# Inteligencia artificial avanzada para la ciencia de datos II(Módulo 5: Estadística Avanzada para ciencia de datos)

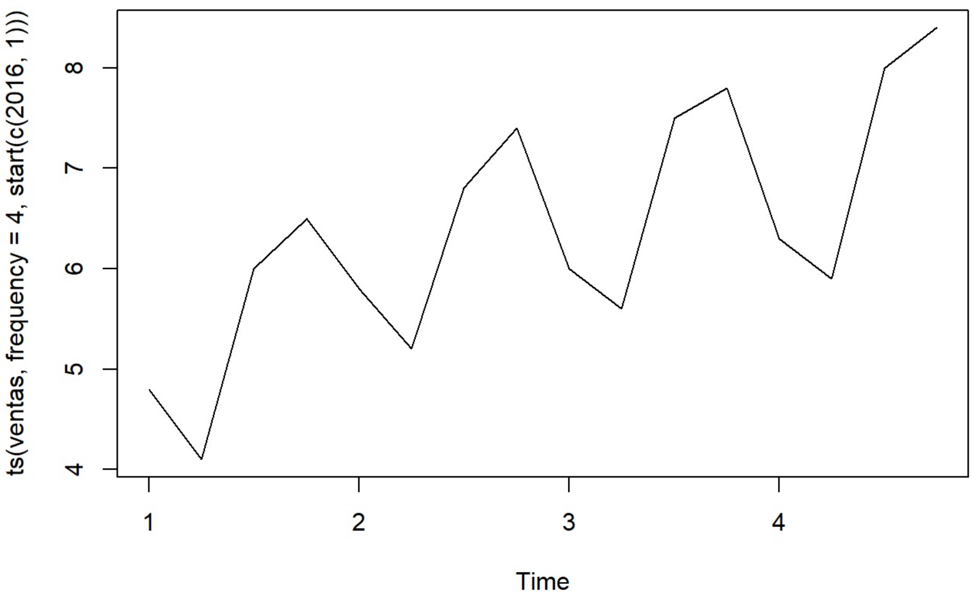
Roel De la Rosa - A01197595

21/11/2022

A través de este reporte se realizará un análisis de series no estacionarias. El objetivo es realizar un análisis de series de tiempo para una tienda de televisores, se busca predecir la cantidad de ventas que se harán para su siguiente año.

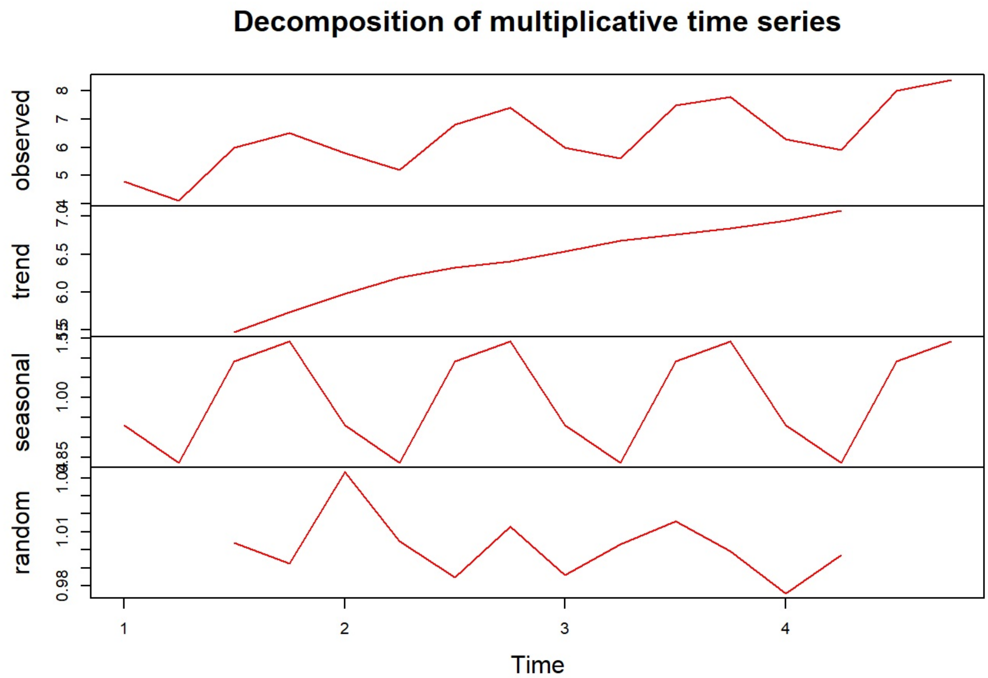
Para poder hacerlo primero se hará una descomposición de series temporales, lo que ocurre aquí es que se separa una serie de tiempo en tres componentes distintos. El primero de ellos es la tendencia que tiene la serie de tiempo, el segundo es el componente estacional el cual es un fenomeno ciclico. El tercero es el componente aleatorio, el cual es el factor necesario para que, sumado a los dos anteriores, se llegue a la observación real.

