act_8_A01742161

Rogelio Lizárraga

2024-11-12

SERIES DE TIEMPO

```r
library(forecast)
```

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

```r
library(tseries)
library(ggplot2)

ventas_data <- data.frame(
  Año = c(1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4),
  Trimestre = c(1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4),
  Ventas_miles = 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)
)

# View the data frame
print(ventas_data)
```

```
##    Año Trimestre Ventas_miles
## 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:

Identifica si es una serie estacionaria

H_0 = La series no es estacionaria. H_1: La serie es estacionaria.

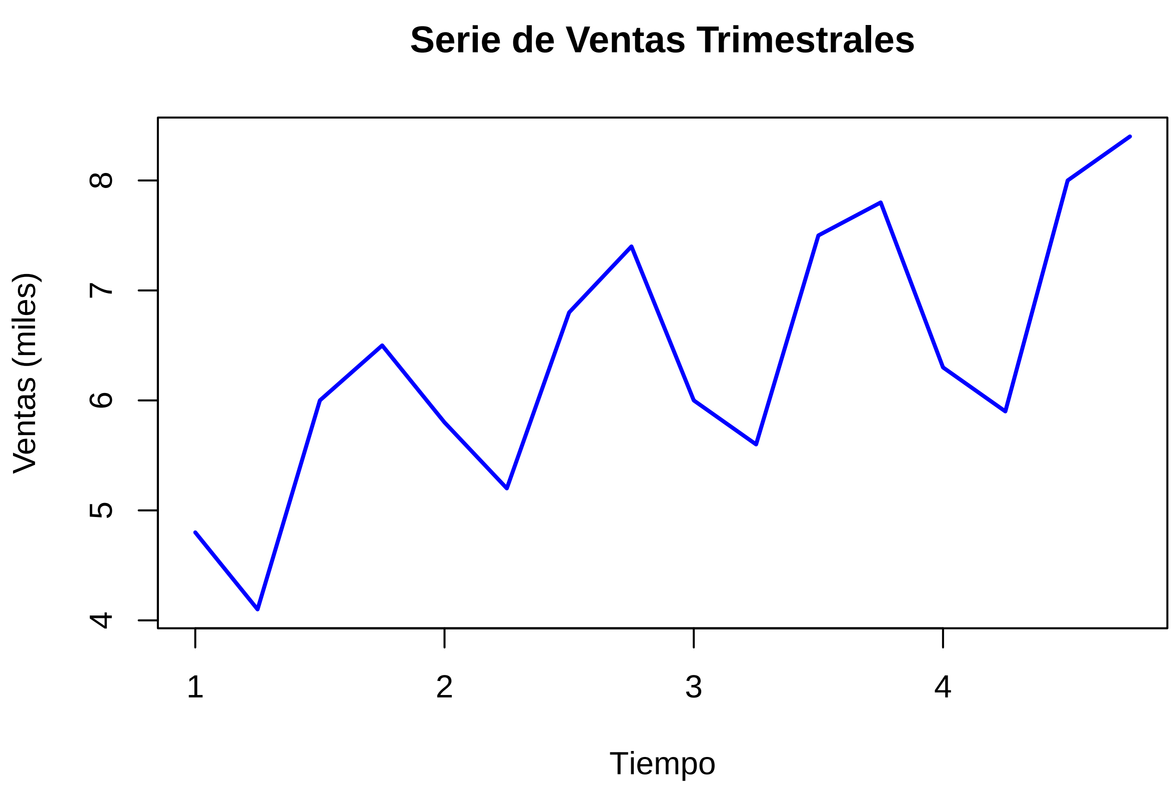
```
ventas_ts <- ts(ventas_data$Ventas_miles, start = c(1, 1), frequency = 4)
adf_test_4 <- adf.test(ventas_ts, k = 0)
print(adf_test_4)
```

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

Como podemos observar, el valor p es < 0.05, por lo que rechazamos la hipótesis inicial que establece que la serie no es estacionaria. Es decir, la serie sí es estacionaria.

