

Reporte final de “Venta de televisores”

Inteligencia artificial avanzada para la ciencia de datos II (Gpo 501)

Módulo 5: Estadística Avanzada

Luis Daniel Cano Irigoyen, A00827178

2022-12-02

Resumen

Problemática

El objetivo del problema es analizar la venta de televisores en un periodo de 4 años, divididos por trimestres, descubriendo tendencias no estacionarias de series de tiempo. Se busca comprender su comportamiento y crear un modelo capaz de realizar un pronóstico de las ventas al futuro.

Métodos y técnicas estadísticas

- Regresión lineal
- Descomposición de serie de tiempo a estacional
- Tendencias
- Residuos y Normalidad
- Pronósticos

Principales resultados

Se logró desestacionalizar la serie y por medio de una regresión lineal crear un modelo que explicara el 92.08% de la variabilidad de las ventas.

Introducción

Problema a resolver

En esta problemática contamos con una serie de tiempo con datos que tienen un comportamiento no estacionario. Los valores suben y bajan de manera periódica siguiendo patrones similares en distintos tiempos, y este patrón parece ir en aumento cada año.

Importancia del problema

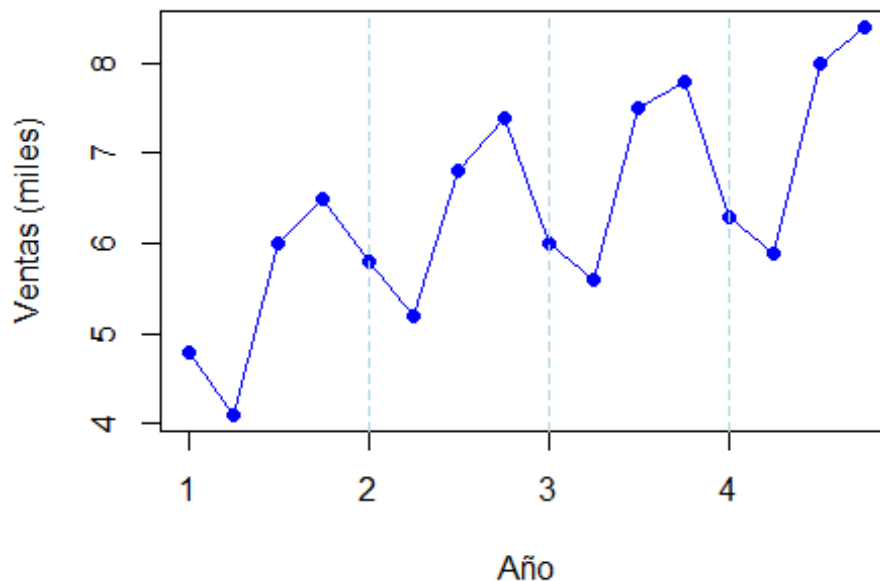
Para una empresa o individuo con interés de vender televisores, conocer la formula que explica estos datos le podría ser de suma importancia al tomar decisiones empresariales, como al hacer estimaciones de ingresos o al comprar un stock de televisores. Estar bien informados con decisiones respaldadas en datos reales podrá incrementar el potencial del negocio.

Análisis de los resultados

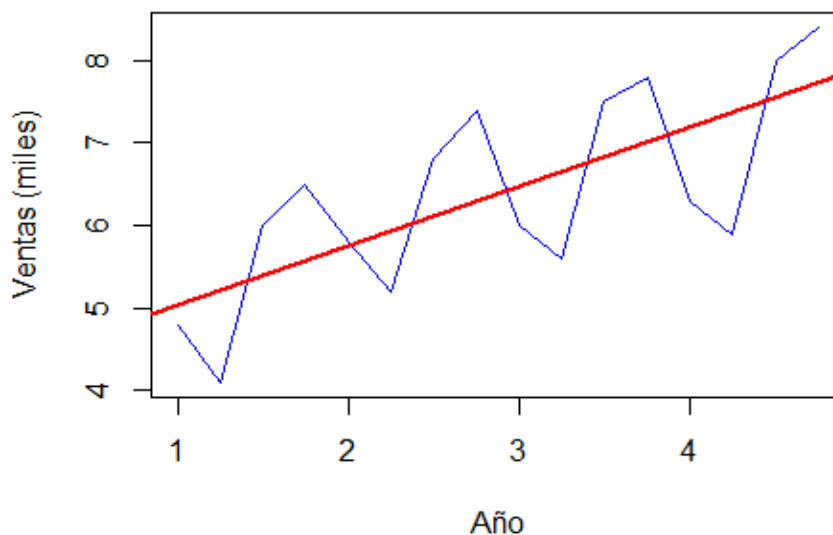
Datos de las ventas de televisores

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

Gráfico de dispersión. Análisis de tendencia, ciclos y estacionalidad.



