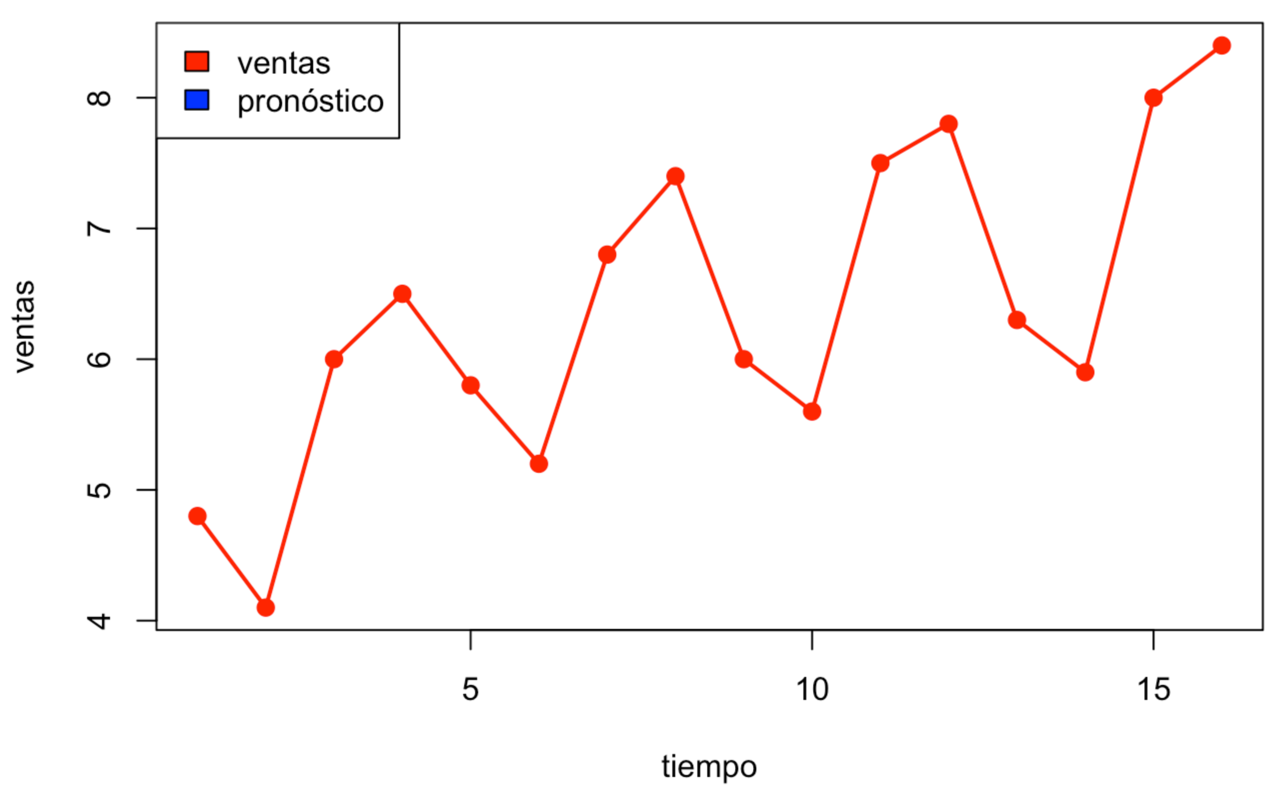
A8. Series de tiempo

Oscar Gutierrez

2024-11-12

Series de tiempo

```
ventas = c(4.8, 4.1, 6, 6.5, 5.8, 5.2, 6.8, 7.4, 6, 5.6, 7.5, 7.8, 6.3, 5.9, 8, 8.4)
x= ts(ventas, frequency = 4, start(c(2016,1)))
tiempo = 1:16
plot(tiempo, ventas, col ="red", type = "o", lwd = 2, pch = 19)
# lines(tiempo, pron, col = "blue", type = "o", lwd =2, lty = 5)
legend("topleft", legend = c("ventas", "pronóstico"), fill = c("red", "blue"))
```



No es una serie estacionaria, en el gráfico se puede ver claramente como tiene una tendencia positiva, es decir, la media de la serie va subiendo con el paso del tiempo. La varianza sí parece mantenerse constante a lo largo de las observaciones.

Se puede confirmar utilizando la prueba de dickey-fuller

H_0: La serie no es estacionaria H_1: La serie es estacionaria

```r
library(tseries)
```

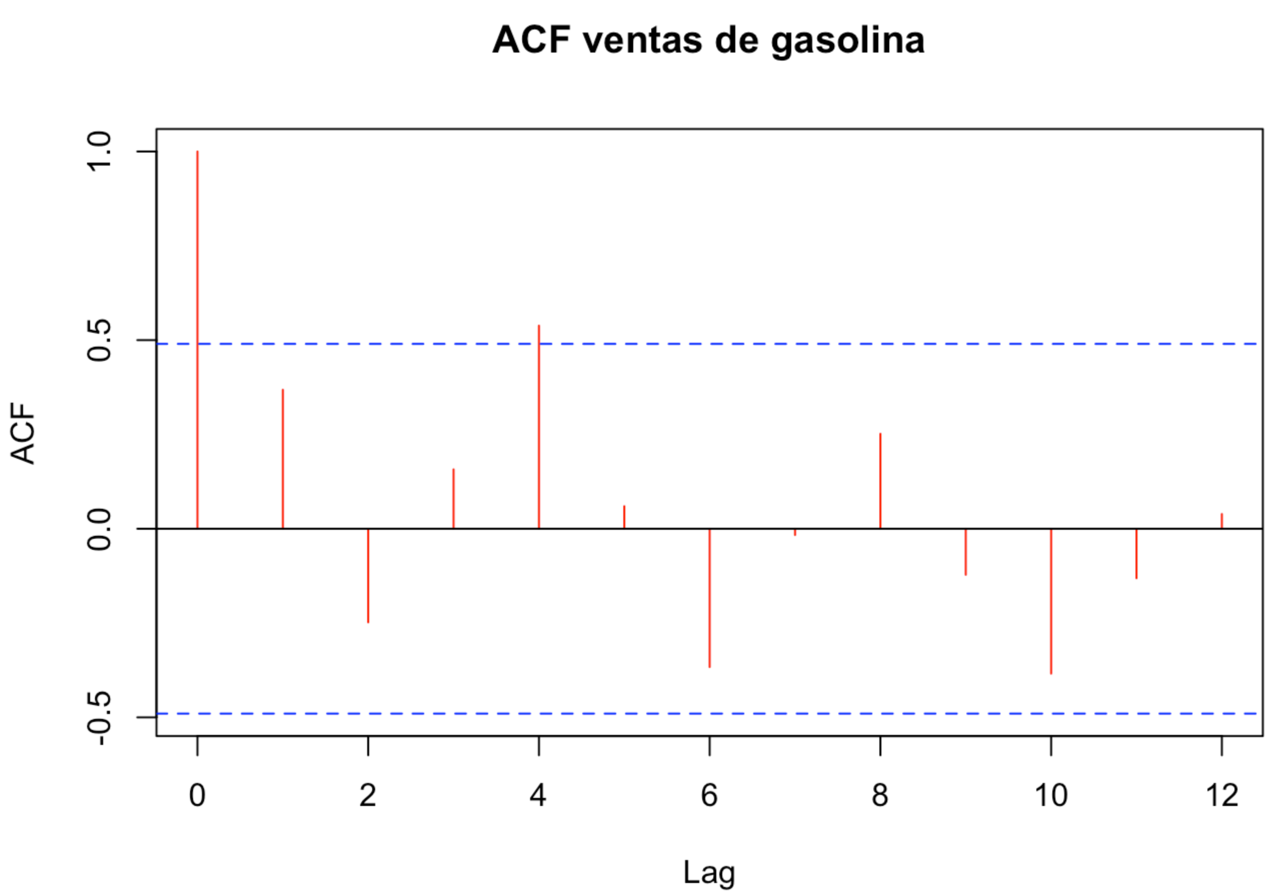
```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
adf.test(ventas, k=0)
```