Grafica la serie para verificar su tendencia y estacionalidad

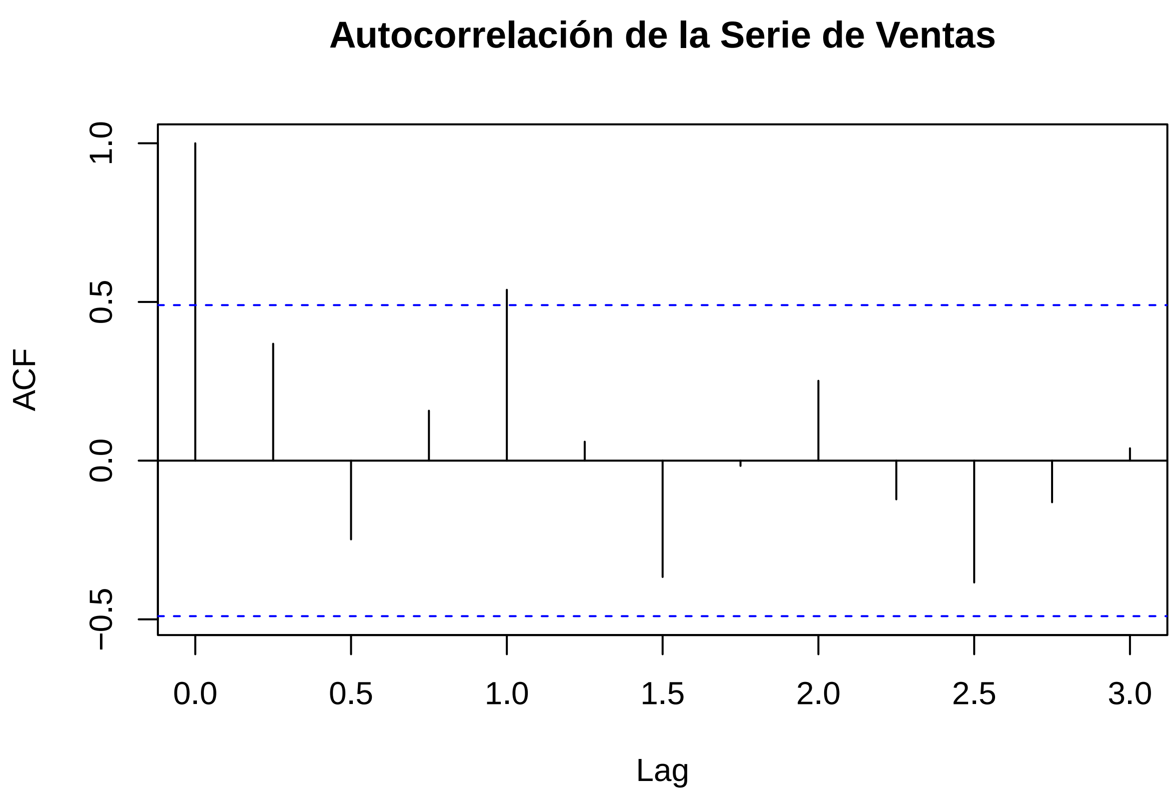
```
plot(ventas_ts, main = "Serie de Ventas Trimestrales", xlab = "Tiempo", ylab = "Ventas (miles)", col =
```



Como podemos observar que la serie de tiempo tiene una tendencia ligera a la alza. Además, observamos que hay cierta estacionalidad pues vemos como fluctúan las ventas cada año con comportamientos similares a lo largo de los trimestres.

Analiza su gráfico de autocorrelación

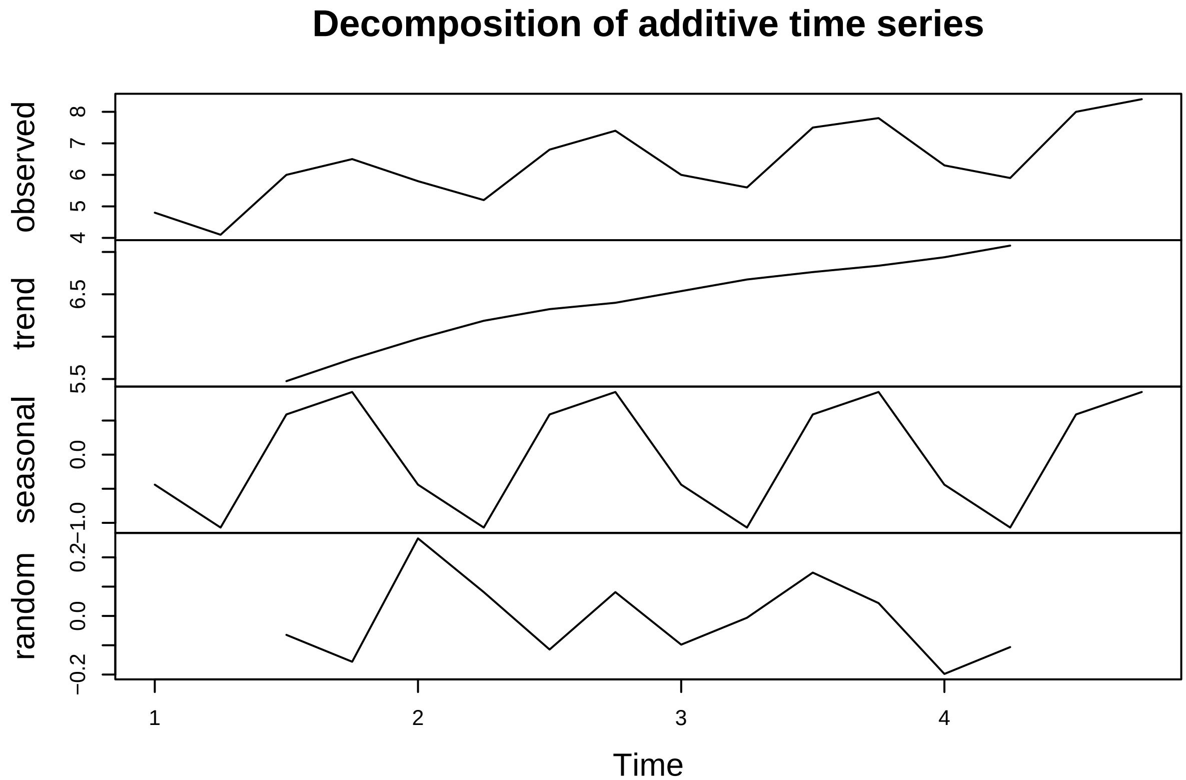
```r
acf(ventas_ts, main = "Autocorrelación de la Serie de Ventas")
```



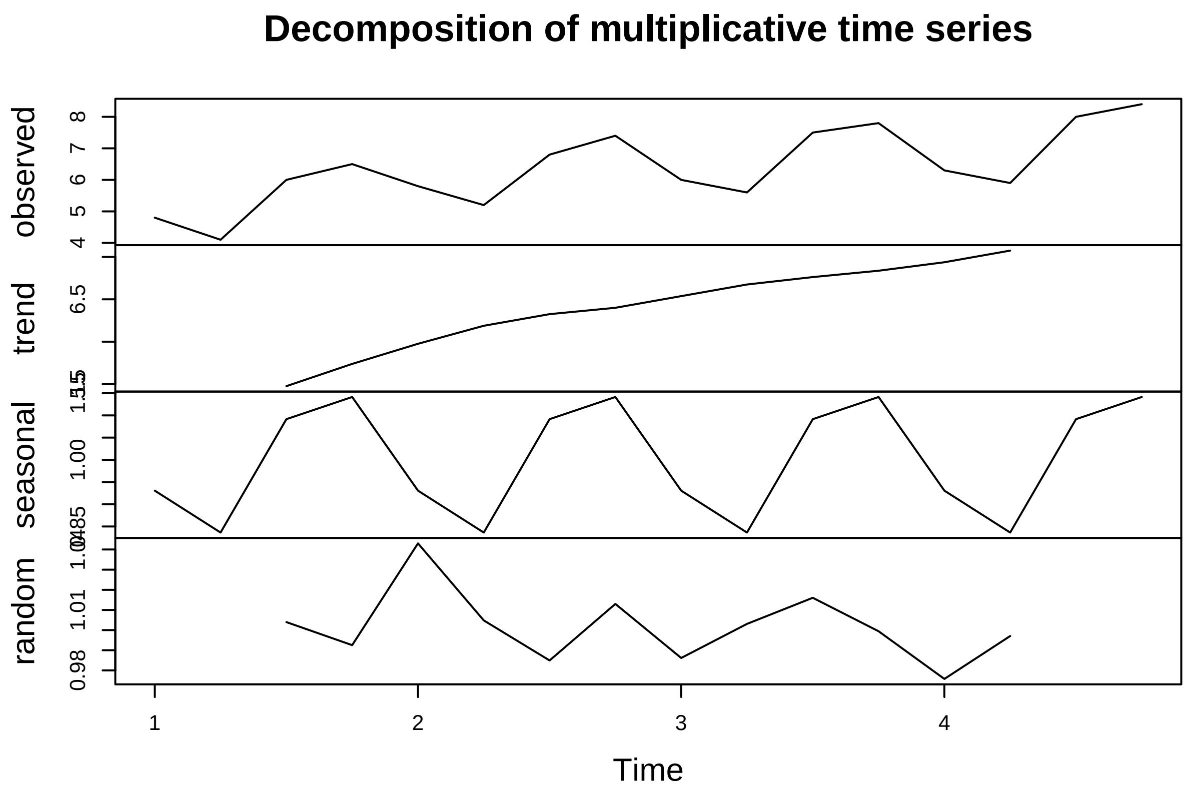
Como podemos observar, el gráfico tiene una autocorrelación de 1 en el lag 0. Sin embargo, vemos que la gran mayoría del resto de los rezagos no superan el intervalo de confianza de 0.5, lo cual indica que la estacionalidad no es tan fuerte.

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

