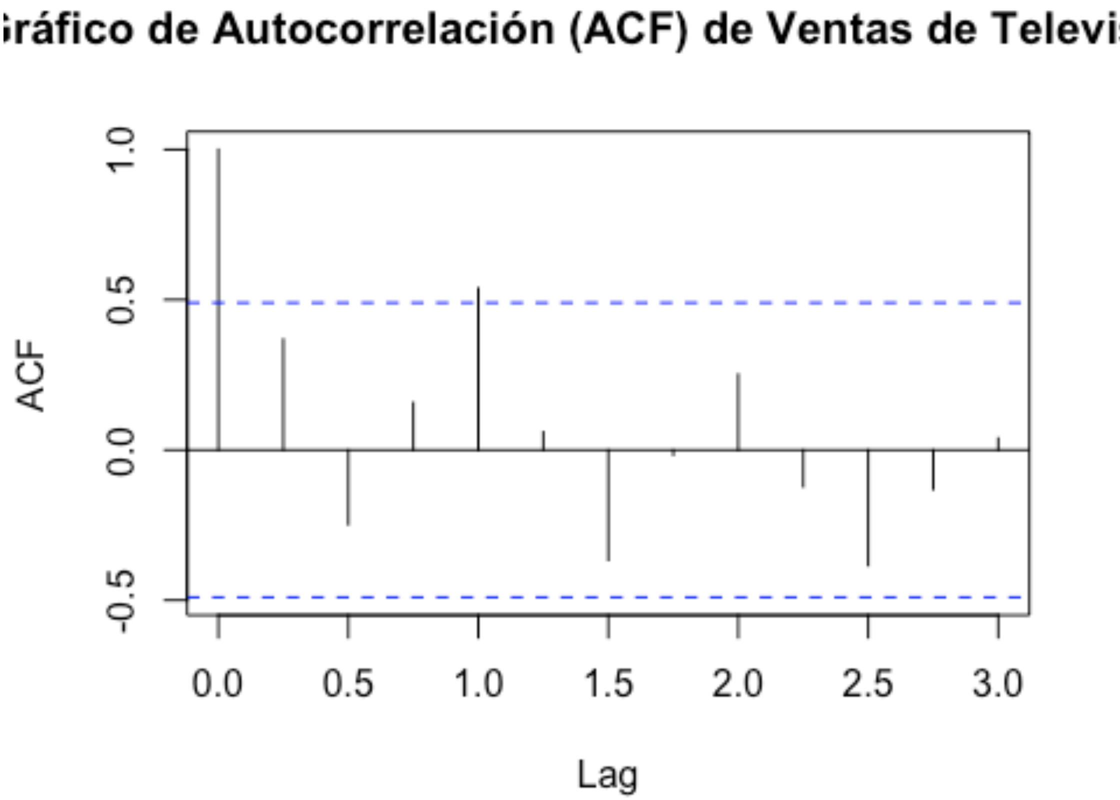
```
# Cargar librerías
if (!require(tseries)) install.packages("tseries")

## Loading required package: tseries

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

library(tseries)
if(!require(forecast)) install.packages("forecast")

## Loading required package: forecast

library(forecast)

# Datos de ventas
ventas <- data.frame(
  Año = rep(1:4, each = 4),
  Trimestre = rep(1:4, times = 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)
)

# Convertir a serie de tiempo trimestral
ventas_ts <- ts(ventas$Ventas, start = c(1, 1), frequency = 4)
```

1. Realiza el análisis de tendencia y estacionalidad:

```
# Prueba de Dickey-Fuller aumentada para verificar estacionariedad
adf.test(ventas_ts)

##
##  Augmented Dickey-Fuller Test
##
## data:  ventas_ts
## Dickey-Fuller = -2.7111, Lag order = 2, p-value = 0.3015
## alternative hypothesis: stationary

# Graficar la serie de tiempo
plot(ventas_ts, main = "Ventas de Televisores (miles)", ylab = "Ventas", xlab
= "Tiempo")
```