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




Realiza el análisis de tendencia y estacionalidad

```
## $x
##   Qtr1 Qtr2 Qtr3 Qtr4
## 1  4.8  4.1  6.0  6.5
## 2  5.8  5.2  6.8  7.4
## 3  6.0  5.6  7.5  7.8
## 4  6.3  5.9  8.0  8.4
##
## $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
##
## $trend
##     Qtr1   Qtr2   Qtr3   Qtr4
## 1     NA     NA 5.4750 5.7375
## 2 5.9750 6.1875 6.3250 6.4000
## 3 6.5375 6.6750 6.7625 6.8375
## 4 6.9375 7.0750     NA     NA
##
## $random
##        Qtr1      Qtr2      Qtr3      Qtr4
## 1        NA        NA 1.0039818 0.9925353
## 2 1.0430335 1.0048157 0.9849340 1.0129944
## 3 0.9861607 1.0030787 1.0160445 0.9994305
## 4 0.9757661 0.9970658        NA        NA
##
## $figure
## [1] 0.9306617 0.8363763 1.0915441 1.1414179
##
## $type
## [1] "multiplicative"
##
## attr(,"class")
## [1] "decomposed.ts"
```

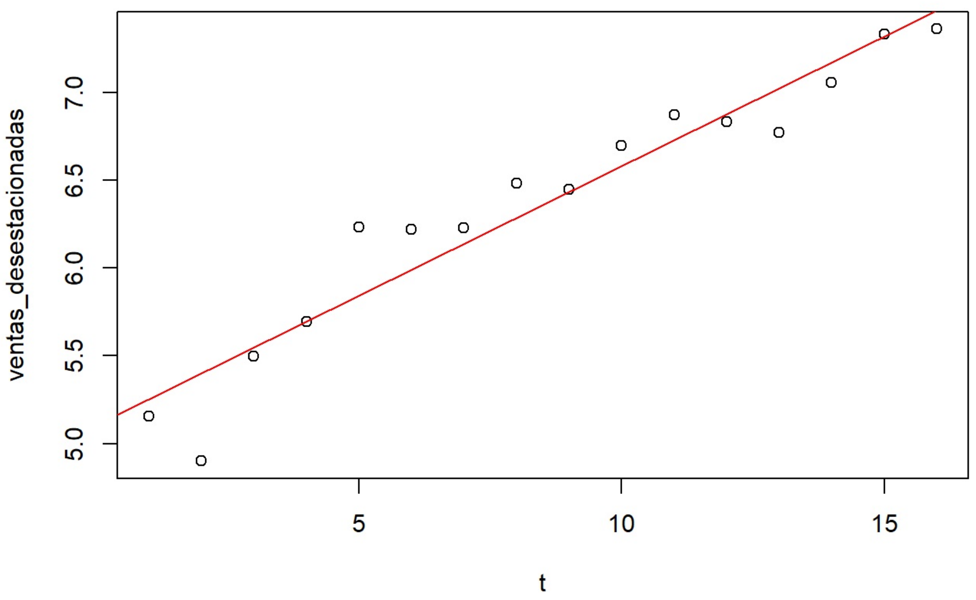



Aquí podemos ver la descomposición de la serie temporal. La primera gráfica, 'observed', muestra los resultados reales que se obtienen. El segundo, 'trend', muestra la línea de tendencía de los datos obtenidos. La tercera gráfica, 'seasonal', muestra las variaciones periodicas que se observan del comportamiento de los datos reales. Por último, la última gráfica muestra el comportamiento 'random', es decir, que no se puede predecir. Esto es importante pues nunca tendremos un modelo que puede predecir todo de forma perfecta.