```r
decomp_add <- decompose(ventas_ts, type = "additive")
decomp_mult <- decompose(ventas_ts, type = "multiplicative")

plot(decomp_add)
```



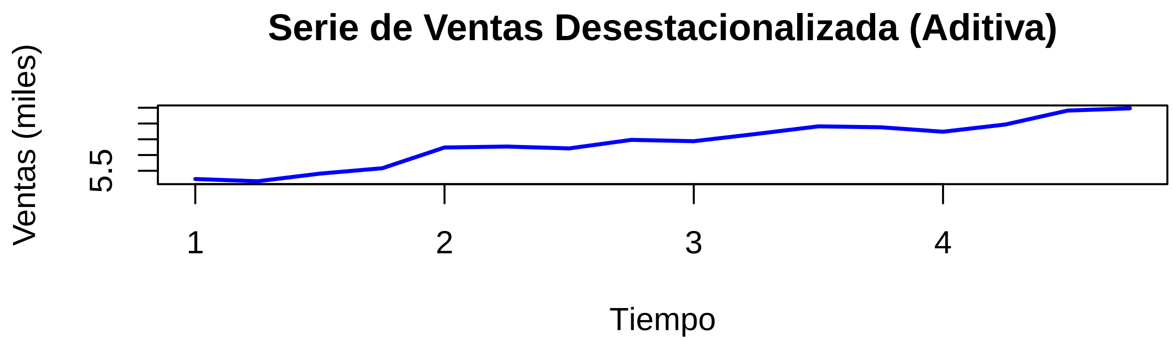

```r
plot(decomp_mult)
```

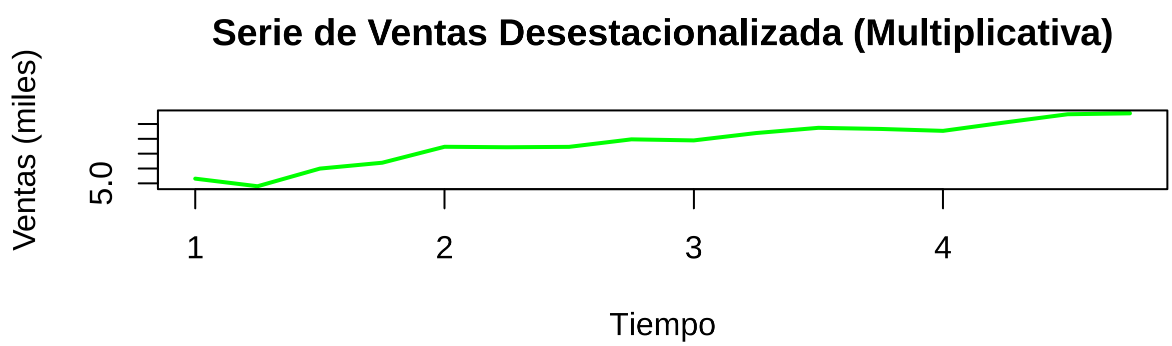


Observamos que la tendencia incrementa levemente a lo largo del tiempo en ambas decomposiciones. Vemos también como el factor estacional se comporta de manera similar para ambas decomposiciones, auque el factor aleatorio sí cambia un poco. Por ello, se probará con ambos modelos.

Calcula los índices estacionales y grafica la serie desestacionalizada

```
seasonal_indices_add <- decomp_add$seasonal
seasonal_indices_mult <- decomp_mult$seasonal
ventas_desestacionalizadas_add <- ventas_ts - seasonal_indices_add
ventas_desestacionalizadas_mult <- ventas_ts / seasonal_indices_mult
par(mfrow = c(2, 1))  # Configurar para dos gráficos verticales
plot(ventas_desestacionalizadas_add, main = "Serie de Ventas Desestacionalizada (Aditiva)", xlab = "Tie
plot(ventas_desestacionalizadas_mult, main = "Serie de Ventas Desestacionalizada (Multiplicativa)", xla
```






```r
print(seasonal_indices_add)
```

```
##         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
```

```r
print(seasonal_indices_mult)
```

```
##        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
```

Como podemos observar, ambas gráficas parecen tener un comportamiento bastante similar, por lo que será importante calcular métricas como MAE o MAPE analizar si los modelos son buenos y cuál es el mejor. ## Analiza el modelo lineal de la tendencia ### Realiza la regresión lineal de la tendencia (ventas desestacionalizadas vs tiempo) ### Analiza la significancia del modelo lineal, global e individual