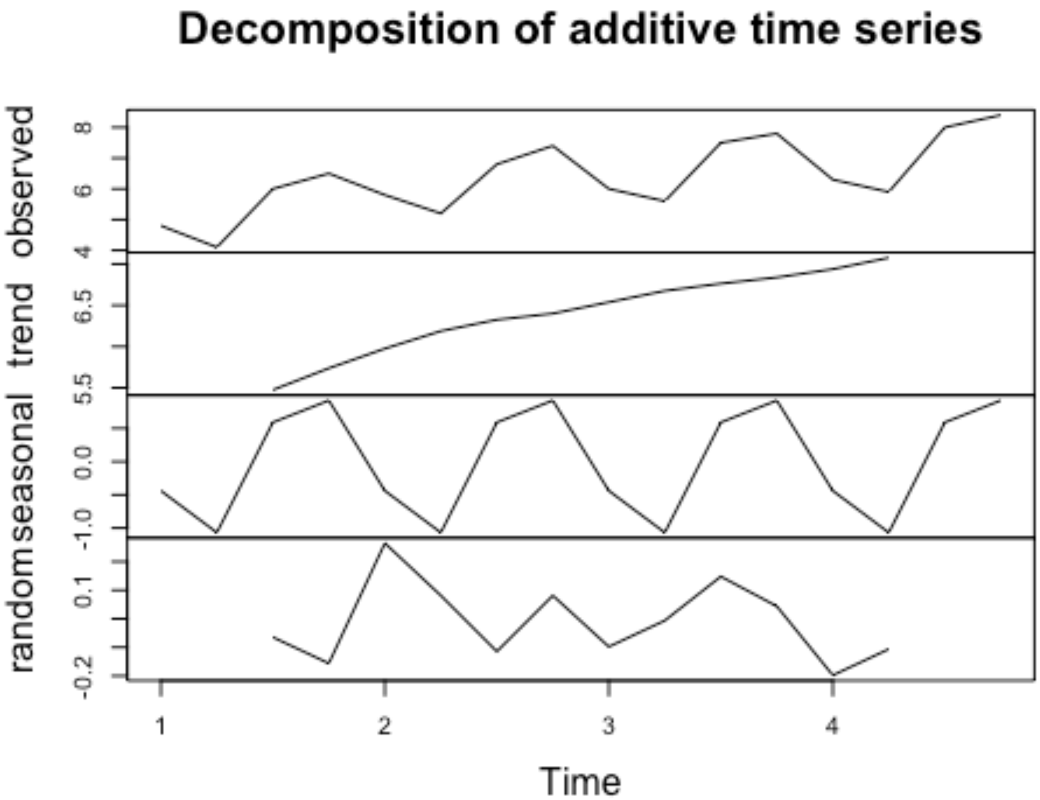

```r
# Graficar el gráfico de autocorrelación (ACF) de la serie original
acf(ventas_ts, main = "Gráfico de Autocorrelación (ACF) de Ventas de Televisores")
```



