

Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Monterrey



Inteligencia artificial avanzada para la ciencia de datos II (Gpo 502)

Módulo 5: Estadística Avanzada para ciencia de datos y nombre de la concentración.

Reporte final de "Venta de televisores"

Angel Corrales Sotelo

| A01562052

Diciembre 2022

Resumen	3
Introducción	3
Análisis de los resultados	3
Exploración de datos	3
Modelo lineal de la tendencia	4
Analiza la pertinencia del modelo lineal	5
Significancia de beta	6
Variabilidad explicada por el modelo	6
Análisis de los residuos	7
Normalidad	7
Media cero	8
Homocedasticidad	8
CME y el EPAM	9
Predicción con el modelo	9
Conclusión del modelo	9
Conclusión	9

Resumen

Se requiere analizar los datos en cuanto a las ventas de televisores de los últimos 5 años con el propósito de predecir mediante un análisis estadístico las ventas de los próximos años y/o trimestres.

En cuanto a los métodos estadísticos, se realizó un análisis de series de tiempo no estacionarias y de tendencia.

Se logró obtener un modelo de regresión lineal que predice dichas ventas.

Introducción

En un mundo tan grande en el que hay tantos datos es necesario saber analizarlos y utilizarlos correctamente. En todo tipo de compañías que se dediquen a vender productos es muy importante poder entender sus ventas y actuar en base a ello de manera que predecirlas es muy útil para saber cuánto y cuándo invertir y de qué manera hacerlo. Como en este ejemplo en venta de televisores.

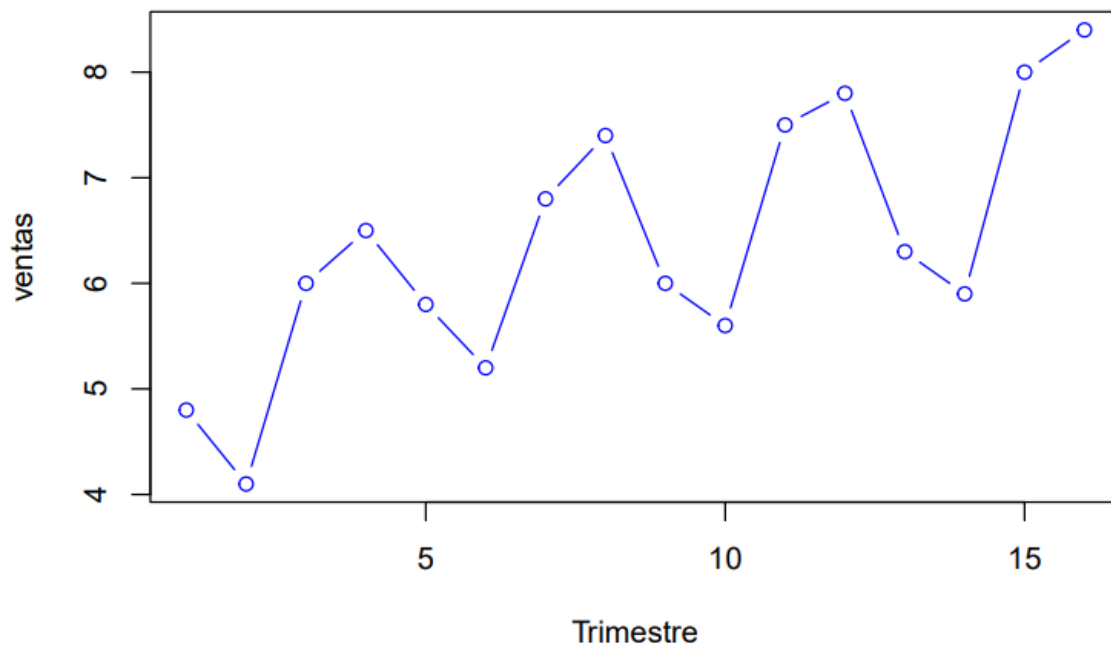
Es importante para que no haya pérdidas y no haya inventario de más ni menos.