```r
tiempo <- 1:length(ventas_desestacionalizadas_add)
data_tendencia_add <- data.frame(tiempo, ventas_desestacionalizadas_add)
data_tendencia_mult <- data.frame(tiempo, ventas_desestacionalizadas_mult)
```

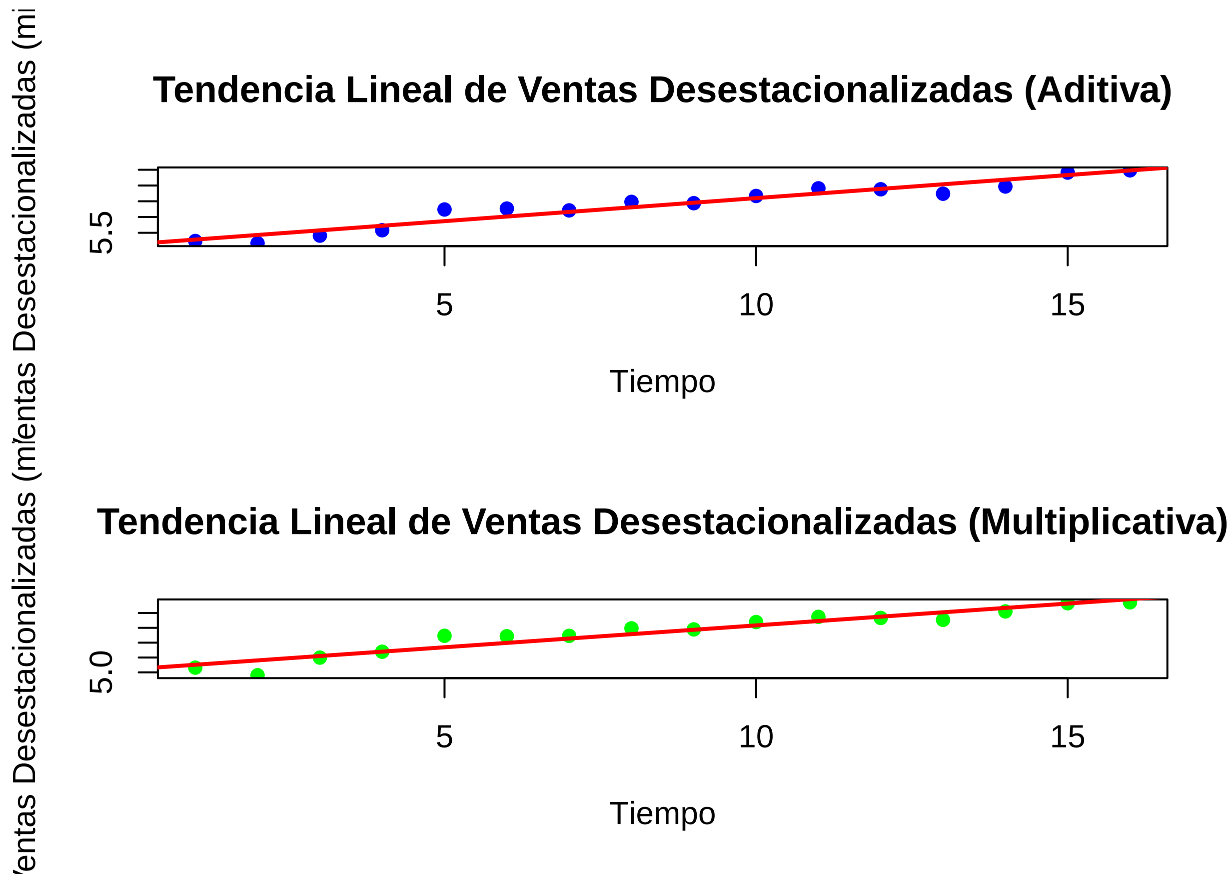
```r
modelo_tendencia_add <- lm(ventas_desestacionalizadas_add ~ tiempo, data = data_tendencia_add)
modelo_tendencia_mult <- lm(ventas_desestacionalizadas_mult ~ tiempo, data = data_tendencia_mult)
summary(modelo_tendencia_add)
```

```
##
## Call:
## lm(formula = ventas_desestacionalizadas_add ~ tiempo, data = data_tendencia_add)
##
## 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
```

```r
summary(modelo_tendencia_mult)
```

```
##
## Call:
## lm(formula = ventas_desestacionalizadas_mult ~ tiempo, data = data_tendencia_mult)
##
## 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
```

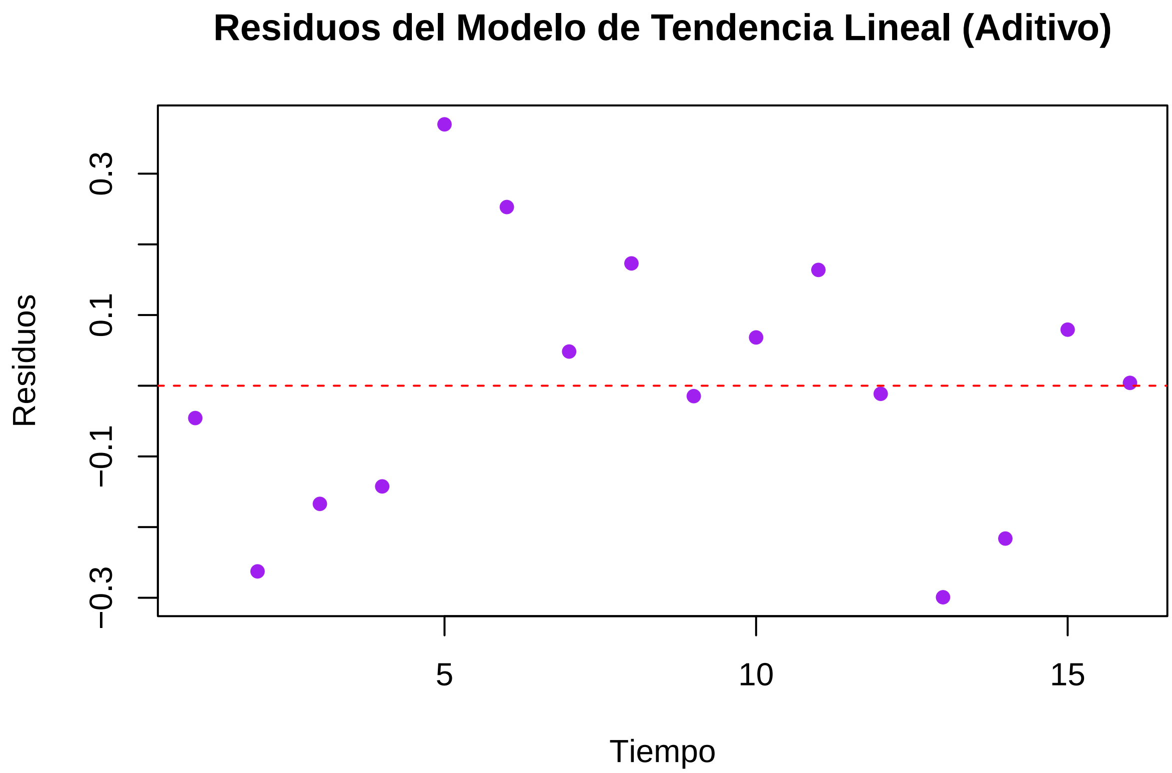
```r
par(mfrow = c(2, 1))
plot(data_tendencia_add$tiempo, data_tendencia_add$ventas_desestacionalizadas_add, main = "Tendencia Li
abline(modelo_tendencia_add, col = "red", lwd = 2)