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

Si se considera un $alpha = 0.05$, la prueba demuestra que no hay evidencia para rechazar H_0, por lo que la serie no es estacionaria.

```r
acf(ventas, col="red", main="ACF ventas de gasolina")
```



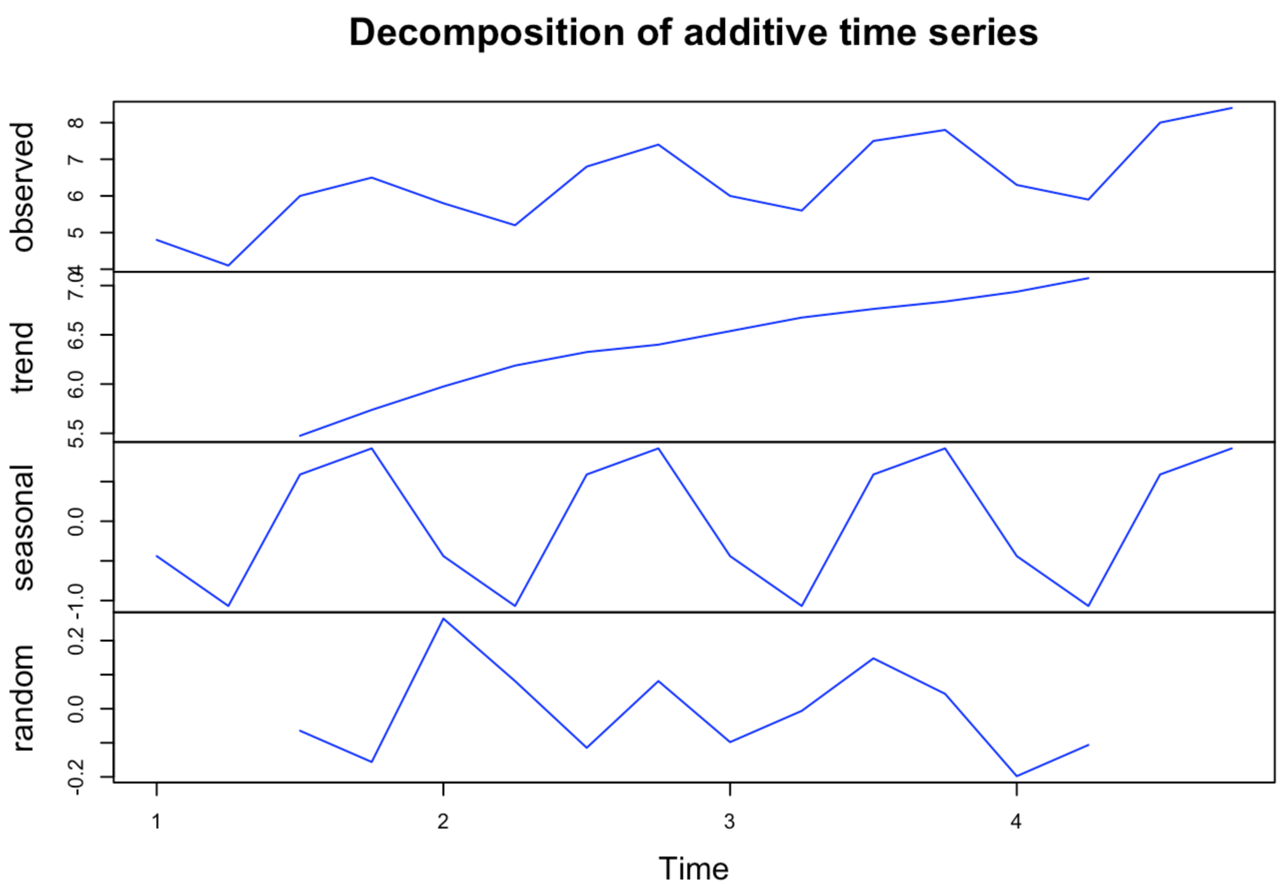

```r
qnorm(1-0.05/2)/sqrt(length(ventas))
```

```
## [1] 0.489991
```

Las autocorrelaciones parecen distribuirse alrededor del 0, solamente hay un lag que obtiene un coeficiente de correlación significativo, siendo este el lag = 1, con un valor de más de 0.5.

