



Instituto Tecnológico de Estudios Superiores de Monterrey
Campus Monterrey

“Apegándome a la Integridad Académica de los Estudiantes del Tecnológico de Monterrey, me comprometo a que mi actuación en esta actividad esté regida por la integridad académica. En congruencia con el compromiso adquirido, realizaré este trabajo de forma honesta y personal, para reflejar, a través de él, mi conocimiento y aceptar, posteriormente, la evaluación obtenida”

Inteligencia artificial avanzada para la ciencia de datos II
Gpo 502

Módulo 5: Estadística
Series de tiempo no estacionarias

Alumno:
A01750164 | Paul Martín García Morfín

Profesor:
Blanca Rosa Ruiz Hernandez

Fecha: 30/11/2022

Venta de televisores

Resumen

En el presente trabajo se realiza el análisis de tendencia de una serie de tiempo, para el cual se utilizó una pequeña base de datos que incluye información sobre la venta trimestral de televisores a lo largo de cuatro años. El proceso realizado comienza con un primer suavizamiento haciendo uso de la técnica de promedios móviles centrados, luego se obtuvieron los índices estacionales por observación, los índices por periodo y, posteriormente, se obtuvo la predicción desestacionalizada. Al final, se construyó un modelo de regresión lineal que permitió describir la tendencia de los datos y se realizó una validación de dicho modelo. Se llegó a la conclusión de que el modelo construido tuvo una falla en las pruebas de homocedasticidad y, por tanto, no es completamente válido. Sería necesario hacer algunos ajustes o transformaciones futuras para mejorar dicho aspecto.

Introducción

Es posible modelar y explicar el comportamiento de una variable aleatoria que sufre cambios a través del tiempo gracias a métodos estadísticos como lo son las series de tiempo. El uso de este tipo de técnicas permite hacer estimaciones o predicciones sobre la forma en que se va a comportar la variable aleatoria analizada en un futuro. Por estas razones, las series de tiempo son de gran utilidad en la planeación y en áreas en las que evaluar el efecto de una variable en el futuro se requiere para disminuir el riesgo en la toma de decisiones o implementación de futuras políticas, como lo son la economía, el marketing, los estudios demográficos o de medio ambiente.

Para poder realizar estos pronósticos se toma en cuenta las observaciones pasadas, así como las variaciones que se puedan haber presentado, y se analiza si esta estructura se conserva en el futuro. Estas variaciones vienen dadas por eventos o fenómenos que pueden ser internos o externos, e influyen directamente en la calidad de las predicciones. Dependiendo de su naturaleza, estos eventos pueden ser o no controlables durante el proceso, por lo que puede ser necesaria la toma de decisiones para actuar sobre ellos.

Las series de tiempo se clasifican como estacionarias o no estacionarias. Las estacionarias son aquellas en las que la media y variabilidad se mantienen constantes a través del tiempo, es decir, tienen la propiedad de ser estables, aunque existan variaciones, estas oscilan alrededor de una media constante.

Por el otro lado, las series no estacionarias tienen una media y varianza que sí cambia con el tiempo, lo que provoca que se observe una tendencia, ya sea creciente o decreciente, y las oscilaciones no se quedan alrededor de un valor constante, sino que se van moviendo.

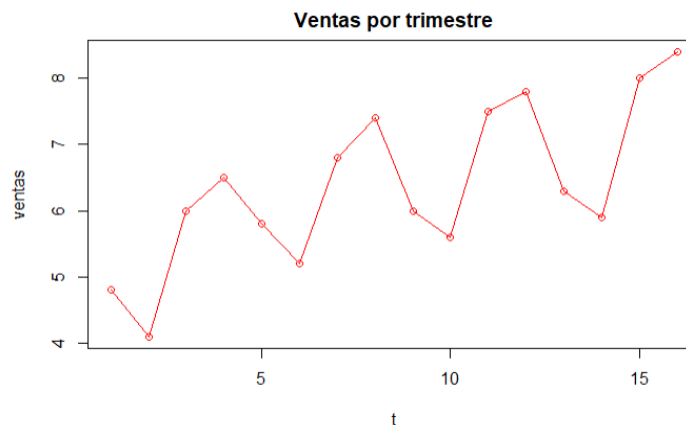
El problema que se analiza en este trabajo pertenece a las series de tiempo no estacionarias, puesto que se trata de información sobre la venta de televisores de varios trimestres a lo largo de cuatro años. Dicho conjunto de datos se presenta a continuación:

Año	1				2				3				4			
Trimestre	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Ventas (miles)	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

Lo que se busca es realizar un análisis de estacionalidad y tendencia para poder crear un modelo predictivo que permita conocer las ventas del quinto año, además de hacer una validación de dicho modelo y conocer qué tan bien se ajusta a nuestros datos a través de métricas como el cuadrado medio del error o el promedio de errores porcentuales.