plot(data_tendencia_mult$tiempo, data_tendencia_mult$ventas_desestacionalizadas_mult, main = "Tendencia
abline(modelo_tendencia_mult, col = "red", lwd = 2)
```



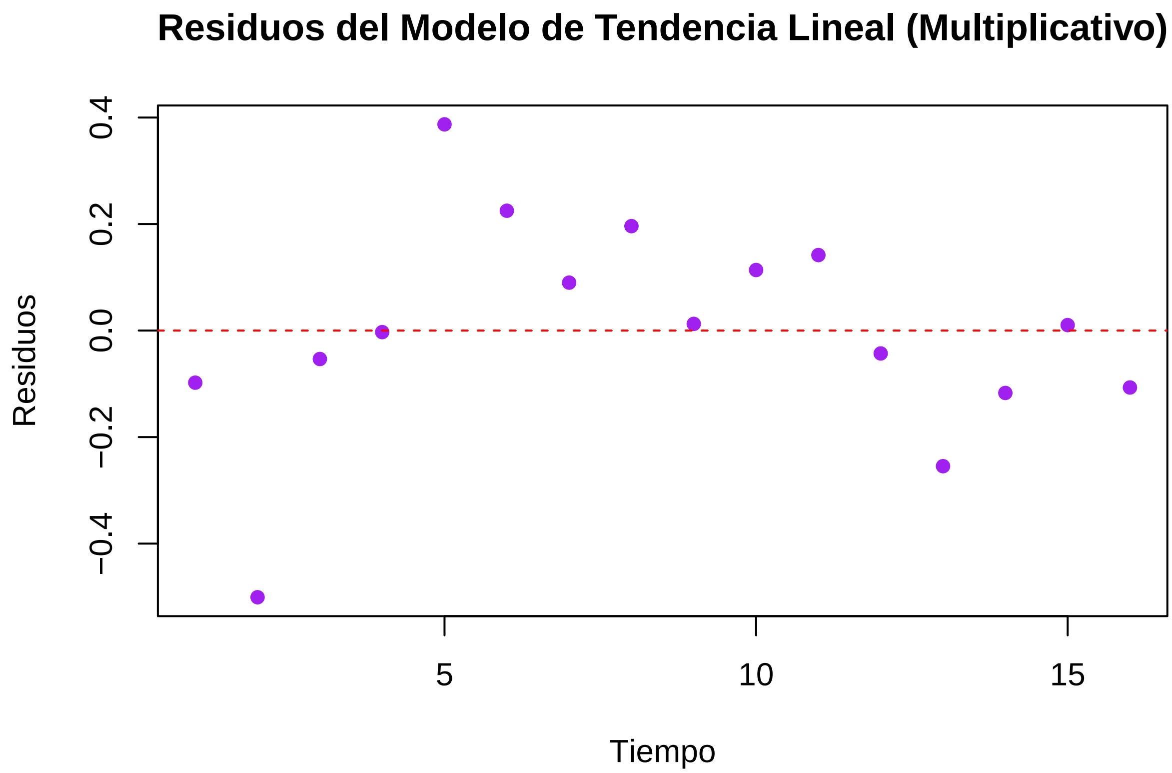
Como podemos observar, nuestros valores p son < 0.05, por lo que la hipótesis inicial se rechaza y podemos observar que nuestro modelo es significativo, así como lo es el intercepto y el tiempo para ambos modelos: aditivos y multiplicativos. Observamos que nuestros coeficientes de determinación son mayores a 90%, siendo el del modelo aditivo un poco mejor con un 92.75%, mientras que el del modelo multiplicativo es 91.51%. Esto quiere decir que el modelo aditivo explica el 92.75% de la variabilidad del tiempo.

Gráficamente, se observa que la tendencia lineal de ambos modelos es similar, con un ligero cambio en sus valores. ### Análisis de residuos

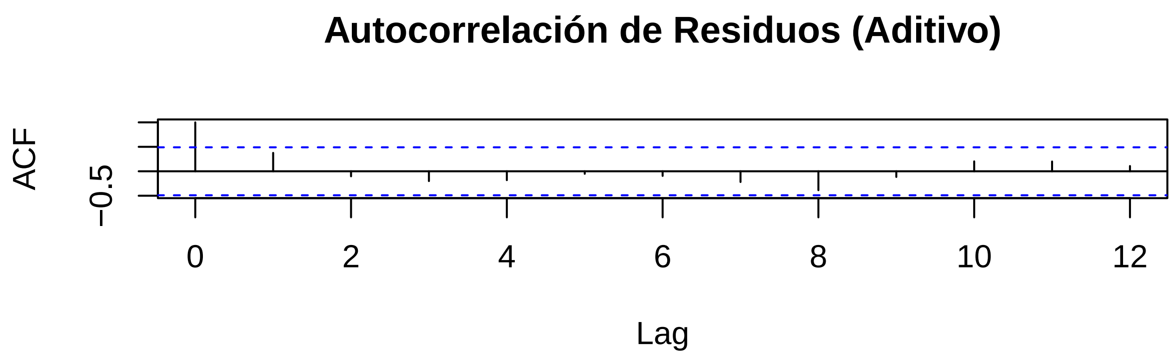
```
residuos_add <- residuals(modelo_tendencia_add)
plot(residuos_add, main = "Residuos del Modelo de Tendencia Lineal (Aditivo)", ylab = "Residuos", xlab =
abline(h = 0, col = "red", lty = 2)
```

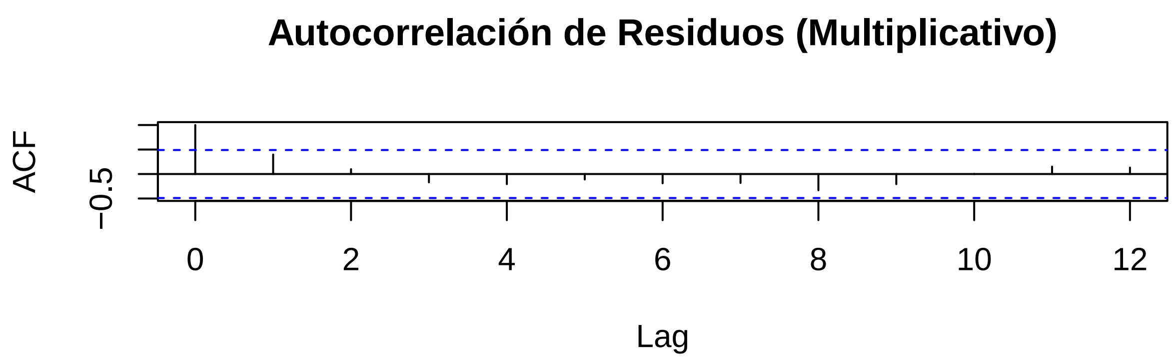