Serie aditiva

```
T1 = decompose(x) #para una serie multiplicativa añade: type="m"
plot(T1, col ="blue")
```

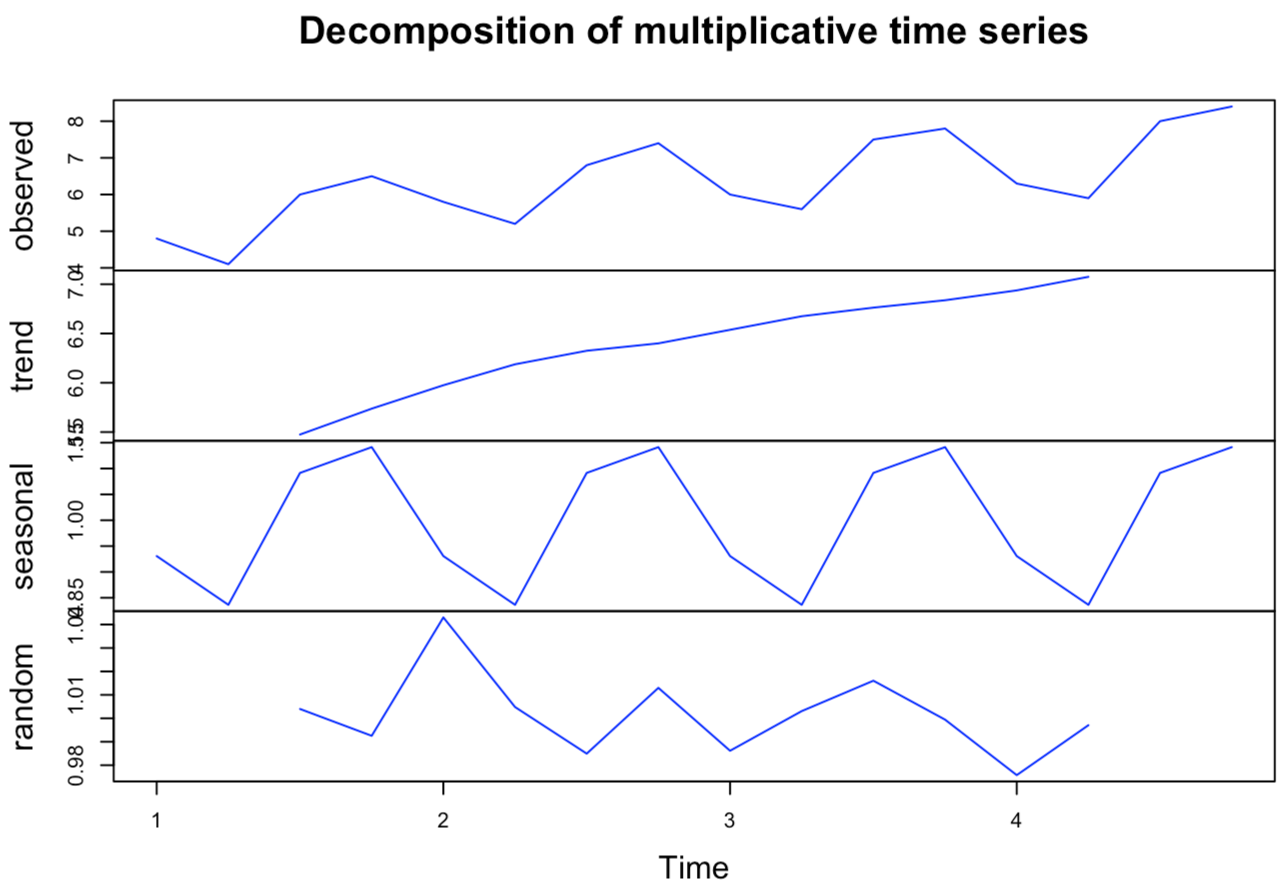



```
T1$seasonal
```

```
##         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 multiplicativa

```
T2 = decompose(x, type = 'm') #para una serie multiplicativa añade: type="m"
plot(T2, col ="blue")
```




```
T2$seasonal
```

```
##        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
```

Observaciones

La forma de las gráficas de la descomposition de las series multiplicativa y aditiva son muy similares, se realizarán ambos modelos para verificar cual es el mejor para predecir las ventas.

Modelos

En esta sección se evaluarán regresiones lineales considerando las series aditiva y multiplicativa, donde se evaluará la significancia de las variables y del modelo considerando un alpha de 0.05.

Hipótesis de variables

$H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0$

Hipótesis de modelo

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_n = 0$

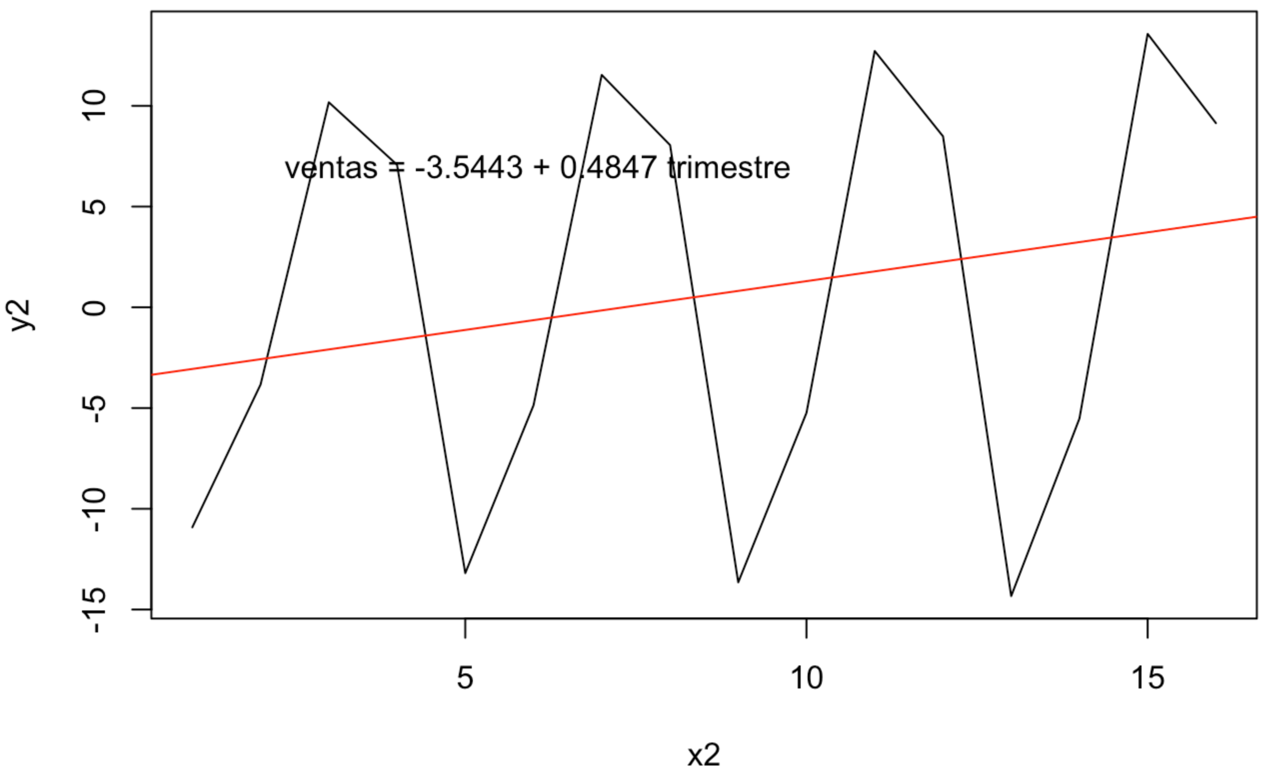
H_1 : Al menos un $\beta_i \neq 0$

Modelo aditivo