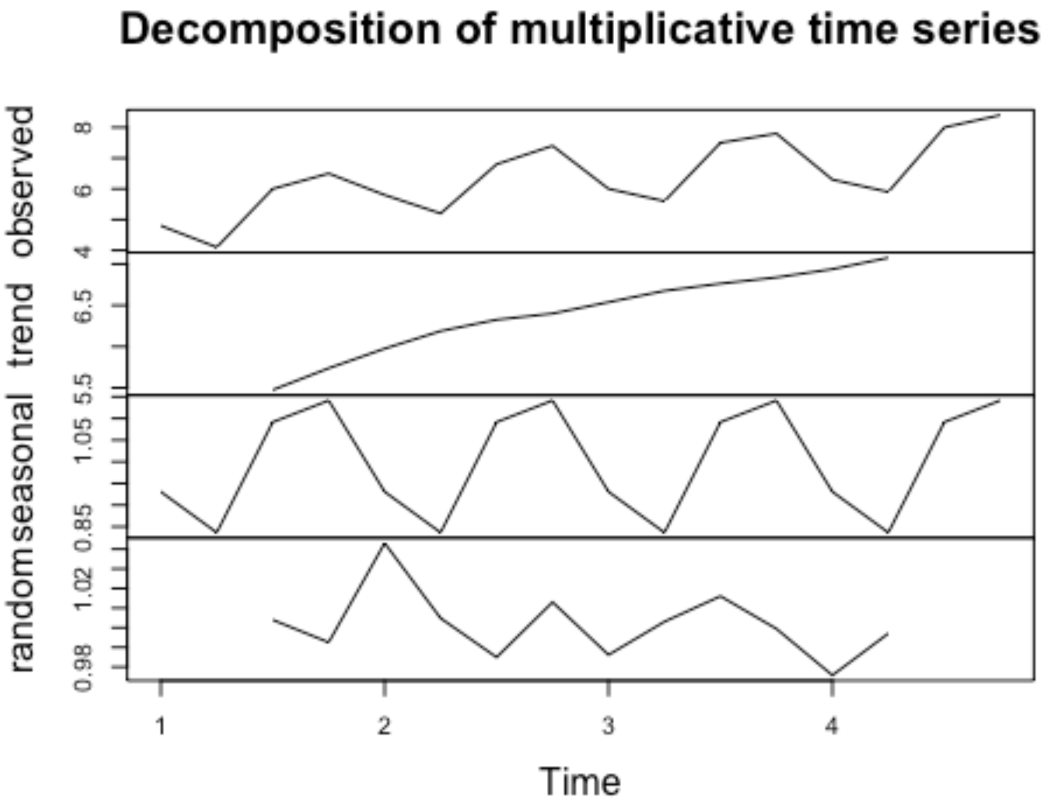
Analiza su gráfico de autocorrelación

El resultado de la prueba de Dickey-Fuller aumentada (ADF) indica que el valor p es de 0.3015, lo cual es superior al nivel de significancia típico (por ejemplo, 0.05). Esto sugiere que no se puede rechazar la hipótesis nula, lo que implica que la serie probablemente no es estacionaria y podría tener una tendencia o componente estacional. El gráfico de autocorrelación (ACF) muestra que las autocorrelaciones en los primeros rezagos son significativas y van disminuyendo gradualmente. Esto sugiere la presencia de una tendencia en la serie, ya que la autocorrelación no cae abruptamente a cero. Este comportamiento es característico de una serie no estacionaria, lo cual concuerda con el resultado de la prueba de Dickey-Fuller, que indicó la no estacionariedad.

```r
# Descomposición de la serie de tiempo
ventas_decomp_add <- decompose(ventas_ts, type = "additive")
ventas_decomp_mult <- decompose(ventas_ts, type = "multiplicative")

# Graficar la descomposición
plot(ventas_decomp_add)
```

