



Tecnológico de Monterrey

REPORTE FINAL “VENTA DE TELEVISORES”

Inteligencia artificial avanzada para la ciencia de datos

TC3007C.502

Módulo 5: Estadística Avanzada para Ciencia de Datos

Carlos David Toapanta Noroña A01657439

03 de diciembre del 2022

Resumen

Se tienen los datos de ventas de televisores (en miles) relativas a los últimos cuatro años (con los registros de cada trimestre). Lo que se busca es analizar la tendencia de las ventas mediante el uso de una regresión lineal de series de tiempo. Para llevar a cabo lo anteriormente planteado se recurrió al uso los siguientes métodos estadísticos: gráfico de dispersión, descomposición de la serie de tiempo en sus componentes, modelo lineal de la tendencia (regresión lineal, pertinencia del modelo, significancia de B1, variabilidad del modelo, análisis de residuos y prueba de normalidad), cálculo de CME/EPAM, y pronóstico del siguiente año. Después del análisis se puede concluir que hay un comportamiento estacional en las ventas y se obtiene un modelo que explica el 92% de la variabilidad de los datos, resultando ser idóneo para predicciones.

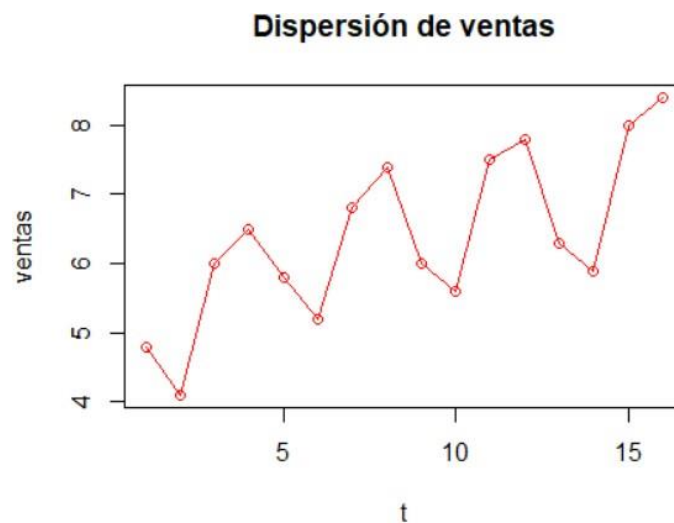
Introducción

Se tiene un set de datos que corresponden a los registros de ventas trimestrales de televisores en los últimos 4 años. Lo que se pretende es analizar la tendencia de la serie de tiempo que representan los datos y a partir de lo anterior responder a lo siguiente: ¿cómo es la tendencia y los ciclos?, ¿cuál es la interpretación de los componentes de la serie?, ¿cómo resulta ser la regresión lineal de la tendencia?, ¿cuál es la pertinencia del modelo creado?, ¿cuál es la significancia de B1?, ¿cómo es la variabilidad explicada por el modelo?, ¿qué dicen los residuos?, ¿cuáles son los resultados de la prueba de normalidad de los datos?, ¿Qué dicen los valores de CME y EPAM del modelo?, y ¿cuál es el pronóstico de las ventas del siguiente año? El problema planteado resulta importante de analizarse debido a que es una muestra de una gran cantidad de sucesos (ventas, cantidad de insumos, etc) presentes en distintos escenarios (finanzas, producción de alimentos, etc) donde los datos de los registros que se tienen siguen series de tiempo no estacionarias.

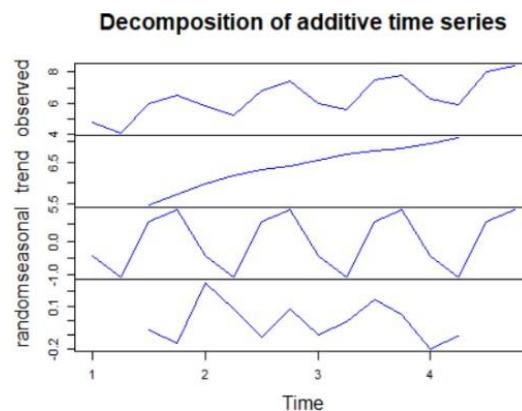
Análisis de resultados.