```r
residuos_mult <- residuals(modelo_tendencia_mult)
plot(residuos_mult, main = "Residuos del Modelo de Tendencia Lineal (Multiplicativo)", ylab = "Residuos
abline(h = 0, col = "red", lty = 2)
```




```
par(mfrow = c(2, 1))  # Configurar para dos gráficos verticales
acf(residuos_add, main = "Autocorrelación de Residuos (Aditivo)")
acf(residuos_mult, main = "Autocorrelación de Residuos (Multiplicativo)")
```





Como podemos observar, ambos gráficos muestran que los residuos de los modelos aditivo y multiplicativo no tienen autocorrelación significativa (las barras están dentro del intervalo de confianza). Esto indica que ambos modelos capturan bien la tendencia y estacionalidad de la serie, dejando residuos que son ruido blanco, lo cual es lo que se busca en estos modelos. # Cálculo de MSE y MAPE

```
ventas_observadas <- ventas_ts
ventas_ajustadas_add <- fitted(modelo_tendencia_add)
ventas_ajustadas_mult <- fitted(modelo_tendencia_mult)

mse_add <- mean((ventas_observadas - ventas_ajustadas_add)^2)
mape_add <- mean(abs((ventas_observadas - ventas_ajustadas_add) / ventas_observadas)) * 100

mse_mult <- mean((ventas_observadas - ventas_ajustadas_mult)^2)
mape_mult <- mean(abs((ventas_observadas - ventas_ajustadas_mult) / ventas_observadas)) * 100

cat("MSE (Aditivo):", mse_add, "\n")
```

```
## MSE (Aditivo): 0.6970679
```

```
cat("MAPE (Aditivo):", mape_add, "%\n")
```

```
## MAPE (Aditivo): 12.66041 %
```

```
cat("MSE (Multiplicativo):", mse_mult, "\n")
```

```
## MSE (Multiplicativo): 0.6957218
```

```r
cat("MAPE (Multiplicativo):", mape_mult, "%\n")
```

```
## MAPE (Multiplicativo): 12.5894 %
```

Como podemos observar, los valores del MSE y el MAPE son un poco menores en el modelo multiplicativo, por lo que el modelo multiplicativo se ajusta un poco mejor a la serie de tiempo. Sin embargo, la diferencia entre ambos modelos es muy pequeña, por lo que los dos modelos puede ser adecuado. Probaremos con otro modelo para ver si se obtienen mejores resultados que los de los dos modelos presentados

Explora un mejor modelo, por ejemplo un modelo cuadrático. Para ello transforma la variable ventas (recuerda que la regresión no lineal es una regresión lineal con una transformación).

```r
# Crear una variable tiempo al cuadrado para el modelo cuadrático
tiempo <- 1:length(ventas_desestacionalizadas_add)
tiempo2 <- tiempo^2

# Datos para el modelo cuadrático aditivo
data_cuadratica_add <- data.frame(tiempo, tiempo2, ventas_desestacionalizadas_add)