```
ventas_desestacionalizadas = (T1$x)/(T1$seasonal)
x2 = 1:16
y2 = ventas_desestacionalizadas
N2 = lm(y2~x2)
summary(N2)
```

```
##
## Call:
## lm(formula = y2 ~ x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.088  -8.085   1.836   8.971  12.267
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.5443     5.5166  -0.642    0.531
## x2            0.4847     0.5705   0.850    0.410
##
## Residual standard error: 10.52 on 14 degrees of freedom
## Multiple R-squared:  0.04902,    Adjusted R-squared:  -0.0189
## F-statistic: 0.7217 on 1 and 14 DF,  p-value: 0.4099
```

```
plot(x2, y2, type = "l")
abline(N2, col = "red")
text(6, 7, " ventas = -3.5443 + 0.4847 trimestre")
```



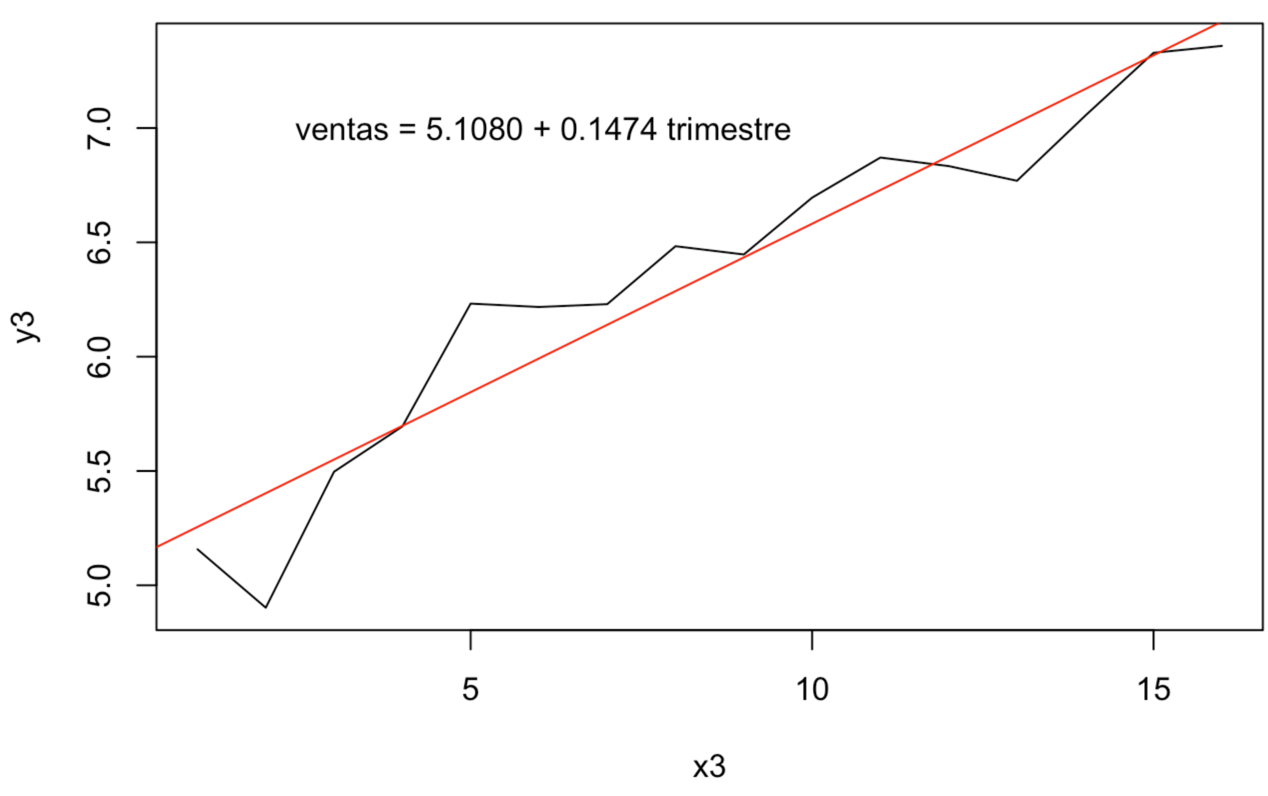
Se puede observar claramente que el modelo aditivo no es adecuado para predecir las ventas, el modelo no cuenta con ninguna variable significativa, además de que el modelo en sí tampoco es significativo considerando un alpha de 0.05.

Modelo multiplicativo

```
ventas_desestacionalizadas = (T2$x)/(T2$seasonal)
x3 = 1:16
y3 = ventas_desestacionalizadas
N3 = lm(y3~x3)
summary(N3)
```

```
##
## Call:
## lm(formula = y3 ~ x3)
##
## 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 ***
## x3           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
```