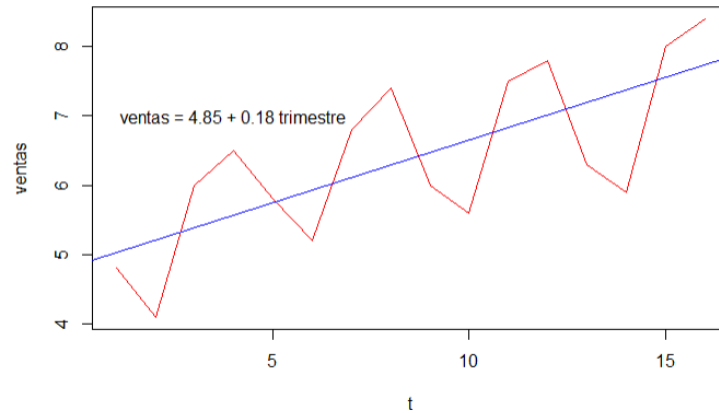
Procedimiento

1. Gráfico de dispersión. Tendencia y ciclos



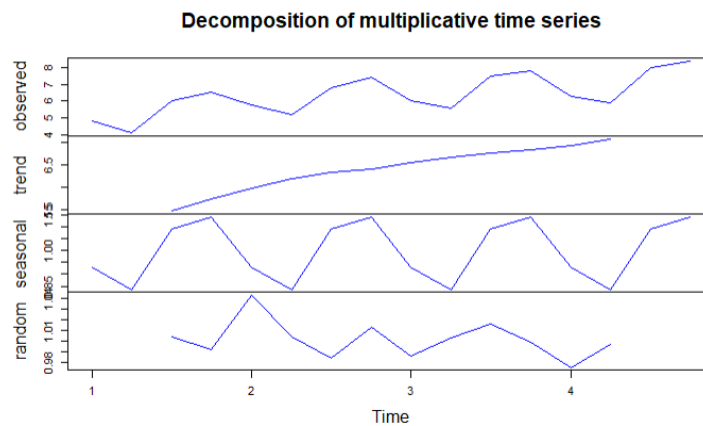
Graficando los datos proporcionados, se puede observar que corresponden a una serie de tiempo no estacionaria ya que se visualiza un claro crecimiento. Cuando se pone un negocio se espera que el negocio prospere y las ventas crezcan, es normal que haya fluctuaciones en el año pero que la tendencia es que crezca.

2. Análisis de tendencia y estacionalidad



Se puede trazar una línea recta y observar cómo la tendencia va creciendo, sin embargo, el problema con esta tendencia es que no está tomando en cuenta la estacionalidad, que son los ciclos a lo largo del año.

A. Descomposición de la serie en sus 3 componentes e interpretación



La primer gráfica son los valores originales observados, mientras que las otras tres son los componentes, los cuales se explican a continuación:

- Tendencia: La tendencia de la serie de tiempo se caracteriza por un patrón gradual y consistente de las variaciones de la propia serie. La tendencia a largo plazo se ajusta al esquema de moverse continuamente hacia arriba. Es la línea recta encontrada.
- Variación estacional: Este componente representa la variabilidad en los datos debido a las estaciones, la cual corresponde a los movimientos de la serie que recurringen año tras año en los mismos trimestres (para este caso) del año, con la misma intensidad.

- Variación irregular: Se debe a factores a corto plazo que afectan a la serie de tiempo y que son imprevisibles. Son los errores, es decir, la diferencia entre lo predicho y lo esperado.

3. Análisis del modelo lineal de la tendencia

A. Regresión lineal de la tendencia (ventas desestacionalizadas vs tiempo)

```
call:
lm(formula = y2 ~ x2)

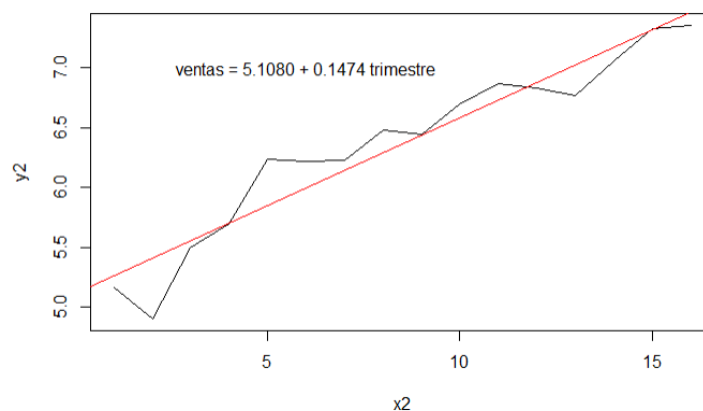
Coefficients:
(Intercept)          x2
      5.1080         0.1474
```

Las ventas desestacionalizadas son las ventas a las que ya se les quitó dos cosas:

- La irregularidad debido al azar, con los promedios móviles.
- La estacionalidad, con los índices estacionales (fluctuación debido a las estaciones del año).