# Ajustar el modelo cuadrático aditivo
modelo_cuadratico_add <- lm(ventas_desestacionalizadas_add ~ tiempo + tiempo2, data = data_cuadratica_ad
summary(modelo_cuadratico_add)
```

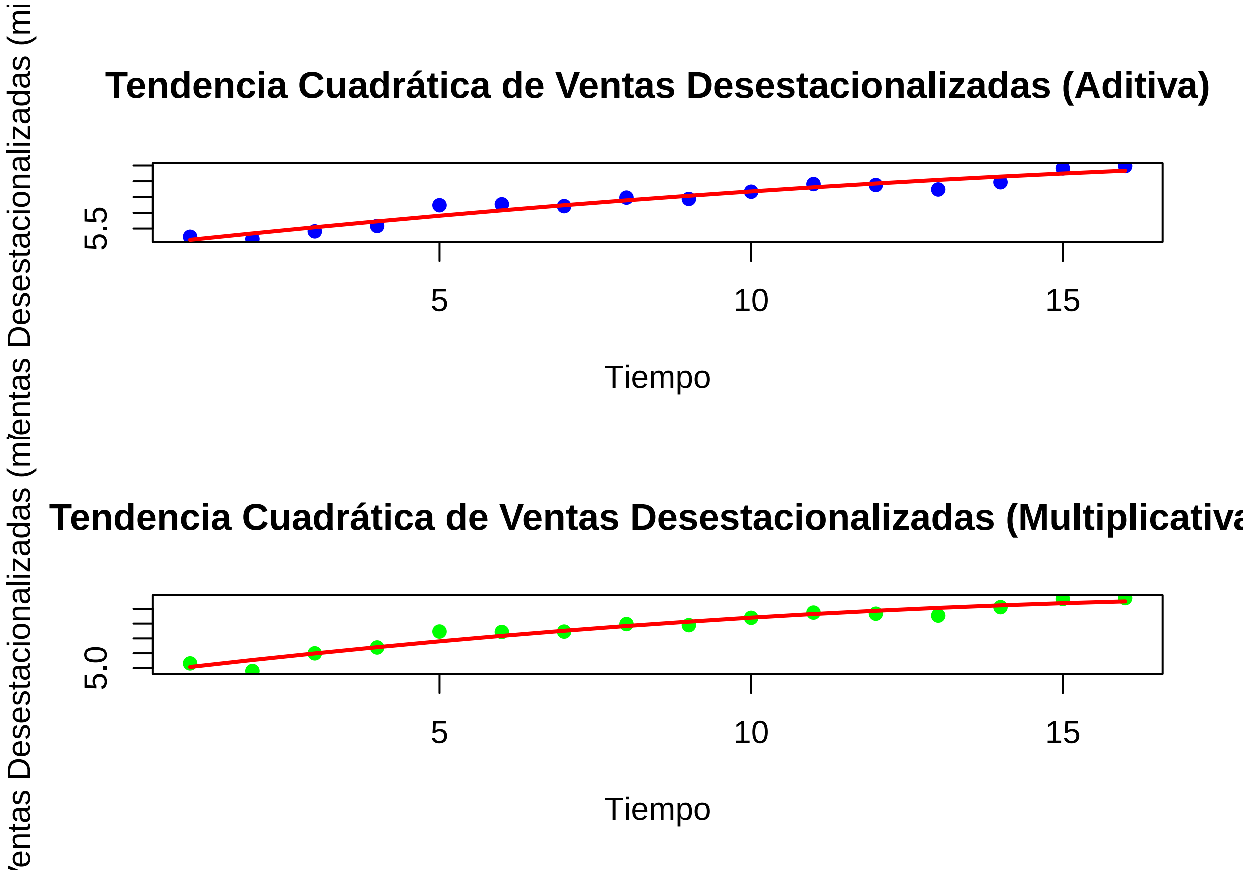
```
##
## Call:
## lm(formula = ventas_desestacionalizadas_add ~ tiempo + tiempo2,
##     data = data_cuadratica_add)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30333 -0.13440 -0.01928  0.11368  0.33301
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (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
```

```r
# Datos para el modelo cuadrático multiplicativo
data_cuadratica_mult <- data.frame(tiempo, tiempo2, ventas_desestacionalizadas_mult)
```

```r
# Ajustar el modelo cuadrático multiplicativo
modelo_cuadratico_mult <- lm(ventas_desestacionalizadas_mult ~ tiempo + tiempo2, data = data_cuadratica_
summary(modelo_cuadratico_mult)
```

```
##
## Call:
## lm(formula = ventas_desestacionalizadas_mult ~ tiempo + tiempo2,
##     data = data_cuadratica_mult)
##
## 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
```

```r
# Graficar el ajuste cuadrático para ambos modelos
par(mfrow = c(2, 1))

# Gráfico para el modelo cuadrático aditivo
plot(data_cuadratica_add$tiempo, data_cuadratica_add$ventas_desestacionalizadas_add,
     main = "Tendencia Cuadrática de Ventas Desestacionalizadas (Aditiva)", xlab = "Tiempo",
     ylab = "Ventas Desestacionalizadas (miles)", col = "blue", pch = 16)
lines(data_cuadratica_add$tiempo, predict(modelo_cuadratico_add), col = "red", lwd = 2)

# Gráfico para el modelo cuadrático multiplicativo
plot(data_cuadratica_mult$tiempo, data_cuadratica_mult$ventas_desestacionalizadas_mult,
     main = "Tendencia Cuadrática de Ventas Desestacionalizadas (Multiplicativa)", xlab = "Tiempo",
     ylab = "Ventas Desestacionalizadas (miles)", col = "green", pch = 16)
lines(data_cuadratica_mult$tiempo, predict(modelo_cuadratico_mult), col = "red", lwd = 2)
```



Como podemos observar, nuestros valores p son < 0.05, por lo que la hipótesis inicial se rechaza y tenemos evidencia para decir que nuestro modelo cuadrático es significativo, así como lo es el intercepto, el *tiempo* y el $tiempo^2$ para ambos modelos: aditivos y multiplicativos. Observamos que nuestros coeficientes de determinación son 93.61% para el modelo aditivo, y 94.05% para el multiplicativo, siendo mejor a los dos modelos previos. Esto quiere decir que el modelo cuadrático multiplicativo explica el 94.05% de la variabilidad del tiempo

ERRORES

```
# Calcular el MSE y MAPE para el modelo cuadrático aditivo
ventas_ajustadas_cuadratico_add <- predict(modelo_cuadratico_add)
mse_cuadratico_add <- mean((ventas_desestacionalizadas_add - ventas_ajustadas_cuadratico_add)^2)
mape_cuadratico_add <- mean(abs((ventas_desestacionalizadas_add - ventas_ajustadas_cuadratico_add) / ven

# Calcular el MSE y MAPE para el modelo cuadrático multiplicativo
ventas_ajustadas_cuadratico_mult <- predict(modelo_cuadratico_mult)
mse_cuadratico_mult <- mean((ventas_desestacionalizadas_mult - ventas_ajustadas_cuadratico_mult)^2)
mape_cuadratico_mult <- mean(abs((ventas_desestacionalizadas_mult - ventas_ajustadas_cuadratico_mult) /