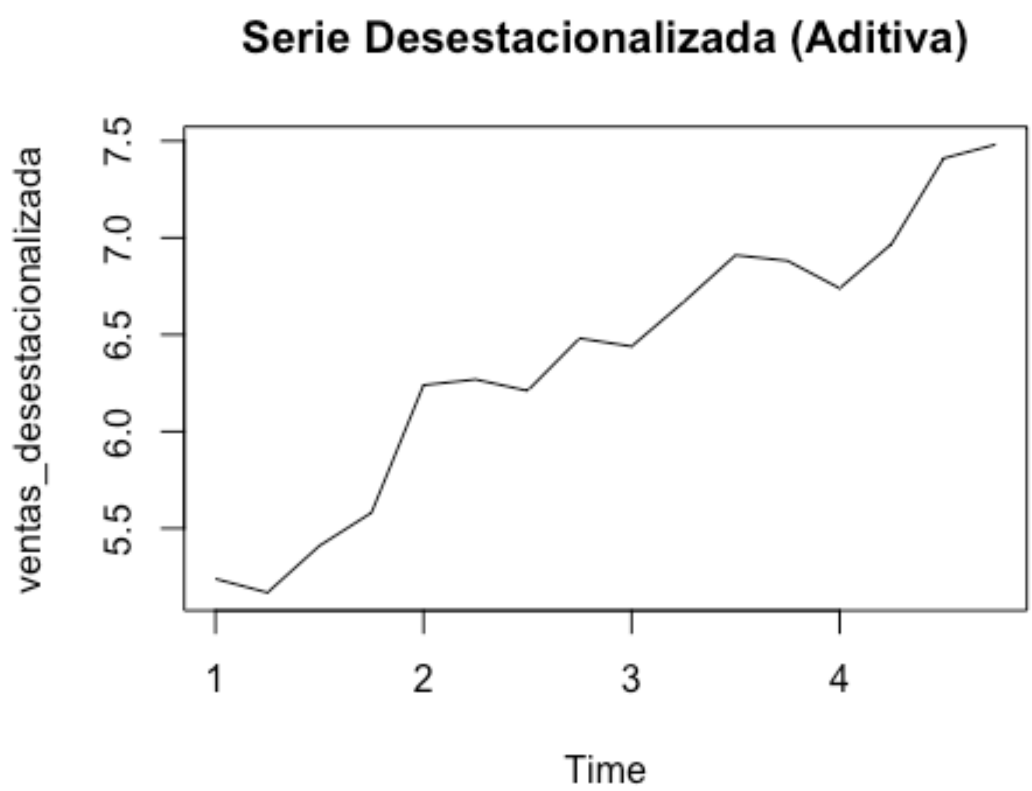



```
plot(ventas_decomp_mult)
```



Calcula los índices estacionales y grafica la serie desestacionalizada

```r
# Calcular índices estacionales
indices_estacionales <- ventas_decomp_add$seasonal

# Serie desestacionalizada (aditiva)
ventas_desestacionalizada <- ventas_ts - indices_estacionales
plot(ventas_desestacionalizada, main = "Serie Desestacionalizada (Aditiva)")
```



Analiza el modelo lineal de la tendencia

```r
# Ajuste de regresión lineal para la tendencia
tiempo <- 1:length(ventas_desestacionalizada)
modelo_tendencia <- lm(ventas_desestacionalizada ~ tiempo)

# Resumen del modelo
summary(modelo_tendencia)

##
## Call:
## lm(formula = ventas_desestacionalizada ~ tiempo)
##
## 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
```

Analiza la significancia del modelo lineal, global e individual

1. Significancia Global del Modelo

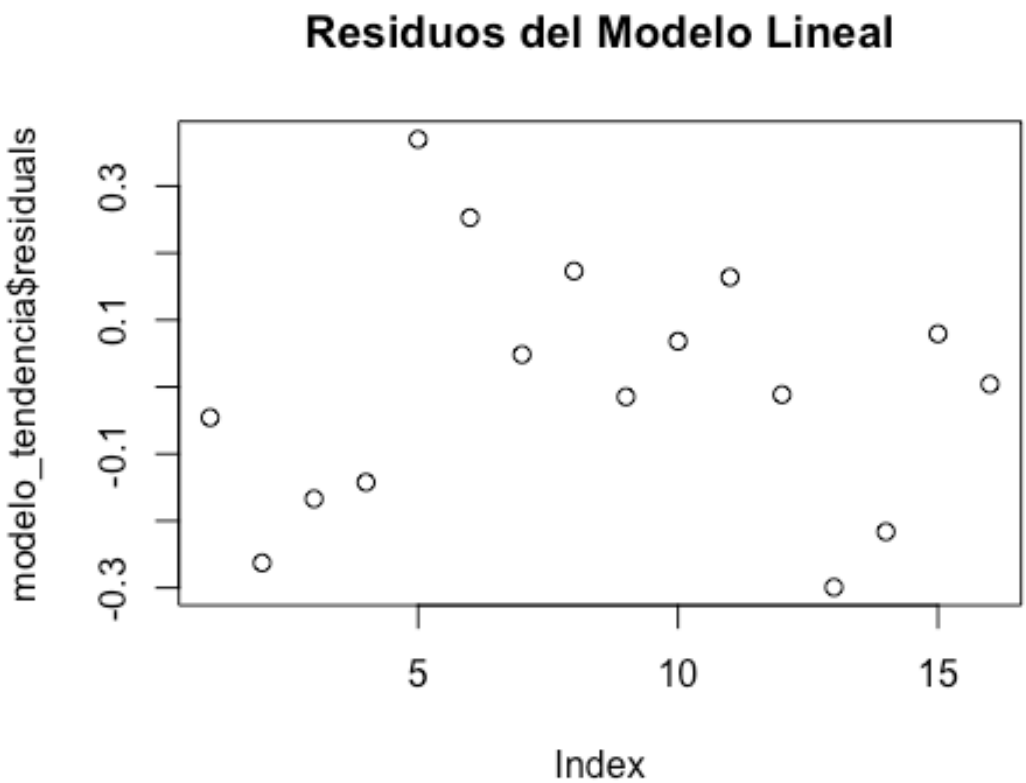
- P-valor del modelo: El valor p asociado al estadístico F es 1.399×10**−9, lo cual es mucho menor que el nivel de significancia común (por ejemplo, 0.05). Esto indica que el modelo lineal es significativo en su totalidad, es decir, que existe una relación significativa entre la variable dependiente (ventas_desestacionalizada) y la variable independiente (tiempo).
- R-cuadrado: El R**2 es 0.9324, lo que indica que el modelo explica aproximadamente el 93.24% de la variabilidad de los datos. Este es un buen ajuste, ya que el modelo es capaz de capturar gran parte de la variabilidad.