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.

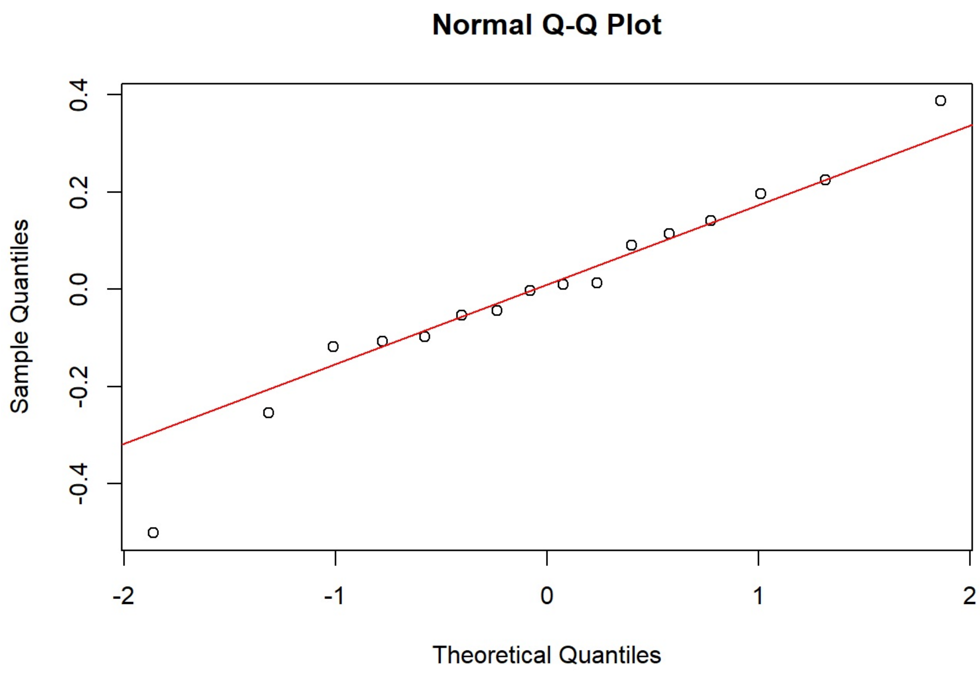



Podemos observar que las ventas desestacionalizadas parece que se comporte de manera lineal con respecto al tiempo. Si bien la línea no es perfecta, parece que se ajusta de manera más que adecuada con las ventas desestacionalzadas.

Analiza la pertinencia del modelo lineal:

```
##
## Call:
## lm(formula = ventas_desestacionadas ~ t)
##
## 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 ***
## t            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
```

A partir de análisis de la regresión lineal de las ventas desestacionalizadas podemos observar que el p-value es de $4.248x10^{-9}$, lo cual indica que el modelo es estadisticamente significativo, es decir, la regresión lineal representa la tendencia de los datos.




```
##
##  Shapiro-Wilk normality test
##
## data:  lr$residuals
## W = 0.96379, p-value = 0.7307
```

Podemos ver que parece ser que no se rechaza la prueba de normalidad de shapiro con respecto a los residuales de la regresion lineal. Esto es importante, pues para poder justificar que una regresión lineal se ajusta correctamente debemos de comprobar que los residuales de dicha regresión lineal se distribuyen bajo una distribución Gaussiana.