Con los datos de los últimos 5 años se analizó la forma en que se comportan las ventas de televisores, resultando en que había una tendencia positiva y estacionalidad no estacionaria, de manera que se pudo encontrar un modelo para predecir las ventas de los próximos años de una manera aproximada.

Análisis de los resultados

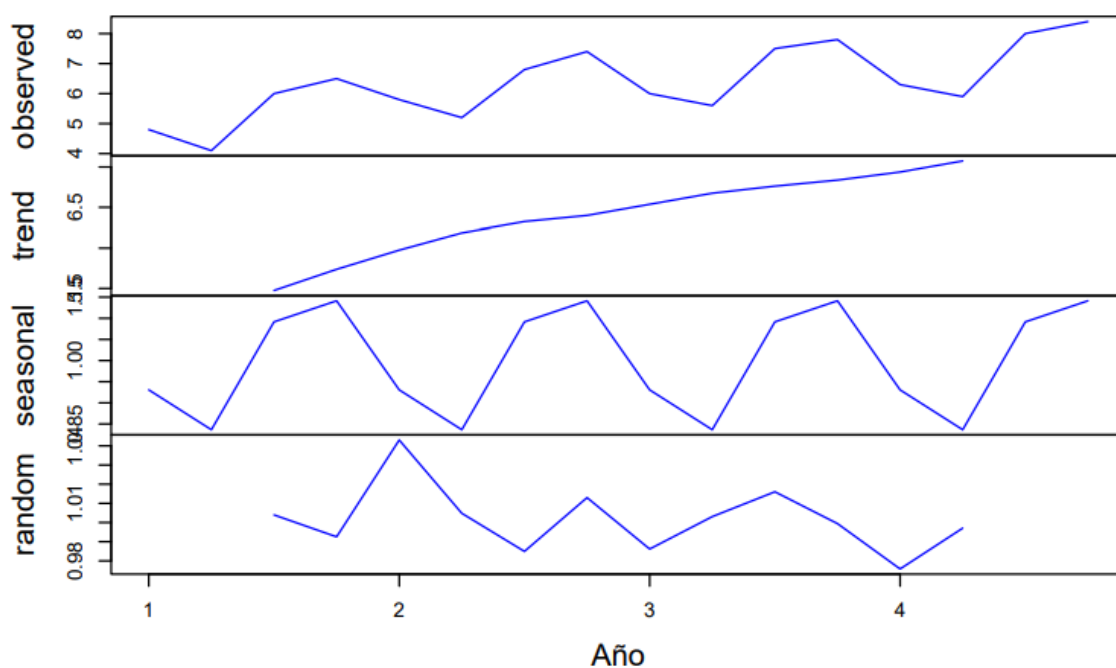
Exploración de datos

Primero se realizó una exploración de los datos en los cuales se logró ver a simple vista que presentaba estacionalidad entre los trimestres así como una aparente tendencia positiva.



Dicha afirmación queda más clara tras observar la gráfica de descomposición en la que la tendencia y estacionalidad se observan con más claridad.