```
plot(x3, y3, type = "l")
abline(N3, col = "red")
text(6, 7, " ventas = 5.1080 + 0.1474 trimestre")
```



A diferencia del modelo aditivo, este modelo sí cuenta con 2 variables significativas (el intercept y x), además de que el modelo también es significativo cuando se considera un alpha de 0.05. El coeficiente R2 ajustado es de 0.9151, esto significa que logra explicar aproximadamente un 92% de la variación en la variable de interés.

Análisis de residuos del modelo multiplicativo

Normalidad de residuos

H_0 : Los residuos se distribuyen normalmente

H_1 : Los residuos no se distribuyen normalmente

```
library(nortest)
ad.test(N3$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  N3$residuals
## A = 0.28897, p-value = 0.5686
```

De acuerdo con la prueba de normalidad de anderson darling, los residuos del modelo se distribuyen normalmente puesto que obtuvo un valor p mayor a alpha, por lo que no se tiene evidencia para rechazar la hipótesis nula.

Comprobar media = 0

$H_0 : \mu = 0$

$H_1 : \mu \neq 0$

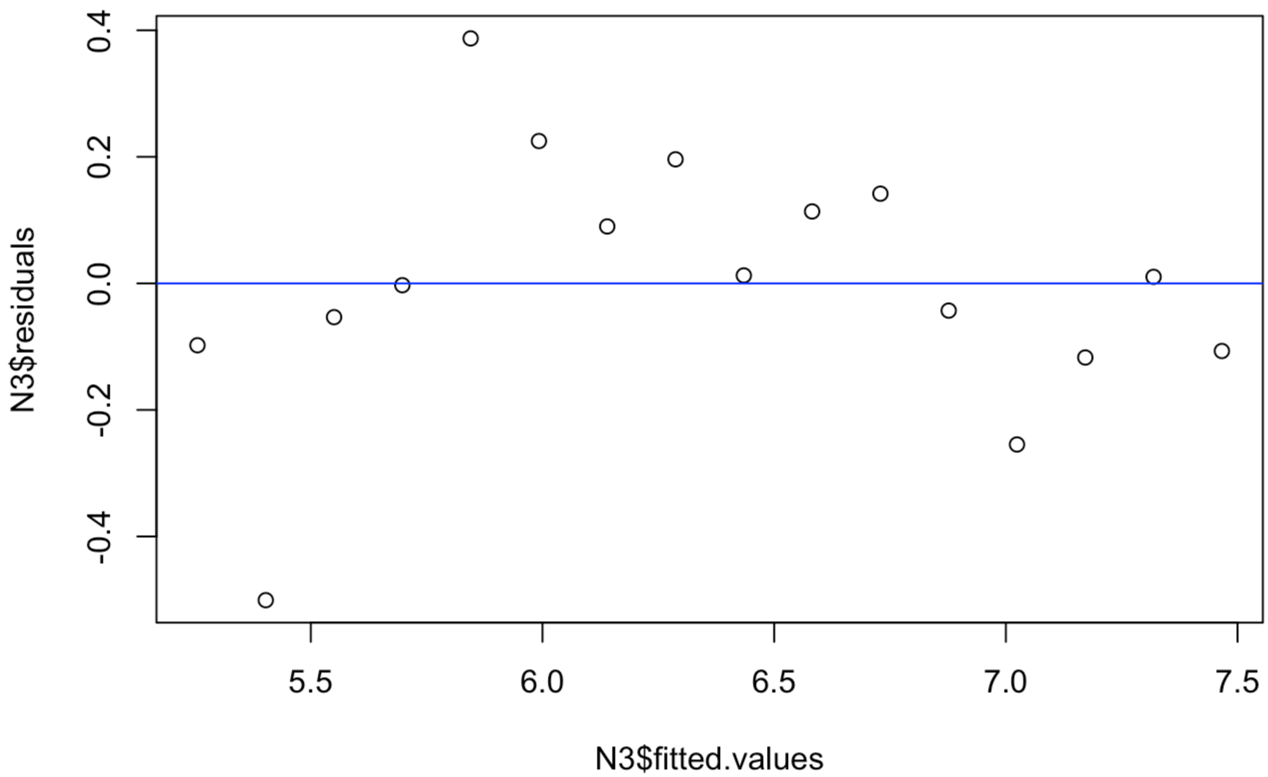
```
t.test(N3$residuals)
```

```
##
##  One Sample t-test
##
## data:  N3$residuals
## t = 1.8544e-16, df = 15, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.1096629  0.1096629
## sample estimates:
##    mean of x
## 9.540979e-18
```

De acuerdo con la prueba de hipótesis realizada, no se tiene evidencia para rechazar la hipótesis nula: la media de los residuos es = 0.

Homocedasticidad e independencia

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



No parece haber ningún patrón, la varianza es constante

Prueba de independencia

H_0: Los errores no están autocorrelacionados.

H_1: Los errores están autocorrelacionados.

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
dwtest(N3)
```

```
##
##  Durbin-Watson test
##
## data:  N3
## DW = 1.1733, p-value = 0.01753
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(N3)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 1
##
## data:  N3
## LM test = 2.5787, df = 1, p-value = 0.1083
```

Una de las pruebas realizadas indica que sí hay autocorrelación de los errores, es importante considerar esto para determinar la validez del modelo.

MSE y MAPE

Se predicen los valores usando el modelo y se calcula el error cuadrático medio y el error porcentual absoluto medio.

```
MSE = mean((ventas - N3$fitted.values)^2)
cat('Error cuadratico medio:', MSE)
```