2. Significancia Individual de los Coeficientes

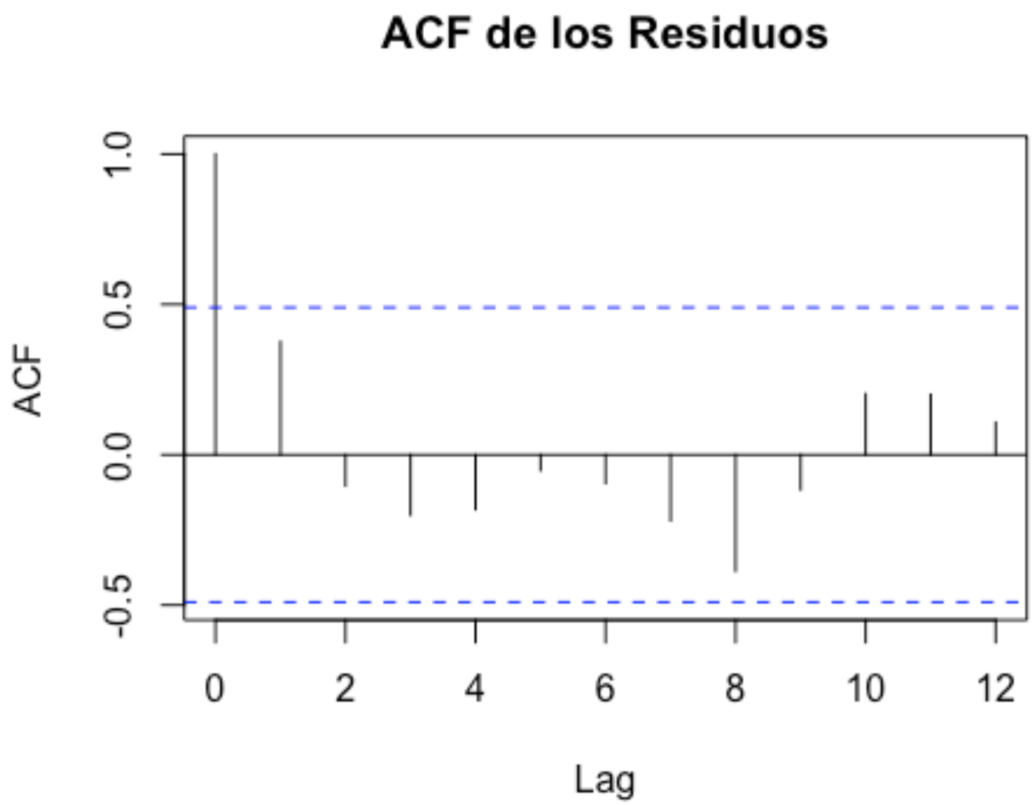
- Intercepto: El valor p del intercepto es menor a 2×10**−16, lo cual es altamente significativo. Esto indica que el intercepto tiene una contribución estadísticamente significativa al modelo.
- Pendiente (tiempo): El coeficiente de la variable tiempo tiene un valor p de 1.4×10**−9, que es también muy bajo y significativo. Esto sugiere que el tiempo tiene una relación significativa con la variable dependiente ventas_desestacionalizada, lo que valida que la tendencia es estadísticamente relevante.