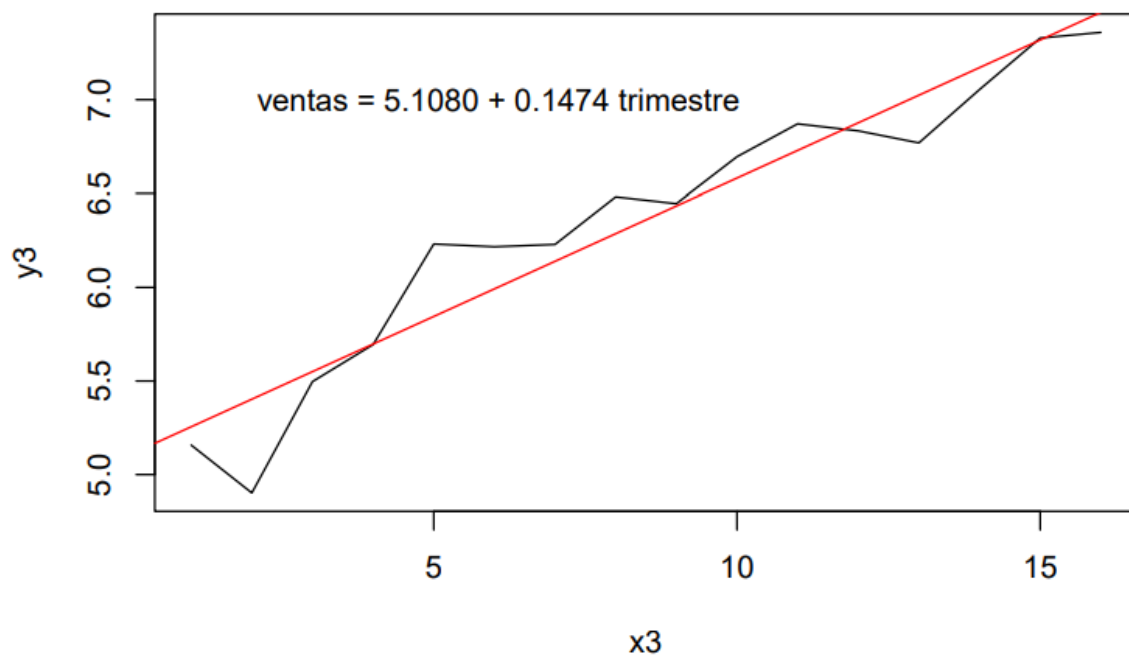
Decomposition of multiplicative time series



Modelo lineal de la tendencia

Se realizó un modelo de regresión lineal de la tendencia con la descomposición anterior de manera que logramos predecir los valores de los próximos años. Se puede observar como el

modelo predice de una forma muy aproximada las ventas que se hicieron a lo largo de los trimestres.



Analiza la pertinencia del modelo lineal

Aunque las gráficas son útiles para saber aproximadamente la calidad de un modelo, es importante realizar los análisis necesarios de este para determinar si realmente es un buen modelo en el que se puede confiar algo tan importante como las decisiones financieras de una empresa.

Para esto se comprobó la significancia de la variable en el modelo y su variabilidad explicada.

Significancia de beta

```

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

```

Hipótesis

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Regla de decisión

Se rechaza H_0 si valor $p < 0.05$

Conclusión de hipótesis

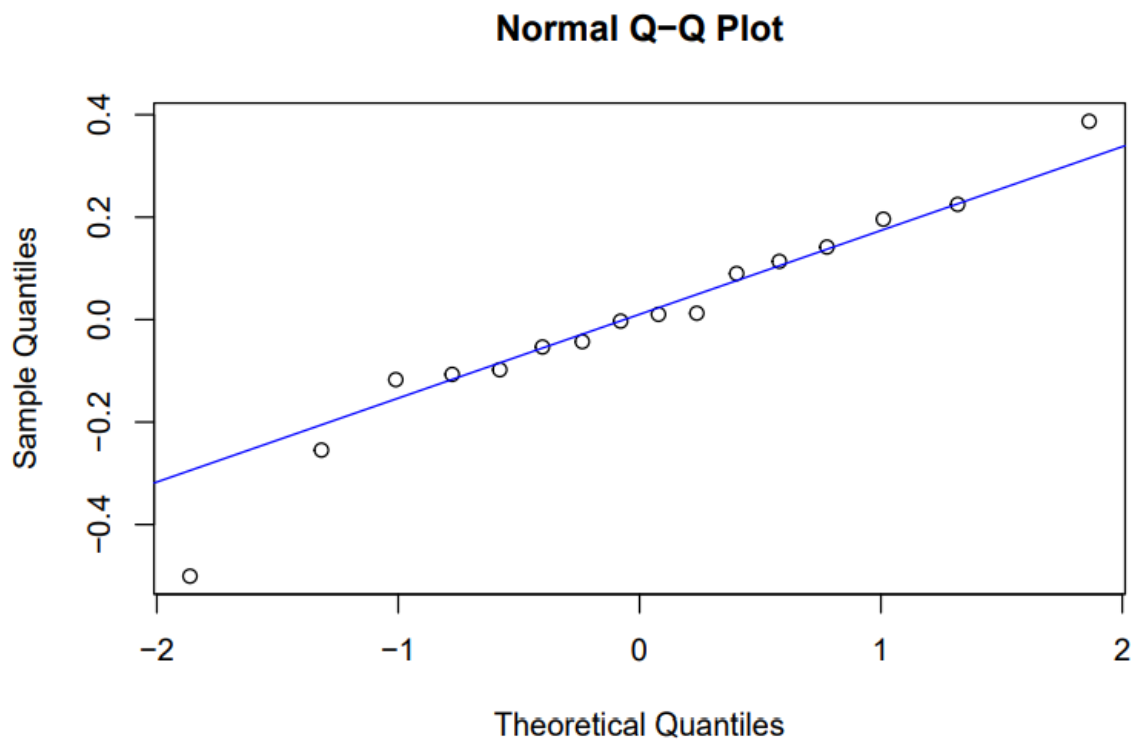
Ya que el valor p es menor a 0.05 , la hipótesis nula se rechaza, por lo tanto la β_1 es significativa.