Se aplica regresión lineal simple y se obtiene la recta, cuya es $y = 5.1080 + 0.1474x$

B. Dibujo de la recta junto con las ventas desestacionalizadas



Como se puede observar en la gráfica, se logró un suavizamiento de la serie diferente. Sobre esta se trabaja la regresión lineal para encontrar la tendencia. Es importante mencionar que se usó un esquema multiplicativo (donde la tendencia se va incrementando) para este suavizamiento.

C. Pertinencia del modelo lineal

- Significancia de β_1

```
call:
lm(formula = y2 ~ x2)

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 ***
x2           0.14738    0.01155   12.76 4.25e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

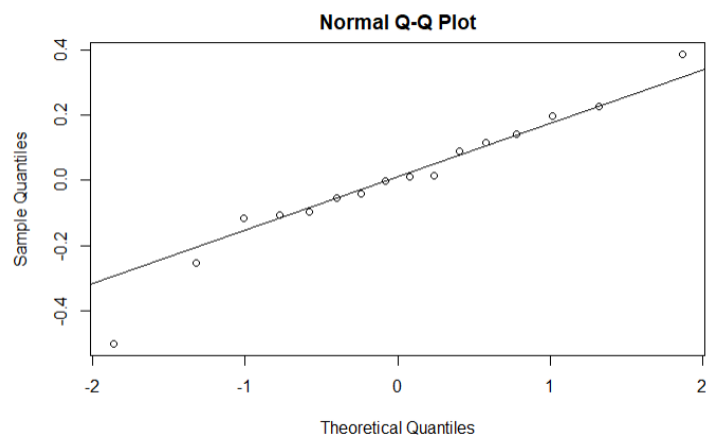
Residual standard error: 0.213 on 14 degrees of freedom
Multiple R-squared:  0.9208,    Adjusted R-squared:  0.9151
F-statistic: 162.7 on 1 and 14 DF,  p-value: 4.248e-09
```

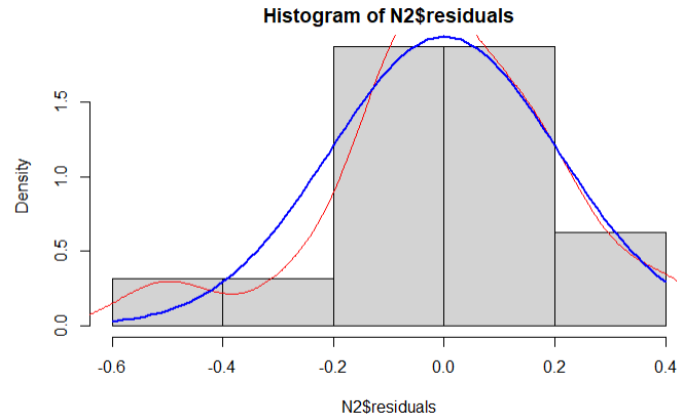
Para este caso, se rechaza la hipótesis nula ya que el valor p es menor a la alfa y se puede decir que la beta es significativamente diferente de 0.

- Variabilidad explicada por el modelo

Para este paso se obtuvo el coeficiente de determinación, el cual nos dice que el modelo explica el 92.08% de la variabilidad. Se deduce que aproximadamente el 8% de las ventas no se relaciona con el tiempo.

- Normalidad de los residuos





El valor p es mayor a alfa, así que se acepta la hipótesis nula y es posible decir que los datos provienen de una normal (los residuos se distribuyen como una normal).