Conclusión - Ambos coeficientes son altamente significativos, lo cual significa que tanto el intercepto como la variable tiempo contribuyen significativamente al modelo. Además, la significancia global del modelo y el alto valor de R**2 respaldan que el modelo lineal es adecuado para explicar la tendencia en la serie desestacionalizada.

```r
# Graficar residuos
plot(modelo_tendencia$residuals, main = "Residuos del Modelo Lineal")
```




```r
acf(modelo_tendencia$residuals, main = "ACF de los Residuos")
```



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

```
# Cálculo del CME
CME <- mean(modelo_tendencia$residuals^2, na.rm = TRUE)

# Cálculo del Error Promedio Absoluto de Medición (EPAM)
EPAM <- mean(abs(modelo_tendencia$residuals), na.rm = TRUE)

CME

## [1] 0.03291917

EPAM

## [1] 0.1449449
```

Explora un mejor modelo

```
# Ajuste de regresión cuadrática
modelo_cuadratico <- lm(ventas_desestacionalizada ~ tiempo + I(tiempo^2))

# Resumen del modelo cuadrático
summary(modelo_cuadratico)
```

```
##
## Call:
## lm(formula = ventas_desestacionalizada ~ tiempo + I(tiempo^2))
##
## 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 ***
## I(tiempo^2) -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
```

Concluye sobre el mejor modelo

Para determinar el modelo más adecuado para la serie de tiempo de las ventas de televisores, se analizaron tanto un modelo lineal como un modelo cuadrático en la serie desestacionalizada. El modelo lineal mostró un valor ajustado de R**2 de 0.9275, mientras que el modelo cuadrático obtuvo un valor ligeramente mayor de 0.9361. Sin embargo, en el modelo cuadrático, el término cuadrático no fue estadísticamente significativo ($p > 0.05$), lo que sugiere que la inclusión de este término no aporta una mejora considerable en la explicación de la variabilidad de las ventas. Además, los residuos de ambos modelos mostraron patrones de autocorrelación aceptables, indicando que no hay una gran dependencia temporal en los residuos.

Por lo tanto, considerando la simplicidad y el buen ajuste del modelo lineal, este se considera el modelo más apropiado para capturar la tendencia en las ventas de televisores. La diferencia en el ajuste entre el modelo lineal y el cuadrático es mínima, y el modelo lineal ofrece una interpretación más sencilla y robusta, especialmente cuando el término cuadrático no es significativo. En conclusión, el modelo lineal es el mejor para este análisis, proporcionando un equilibrio adecuado entre simplicidad y precisión en las predicciones.

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

```r
# Predecir para el siguiente año usando el modelo cuadrático
nuevos_tiempos <- (length(ventas_desestacionalizada) +
1):(length(ventas_desestacionalizada) + 4)
predicciones <- predict(modelo_cuadratico, newdata = data.frame(tiempo =
nuevos_tiempos))
```

```r
# Graficar los datos originales y las predicciones
plot(ventas_ts, xlim = c(1, 5), ylim = c(min(ventas_ts), max(predicciones)),
main = "Pronóstico de Ventas")
lines(c(ventas_desestacionalizada, predicciones), col = "purple", lty = 2)

# Leyenda para diferenciar las series
legend("topleft", legend = c("Datos Originales", "Pronóstico"),
       col = c("black", "purple"), lty = c(1, 2, 2))
```