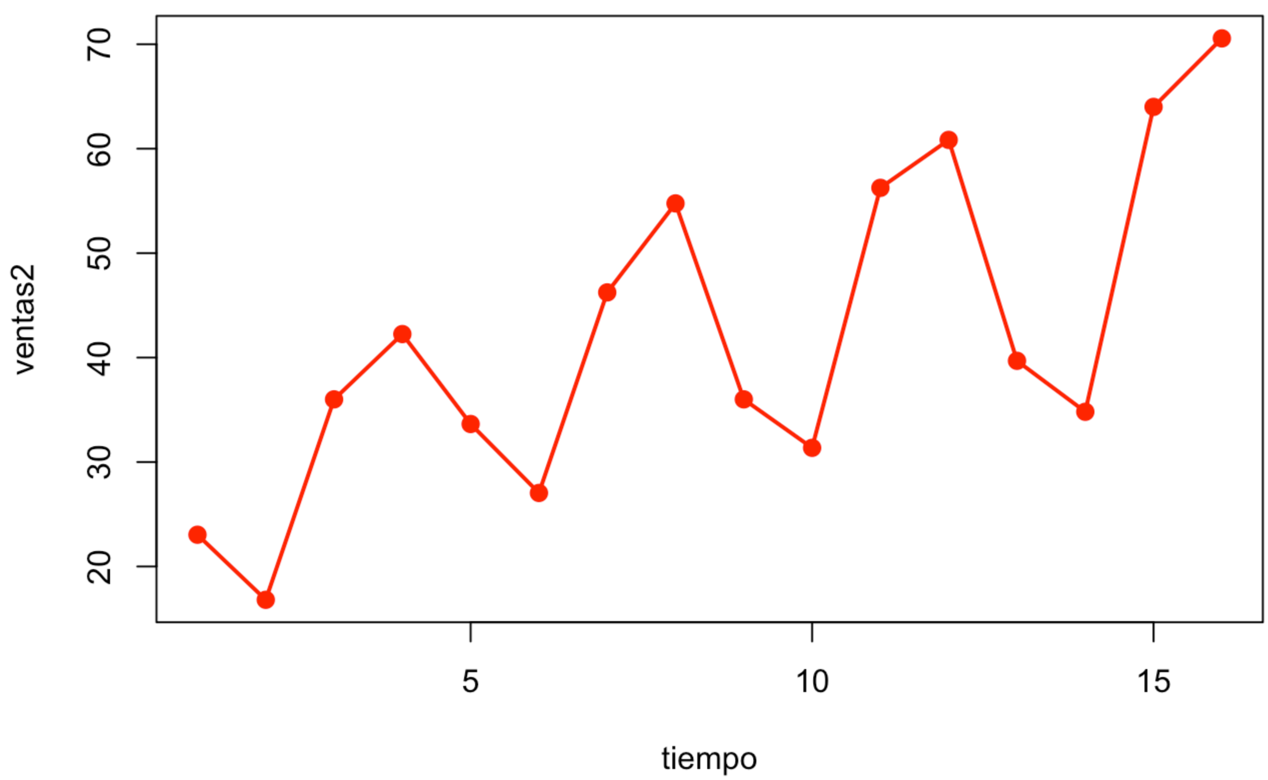
```
## Error cuadratico medio: 0.6957218
```

```
MAPE = mean(abs((ventas - N3$fitted.values)) / ventas) * 100
cat('\nError porcentual absoluto medio: ', MAPE,'%', sep = '')
```

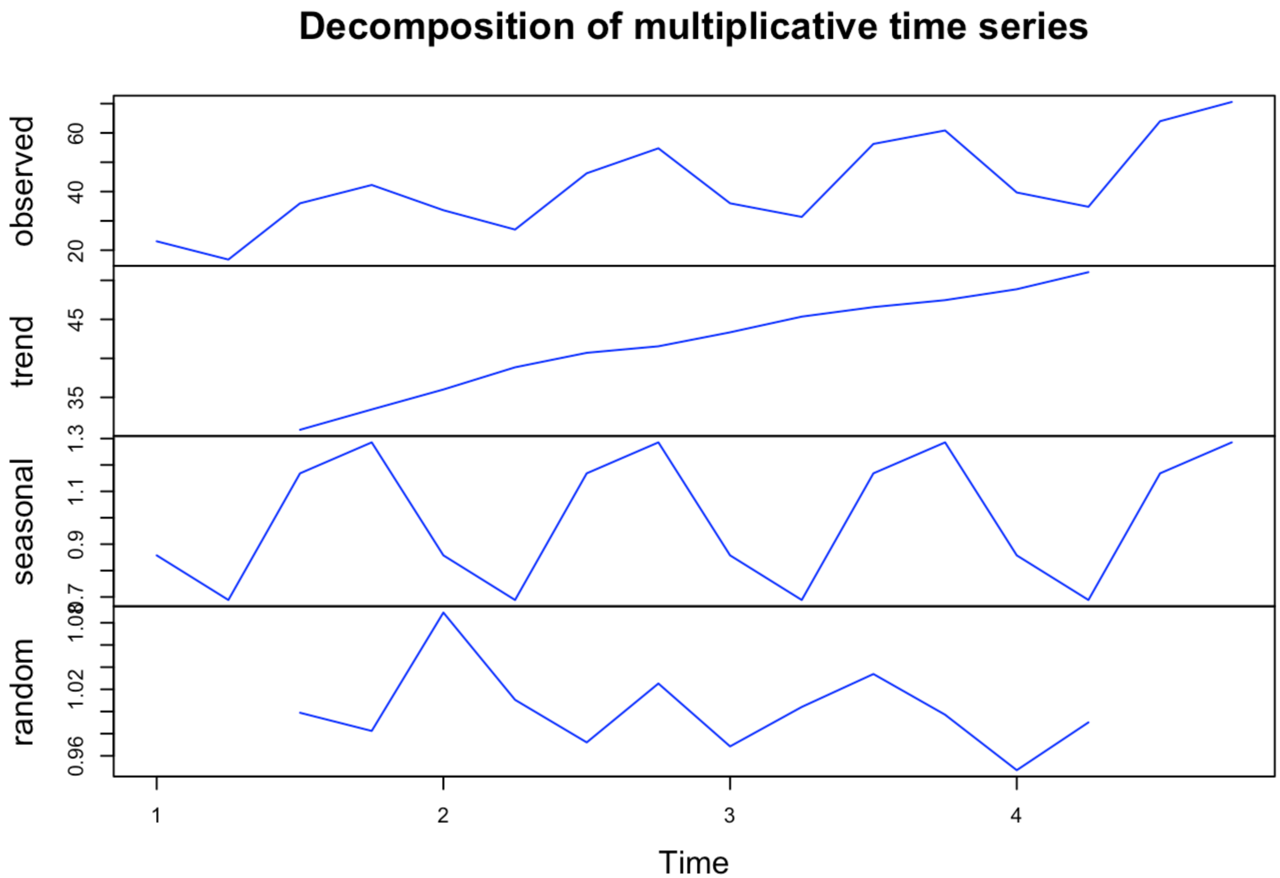
```
##
## Error porcentual absoluto medio: 12.5894%
```