Dispersión de ventas

Al realizar el gráfico de dispersión podemos observar algunas fluctuaciones periódicas que se repiten de manera sistemática, lo cual nos puede indicar que la serie es estacional. Por ejemplo, en los trimestres 6, 10 y 14, las ventas tienden a bajar, es decir, en el segundo trimestre de cada año (abril-junio) las ventas de televisiones bajan.



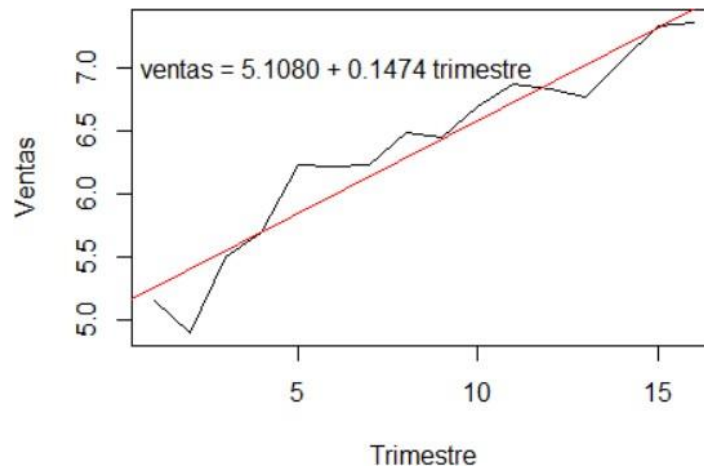
Tendencia y estacionalidad.



Para realizar el análisis de una serie de tiempo es necesario realizar observar sus componentes de manera aislada, puesto que nos permitirá encontrar tendencias, estacionariedad y/o aleatoriedad. En el gráfico 2 comprobamos que existe una tendencia, es decir, un movimiento a largo plazo (durante tres años), es ente caso, del aumento de las ventas de televisores.

Además, se confirma como se mencionó anteriormente, que la serie presenta estacionariedad al tener repeticiones periódicas de bajas ventas de televisores en el segundo trimestre de cada año.

Modelo lineal de la tendencia.



La línea de tendencia del gráfico lineal nos muestra que las ventas de televisores aumentan a un ritmo constante tras el paso de trimestres. El total de las ventas crece a un ritmo de 0.1474 por el número de trimestres, más una constante de 5.1080.

Pertinencia del modelo lineal

```

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

```

- Significancia de B1

Tomando en consideración la hipótesis:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Y la regla de decisión con un nivel de significancia del 5%:

Se rechaza H_0 si: el valor p es menor a $\alpha = 0.05$

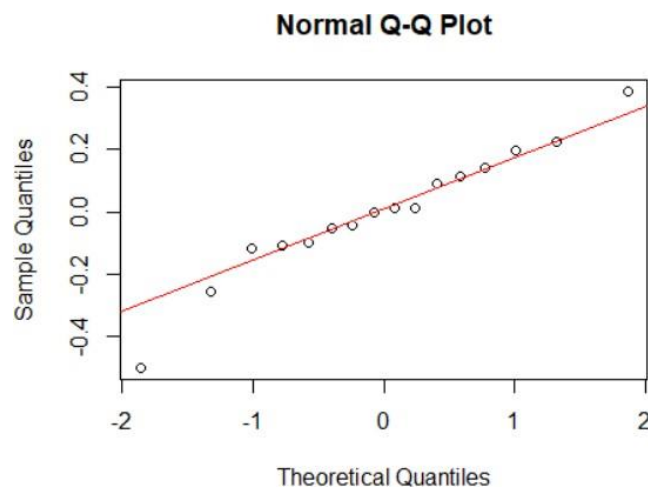
La conclusión es que ya que el p-value de $B1=4.25e-9 < 0.05$, menor a Alpha; hipótesis nula se rechaza y se concluye que B1 es estadísticamente significativa.

- Variabilidad explicada por el modelo.

El coeficiente de determinación o R^2 nos dice que el regresor B1 explica el 92.08% de la variabilidad de los datos.

- Análisis de residuos.

La gráfica 4 demuestra un comportamiento normal de los residuos. En el siguiente punto se comprobará con una prueba de normalidad.