Variabilidad explicada por el modelo

El coeficiente de determinación es de 0.9208 lo cual indica que hay una alta variabilidad explicada por el modelo lo cual lo hace más confiable.

Análisis de los residuos

Normalidad



La gráfica normal qqplot muestra que los residuos aparentemente se comportan de una manera normal, sin embargo es necesario confirmarlo con una prueba estadística, en este caso Shapiro-Wilk.

Shapiro-Wilk normality test

```
data: residuals  
W = 0.96379, p-value = 0.7307
```

Hipótesis

H0: Los residuos se comportan de manera normal

H1: Los residuos no se comportan de manera normal

Regla de decisión

Se rechaza H0 si valor $p < 0.05$

Conclusión

Ya que el valor p es mayor a 0.05 , la hipótesis nula no se rechaza, por lo tanto los residuos se comportan de manera normal

Media cero

One Sample t-test

```
data: residuals  
t = -2.9502e-17, df = 15, p-value = 1
```

Hipótesis

$H_0: \mu = 0$

$H_1: \mu \neq 0$

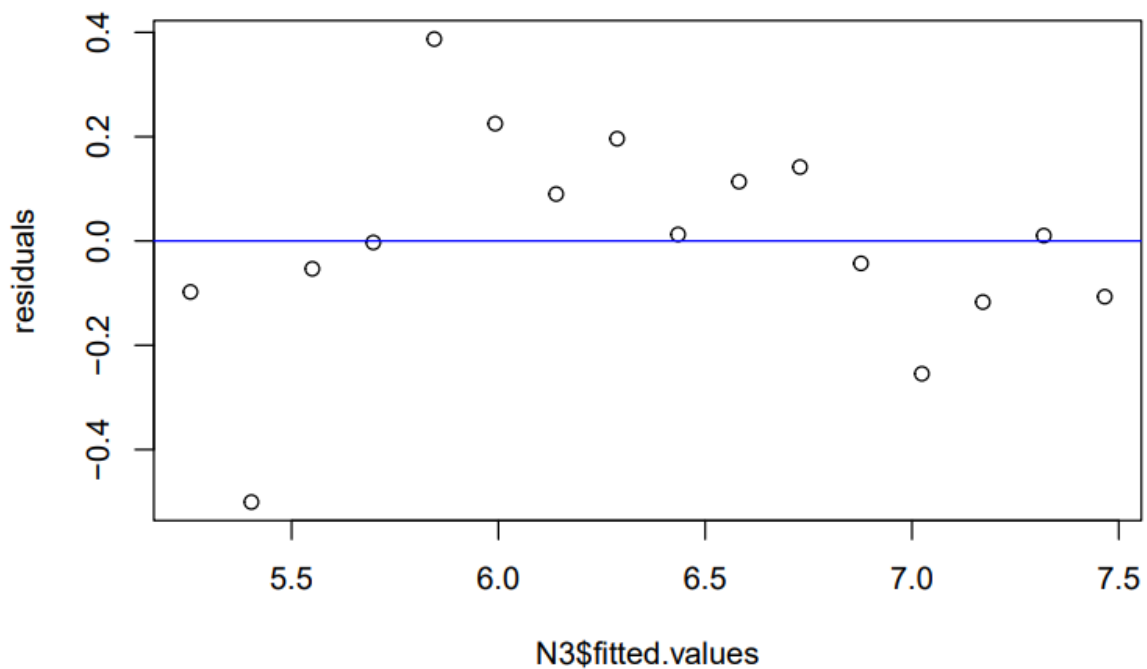
Regla de decisión

Se rechaza H_0 si valor $p < 0.05$

Conclusión

Ya que el valor p es 1, mayor a $\alpha = 0.05$, la hipótesis nula no se rechaza, por lo tanto la media de los residuos es 0

Homocedasticidad



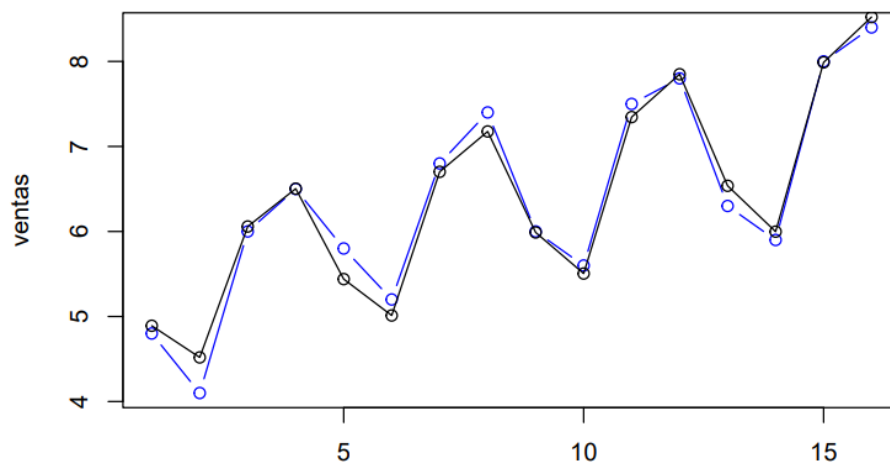
No se observa homocedasticidad lo cual podría indicar que hay dependencia entre los residuos.

CME y el EPAM