Modelo cuadrático

```
ventas = c(4.8, 4.1, 6, 6.5, 5.8, 5.2, 6.8, 7.4, 6, 5.6, 7.5, 7.8, 6.3, 5.9, 8, 8.4)
ventas2 = ventas^2
x= ts(ventas2, frequency = 4, start(c(2016,1)))
tiempo = 1:16
plot(tiempo, ventas2, col ="red", type = "o", lwd = 2, pch = 19)
```




```r
T3 = decompose(x, type = 'm') #para una serie multiplicativa añade: type="m"
plot(T3, col ="blue")
```




```
T3$seasonal
```

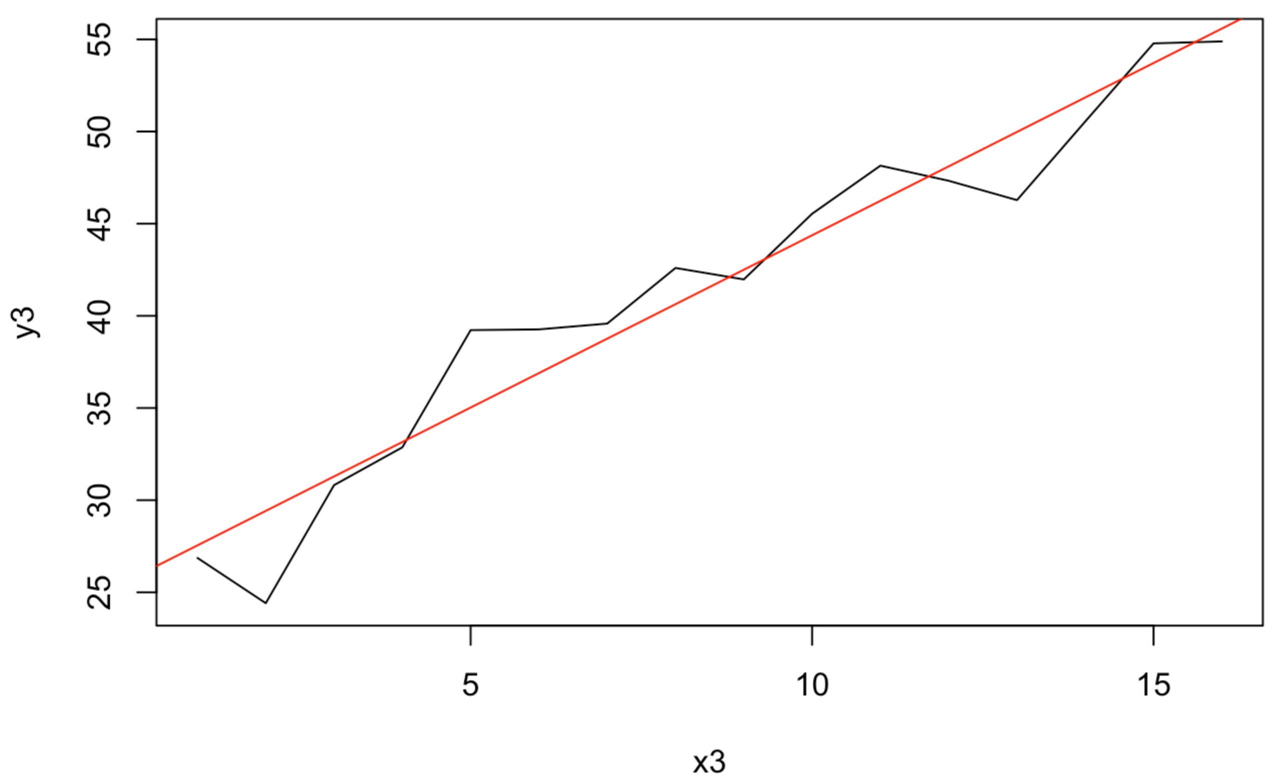
```
##        Qtr1      Qtr2      Qtr3      Qtr4
## 1 0.8576109 0.6886472 1.1682611 1.2854808
## 2 0.8576109 0.6886472 1.1682611 1.2854808
## 3 0.8576109 0.6886472 1.1682611 1.2854808
## 4 0.8576109 0.6886472 1.1682611 1.2854808
```

```
ventas_desestacionalizadas = (T3$x)/(T3$seasonal)
x3 = 1:16
y3 = ventas_desestacionalizadas

modelo_cuadratico <- lm(y3 ~ x3) # y3 son las ventas desestacionalizadas de la serie mul
tiplicativa
summary(modelo_cuadratico)
```

```
##
## Call:
## lm(formula = y3 ~ x3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0069 -0.7233 -0.3806  1.3491  4.1991
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  25.6778     1.2260   20.94 5.75e-12 ***
## x3            1.8697     0.1268   14.75 6.38e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.338 on 14 degrees of freedom
## Multiple R-squared:  0.9395, Adjusted R-squared:  0.9352
## F-statistic: 217.4 on 1 and 14 DF,  p-value: 6.378e-10
```

```
plot(x3, y3, type = "l")
abline(modelo_cuadratico, col = "red")
```



Hipótesis de variables

$H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0$

Hipótesis de modelo

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_n = 0$

$H_1 :$ Al menos un $\beta_i \neq 0$

Tanto las variables como el modelo son significativos considerando un alpha de 0.05.

Este modelo obtiene un coeficiente r2 ajustado de 0.9352, esto quiere decir que logra explicar aproximadamente un 94% de la variación en la variable de interés (ventas).

Análisis de residuos del modelo cuadratico

Normalidad de residuos

H_0 : Los residuos se distribuyen normalmente

H_1 : Los residuos no se distribuyen normalmente

```
library(nortest)
ad.test(modelo_cuadratico$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  modelo_cuadratico$residuals
## A = 0.42454, p-value = 0.2788
```

De acuerdo con la prueba de normalidad de anderson darling, los residuos del modelo se distribuyen normalmente puesto que obtuvo un valor p mayor a alpha, por lo que no se tiene evidencia para rechazar la hipótesis nula.