- Prueba de normalidad.

Tomando en consideración la hipótesis:

$$H_0: \beta_i = 0 \text{ los datos representan una población normal}$$
$$H_1: \beta_i \neq 0 \text{ los datos no representan una población normal.}$$

Y la regla de decisión con un nivel de significancia del 5%:

Se rechaza H_0 si valor $p < \alpha$ donde $\alpha = 0.05$

La conclusión es que ya que $p\text{-value} = 0.7306 > 0.05$, mayor a Alpha, no se rechaza la hipótesis nula y se concluye que los datos representan una población normal.

CME y EPAM de la exactitud de predicción de serie de tiempo.

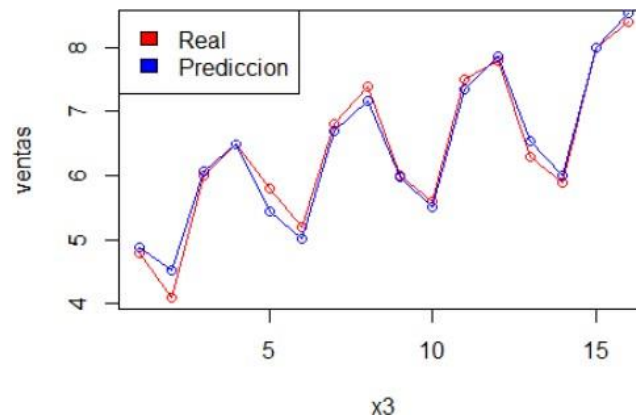
CME = 0.03970642

EPAM = 2.641619 %

El cuadrado medio debido al error (CME) o error cuadrático medio, es el promedio de los errores de pronóstico cuadrados, mientras que Error porcentual absoluto medio (EPAM) es una medida útil para comparar diferentes escalas. Ambos nos indican cuán exactos serán los pronósticos de nuestro modelo y así determinar si éste es capaz de reproducir los datos de las series de tiempo que están disponibles.

El CME se puede interpretar como la desviación estándar de la varianza inexplicada, y tiene la propiedad útil de estar en las mismas unidades que la variable de respuesta, por lo que, entre un valor más bajo, se indica un mejor ajuste. En nuestro caso el CME es bajo. Por su parte, el EPAM es un indicador del desempeño del pronóstico de las ventas de televisores, y en este caso nos dice que el error porcentual promedio del pronóstico es de 2.64%.

Ventas y predicciones vs tiempo



En la gráfica se percibe la superposición de la tendencia de las ventas reales (color rojo) a lo largo del tiempo en contraste con las ventas que resultan de las predicciones realizadas a partir del modelo lineal planteado (color azul); se puede apreciar como ambas presentan la misma tendencia y siguen una misma estacionalidad, a partir de lo cual es posible denotar la precisión del modelo.

Pronóstico del siguiente año

##	Trimestre	Pronostico
## 1	1	7.085872
## 2	2	6.491284
## 3	3	8.632585
## 4	4	9.195263

Utilizando el modelo lineal generado se obtiene el pronóstico de los siguientes cuatro trimestres de ventas (equivalentes al siguiente año). Se puede apreciar cómo entre los cambios de trimestre las ventas bajan y suben (siguiendo la estacionalidad de los datos utilizados para crear el modelo) y cómo los valores aumentan en comparación con los de registros pasados correspondientes a semestres anteriores (denotando la tendencia de las ventas). Además, se puede apreciar que el trimestre con ventas más bajas es el segundo, siguiendo la estacionariedad de los datos.

Conclusión

En conclusión, se analizó una situación donde el comportamiento de los datos referentes a los registros de ventas de televisores sigue una tendencia de serie de tiempo estacionaria. Se implementó un modelo lineal donde la variabilidad que explica es de 92%, betha es significativo y los residuos presentan normalidad, pudiendo concluir que se trata de un buen modelo, corroborado por la precisión de las predicciones que se llegan a realizar para el año siguiente.

Anexos

Código en R:

<https://drive.google.com/drive/folders/1dnXchXBpEhKbnm6AO7jTo2oNejHn2kF3?usp=sharing>