# Imprimir los resultados
cat("MSE (Cuadrático Aditivo):", mse_cuadratico_add, "\n")
```

```
## MSE (Cuadrático Aditivo): 0.02696192
```

```
cat("MAPE (Cuadrático Aditivo):", mape_cuadratico_add, "%\n")
```

```
## MAPE (Cuadrático Aditivo): 2.229868 %
```

```r
cat("MSE (Cuadrático Multiplicativo):", mse_cuadratico_mult, "\n")
```

```
## MSE (Cuadrático Multiplicativo): 0.02584767
```

```r
cat("MAPE (Cuadrático Multiplicativo):", mape_cuadratico_mult, "%\n")
```

```
## MAPE (Cuadrático Multiplicativo): 1.949349 %
```

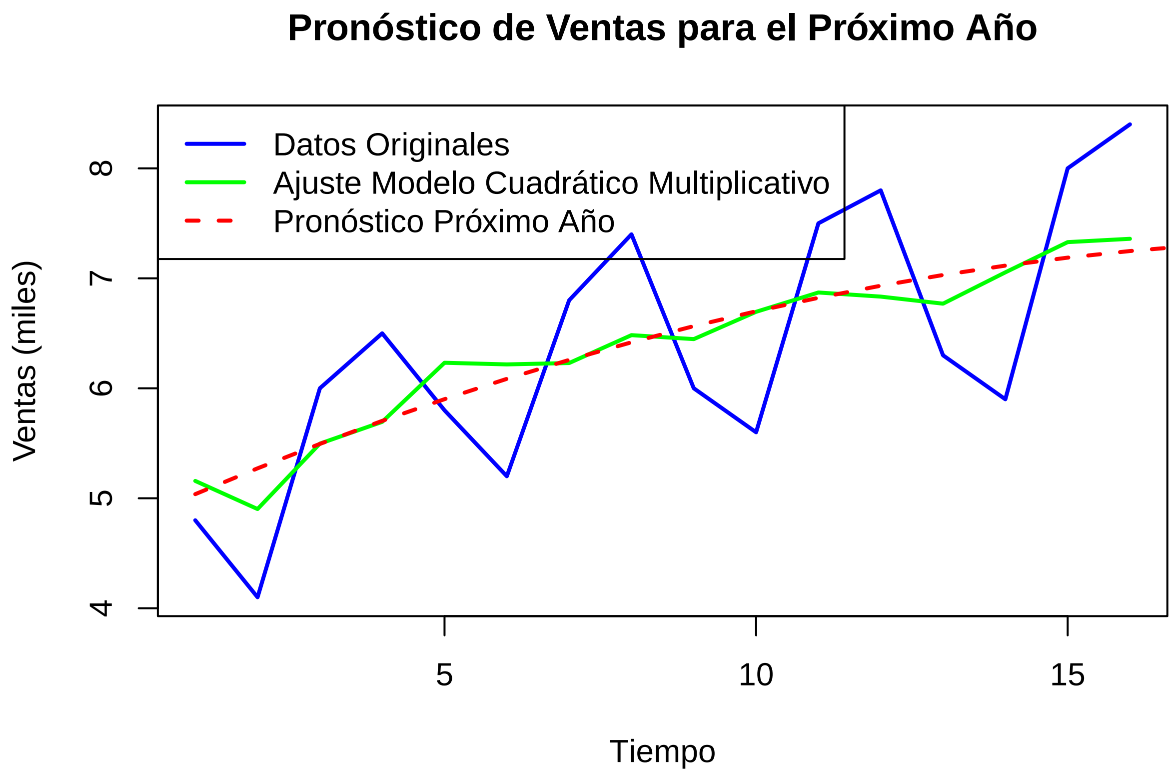
Los errores cuadráticos medios y absolutos porcentuales son mucho menores para los modelos cuadráticos que para los primeros dos modelos lineales que se realizaron. Vemos que estos errores se minimizaron de manera significativa. Dentro de los modelos cuadráticos, el multiplicatico tiene errores menores, con un MSE de 0.0258 y un MAPE del 1.9493%, siendo este el mejor modelo

Concluye sobre el mejor modelo

El modelo cuadrático multiplicativo es el mejor entre todos los modelos evaluados (lineal aditivo, lineal multiplicativo, cuadrático aditivo, y cuadrático multiplicativo) en términos de MSE, MAPE y R^2. Esto sugiere que la serie de ventas trimestrales tiene una tendencia no lineal, y un modelo cuadrático multiplicativo es el que mejor captura tanto la tendencia como la estacionalidad de la serie.

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

```r
nueva_longitud <- length(ventas_desestacionalizadas_mult) + 4
tiempo_extendido <- 1:nueva_longitud
tiempo2_extendido <- tiempo_extendido^2
data_extendido <- data.frame(tiempo = tiempo_extendido, tiempo2 = tiempo2_extendido)
pronostico_cuadratico_mult <- predict(modelo_cuadratico_mult, newdata = data_extendido)
pronostico_anio_siguiente <- pronostico_cuadratico_mult[(length(ventas_desestacionalizadas_mult) + 1):nu
par(mfrow = c(1, 1))
plot(tiempo_extendido[1:length(ventas_ts)], ventas_ts, type = "l", col = "blue", lwd = 2,
     xlab = "Tiempo", ylab = "Ventas (miles)", main = "Pronóstico de Ventas para el Próximo Año")
lines(tiempo_extendido[1:length(ventas_desestacionalizadas_mult)], ventas_desestacionalizadas_mult, col
lines(tiempo_extendido, pronostico_cuadratico_mult, col = "red", lwd = 2, lty = 2)
legend("topleft", legend = c("Datos Originales", "Ajuste Modelo Cuadrático Multiplicativo", "Pronóstico
       col = c("blue", "green", "red"), lwd = 2, lty = c(1, 1, 2))
```



Como se puede observar, el modelo cuadrático multiplicativo proyecta un aumento gradual en las ventas para el próximo año, lo cual hace sentido, siguiendo la tendencia de la serie de tiempo Sin embargo, dado que el modelo suaviza la serie, el pronóstico no refleja la estacionalidad trimestral presente en los datos originales. Para ello, sería mejor implementar distintos modelos que tomen en cuenta estos factores, como pudiese ser ARIMA, o incluso mejor, SARIMA.