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

Posteriormente realziamos unas predicciones para una serie de tiempo y podemos ver a continuación los resultados obtenidos

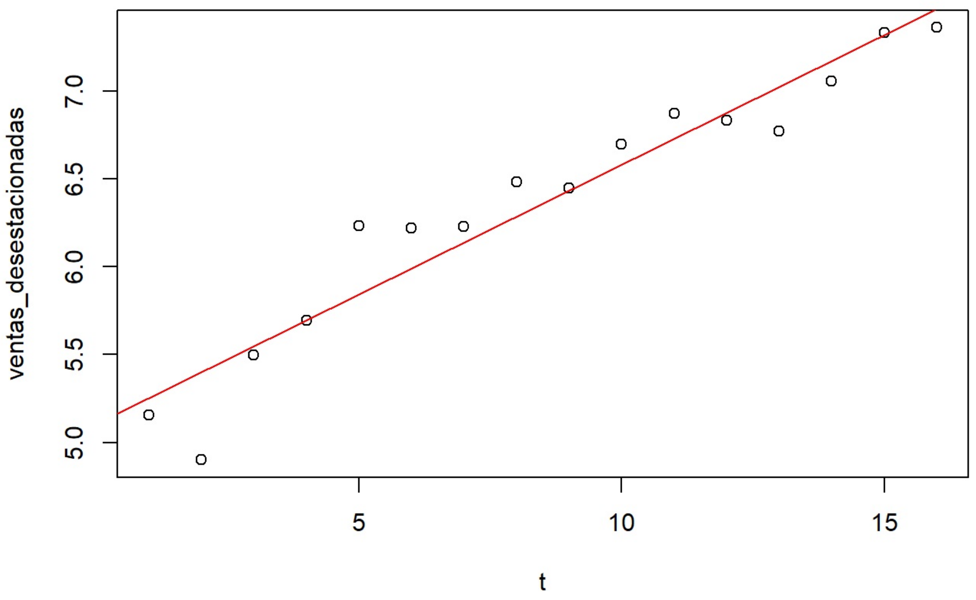
```
e <- NA
p <- NA
n <- 16
error <- NA
y = function(k) {5.108 + 0.147*k}

for(i in 1:(n-3)){p[i+3]=y(i); e[i+3] = p[i+3] - ventas[i+3]; error[i+3] = ((ventas[i+3]-p[i+3])/ventas[i+3])*100
}
CME2 <-mean(e^2,na.rm=TRUE)
message('El CME^2 es de: ', round(CME2,5))
```

```
## El CME^2 es de: 1.004
```

Vemos que el CME^2 obtenido es de 1.004, el cual es un valor relativamente muy bajo, lo que puede indicar que la estimación que hemos elaborado podrían ser consideradas como adecuadas.

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




En el gráfico anterior podemos observar que la línea recta se ajusta muy bien a los datos, lo cual nos indica nuevamente que los resultados obtenidos son bastante buenos.

Realiza el pronóstico para el siguiente año.

```
##   Trimestres predicciones
## 1          1     7.079544
## 2          2     6.485262
## 3          3     8.624290
## 4          4     9.186131
```

A partir de lo anteriormente mencionado se realiza las predicciones para los siguientes trimestres y podemos observar los resultados obtenidos. Obviamente no se pueden vender un número no entero de televisores, por lo que, en un caso pesisimista, redondeamos al entero menor más cercano.

```
##   Trimestres predicciones
## 1          1            7
## 2          2            6
## 3          3            8
## 4          4            9
```