Comprobar media = 0

$H_0 : \mu = 0$

$H_1 : \mu \neq 0$

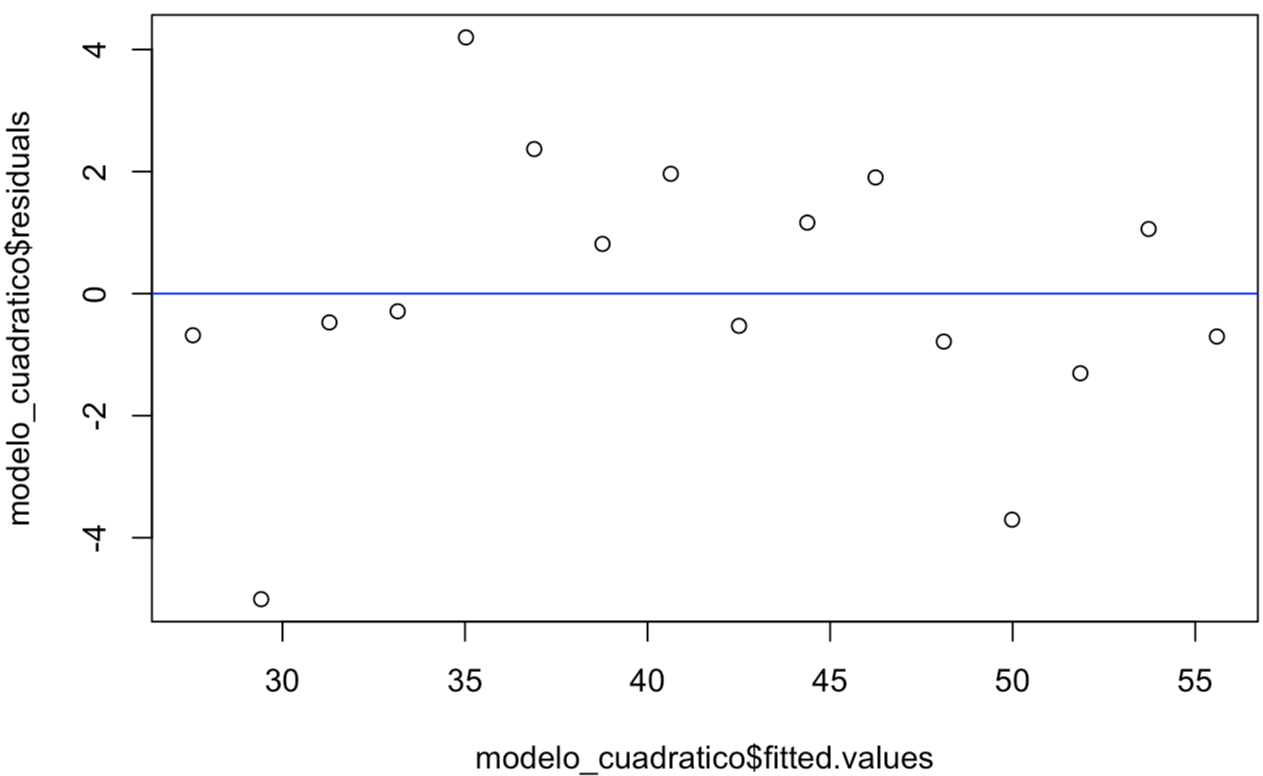
```
t.test(modelo_cuadratico$residuals)
```

```
##
##  One Sample t-test
##
## data:  modelo_cuadratico$residuals
## t = -4.9155e-17, df = 15, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -1.203541  1.203541
## sample estimates:
##     mean of x
## -2.775558e-17
```

De acuerdo con la prueba de hipótesis realizada, no se tiene evidencia para rechazar la hipótesis nula: la media de los residuos es = 0.

Homocedasticidad e independencia

```
plot(modelo_cuadratico$fitted.values, modelo_cuadratico$residuals)
abline(h=0, col= 'blue')
```



No parece haber ningún patrón, la varianza es constante

Prueba de independencia

H_0: Los errores no están autocorrelacionados.

H_1: Los errores están autocorrelacionados.

```
library(lmtest)
dwtest(modelo_cuadratico)
```

```
##
##  Durbin-Watson test
##
## data:  modelo_cuadratico
## DW = 1.3899, p-value = 0.05476
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(modelo_cuadratico)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 1
##
## data:  modelo_cuadratico
## LM test = 1.4403, df = 1, p-value = 0.2301
```

En este caso, ambas prueban obtienen un valor mayor a alpha, por lo que no se rechaza la hipótesis nula, en este caso los errores no están correlacionados.

MSE y MAPE

Se predicen los valores usando el modelo y se calcula el error cuadrático medio y el error porcentual absoluto medio.

```
MSE = mean((ventas2 - modelo_cuadratico$fitted.values)^2)
cat('Error cuadratico medio:', MSE)
```

```
## Error cuadratico medio: 114.902
```

```
MAPE = mean(abs((ventas2 - modelo_cuadratico$fitted.values)) / ventas2) * 100
cat('\nError porcentual absoluto medio: ', MAPE,'%', sep = '')
```

```
##
## Error porcentual absoluto medio: 26.38411%
```

Conclusión

El mejor modelo es el modelo cuadrático, puesto que es el único que cumple con todos los supuestos de la regresión lineal, este modelo cuenta con un R2 de 0.94, lo cual significa que logra explicar casi toda la variación en la variable de interés.

La ecuación resultante del modelo es:

$$\text{ventas}^2 = 25.6778 + 0.1268 \times \text{tiempo}$$

Pronóstico para el siguiente año

```
# Crear el rango extendido de tiempo
tiempo <- 1:16
tiempo2 <- 1:20
datos <- data.frame(x3 = tiempo2)

# Realizar predicciones para los próximos 4 trimestres
y_pred <- predict(modelo_cuadratico, newdata = datos)
y_pred <- sqrt(y_pred)

# Crear la gráfica
plot(tiempo, ventas, type = "b", col = "blue", pch = 16,
     xlab = "Tiempo", ylab = "Ventas", main = "Datos Originales y Predicción del Modelo
Cuadrático",
     xlim = c(1, 20), ylim = range(c(ventas, y_pred)))

# Agregar las predicciones al gráfico
lines(tiempo2, y_pred, col = "red", lwd = 2)
points(tiempo2, y_pred, col = "red", pch = 17)

# Agregar leyenda
legend("topleft", legend = c("Datos Originales", "Predicción del Modelo Cuadrático"),
       col = c("blue", "red"), pch = c(16, 17), lwd = 2)
```