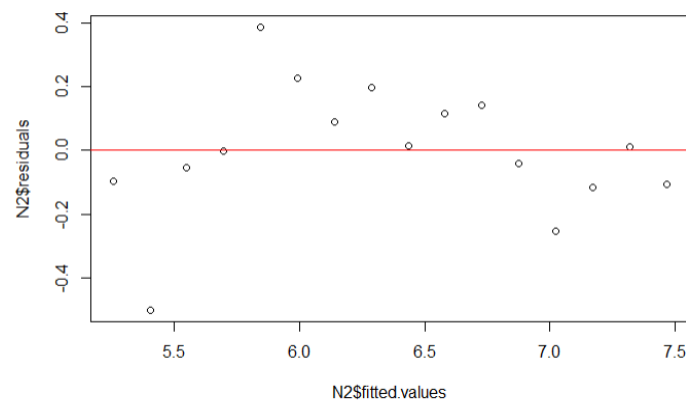
- Verificación de media cero

```
One Sample t-test

data:  N2$residuals
t = -1.4751e-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
-7.589415e-18
```

Se aplicó una prueba de hipótesis para medias usando la t de Student, la cual arrojó un valor p mayor a alfa, por lo cual se acepta la hipótesis nula y se dice que la media de los residuos no es significativamente diferente de cero.

- Homocedasticidad



De acuerdo a la gráfica, parece no existir homocedasticidad en los residuos, presenta asimetría, por lo que presenta heterocedasticidad y sesgo.

4. Cálculo del CME y del EPAM

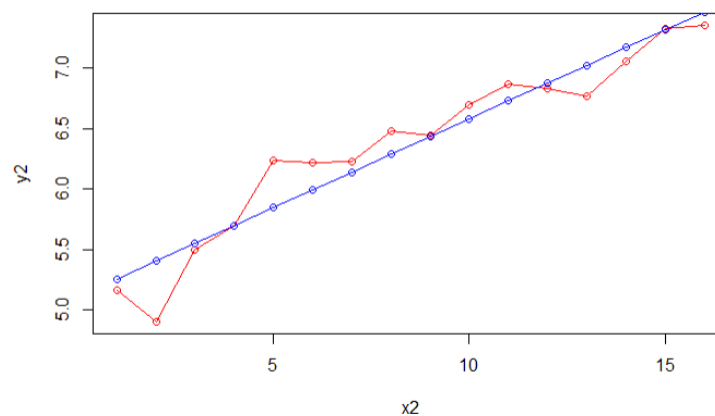
Se realizó el cálculo de estas dos métricas para evaluar las predicciones de la serie de tiempo y se obtuvo lo siguiente:

- CME = 0.03970642
- EPAM = 2.439396

El mejor modelo es aquel en donde se obtenga menos cuadrado medio del error.

5. Gráfico de los valores de las ventas y las predicciones vs el tiempo

El modelo resultante se observa a continuación:



6. Pronóstico para el siguiente año

Los resultados de las predicciones de cada trimestre para el siguiente año son:

- [17] 7085.872
- [18] 6491.284
- [19] 8632.585
- [20] 9195.263

Para este cálculo se estableció la función lineal obtenida anteriormente y se sacaron los índices estacionales de la matriz al hacer la descomposición (para así tomar en cuenta la estacionalidad), se multiplica por la función evaluada en el tiempo de cada trimestre y, para obtener las unidades correctas se multiplica por 1000.

Conclusión

Gracias a las métricas utilizadas y a los análisis realizados se podría concluir que es un buen modelo, ya que se obtuvo un CME pequeño y el coeficiente de determinación indica un alto porcentaje de explicación, además de que en la mayoría de pruebas de validación de residuos se cumplió con los supuestos, sin embargo, en los tests de homocedasticidad no salió como lo esperado, es decir, presentó sesgo y heterocedasticidad en los residuos. Esto se puede deber a que el comportamiento no es del todo lineal, o hace falta considerar otro factor para explicar la variabilidad completa. Sería necesario hacer algunos ajustes o transformaciones para mejorar en ese sentido.

Por otro lado, se remarca la importancia de estas técnicas para el análisis de variables que varían con el tiempo, especialmente cuando se trata de casos parecidos al presentado en este trabajo como lo son el análisis de ventas o problemas de economía en general, ya que pueden ayudar al manejo de riesgos y toma de decisiones. Sin embargo, también es importante validar el modelo y recordar que lo que se obtendrá son sólo pronósticos que pueden no cumplirse debido a factores externos.

Referencias

EcuRed. (2022). Series Temporales.

https://www.ecured.cu/Discusi%C3%B3n:Series_Temporales

Hernández, C; Pedraza L; Escobar A. (2008, junio). APLICACIONES DE LAS SERIES DE TIEMPO EN MODELOS DE TRÁFICO PARA UNA RED DE DATOS. Universidad Tecnológica de Pereira. [PDF]

Anexos

Liga a los documentos de análisis:

https://drive.google.com/drive/folders/1FrhDej6-UmzN_YvVlbT-7nlkSORvnMmw?usp=sharing