Para conocer el error de las predicciones de nuestro modelo se calcularon el CME y EPAM, de los cuales se obtuvo 0.033 y 0.024, respectivamente. Estos errores son muy bajos, lo cual es un buen indicador de que el modelo es confiable.

Predicción con el modelo

Finalmente se predijeron los valores con el modelo para compararlo con los datos reales.



Conclusión del modelo

Ya que el modelo pasó casi todas las pruebas de residuos y además la predicción de ventas se ajusta casi idealmente a las ventas reales se concluye que el modelo es una buena herramienta para realizar el pronóstico de ventas en este caso.

Conclusión

En cuanto a lo que esto significa para el problema en cuestión, podemos concluir que para el análisis de los datos y la construcción de un modelo es importante reconocer la manera en la que los datos se comportan y de esta manera reconocer la aproximación necesaria para su manejo. En este caso distinguir que había una tendencia, estacionalidad y que no era estacionaria permitió construir un modelo con muy poco error y un nivel alto de confiabilidad, de manera que las empresas podrían anticiparse al futuro y tomar decisiones de inventario dependiendo de sus intereses y el pronóstico realizado.

Anexos

<https://github.com/AngelCorso/Evidencia-TC3007C-502/tree/main/5.%20Estad%C3%ADstica%20Avanzada/Venta%20de%20televisores>