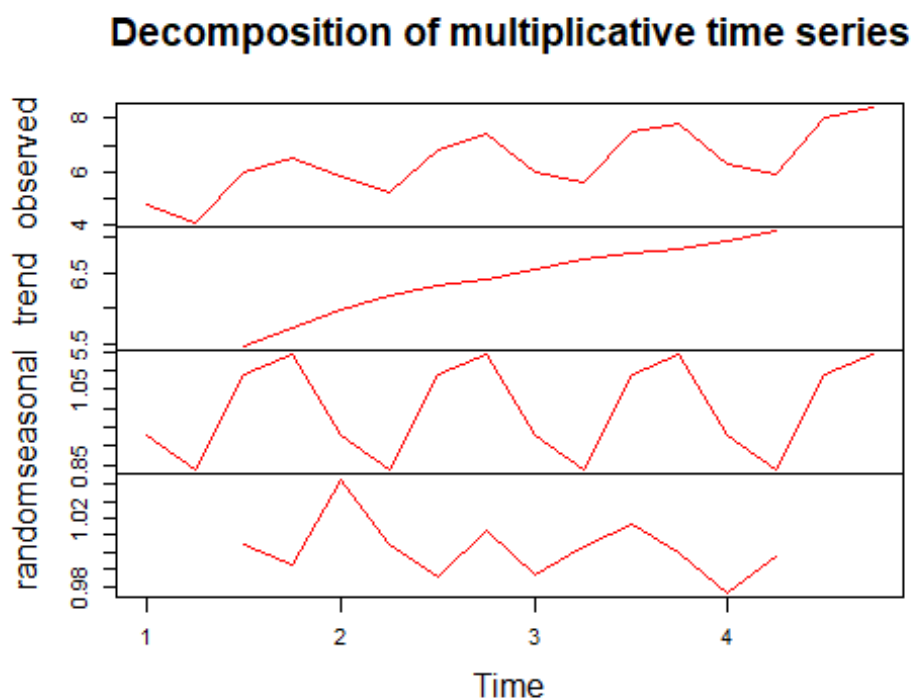
Al realizar el gráfico de dispersión sobre las ventas a lo largo de 16 trimestres, podemos notar un claro patrón anual (dividido por línea punteada) en donde las ventas empiezan en un punto medio-bajo, luego bajan a su punto menor, suben en el tercer trimestre, y llegan a su punto máximo en el cuarto, y este ciclo se repite anualmente, pero con una tendencia a aumentar entre años.



Si realizamos una regresión lineal sobre los mismos datos (línea roja) podemos ver claramente como las ventas, a pesar de su patrón no estacionario, tienden a aumentar con el tiempo. Esta regresión no cuenta con estacionalidad, por lo cual sólo nos es útil para analizar los datos, pero no para predecir los valores de ventas del pronóstico.

Descompongamos la serie en sus componentes

Al realizar la descomposición para un esquema multiplicativo obtenemos los siguientes resultados:



Observed son los datos originales de la tabla, mismos que mostramos en el gráfico de dispersión.

Trend representa la tendencia de los datos con el tiempo. Se comporta similar a nuestra regresión lineal sin estacionalidad. Se muestra un constante aumento a lo largo de la tendencia, solo con ligeros cambios en la velocidad que aumenta.

Seasonal muestra el patrón que siguen los datos cada estación o temporada. Podemos ver como se repite cada 4 trimestres, que bien es cada año. Se esperaría una repetición de patrón anual normalmente en la venta de cualquier producto, ya que los comportamientos de compra de la sociedad siguen ciertos patrones temporales anuales rondando fechas festivas, vacaciones, eventos de recurrencia anual como el aguinaldo y el Buen Fin, etc. Esta gráfica tiene forma similar a la dispersión de los datos original, pero no muestra el incremento de los valores anual de la primera.

Random muestra la gráfica de los residuos de la serie de tiempo. Se puede ver un comportamiento más aleatorio en esta.

Análisis el modelo lineal de la tendencia:

Regresión lineal de la tendencia, de ventas desestacionalizadas vs tiempo

Para ello dividimos de los componentes de nuestra serie descompuesta la X original y las estaciones (\$x / \$seasonal) creando las ventas desestacionalizadas, y de estas hacemos la regresión lineal contra el tiempo de 16 trimestres,

```
lm(ventas desestacionalizadas ~ tiempo)
```

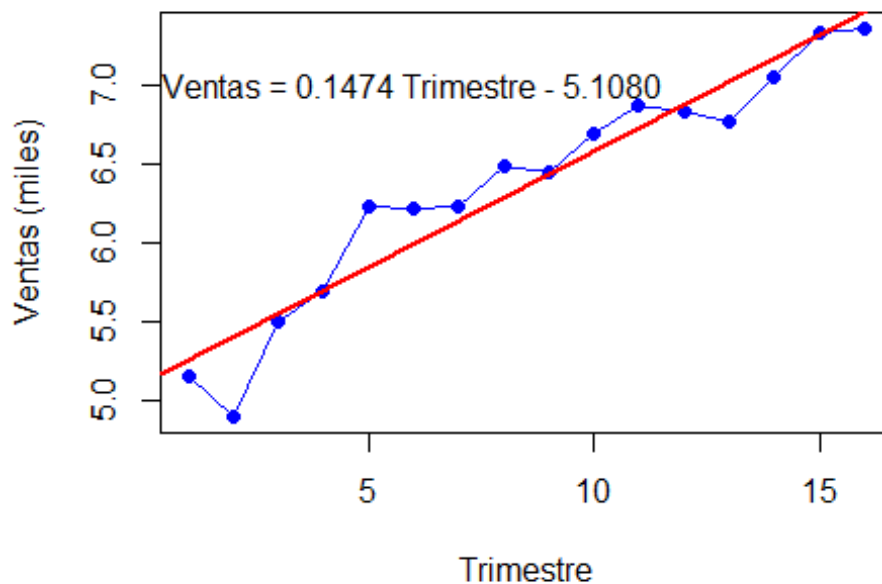
Coefficients:

(Intercept)	x1
5.1080	0.1474

obteniendo como resultado la fórmula:

$$\text{Ventas} = 0.1474 \text{ Trimestre} - 5.1080$$

Recta junto con las ventas desestacionalizadas.



Este conjunto no parece hacer mucho sentido, pero más adelante veremos cómo podemos utilizarlo para procrastinar los valores de ventas del siguiente año.

Pertinencia del modelo lineal: Significancia de β_1

Al realizar un summary de la regresión obtenemos los siguientes resultados:

```
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
```

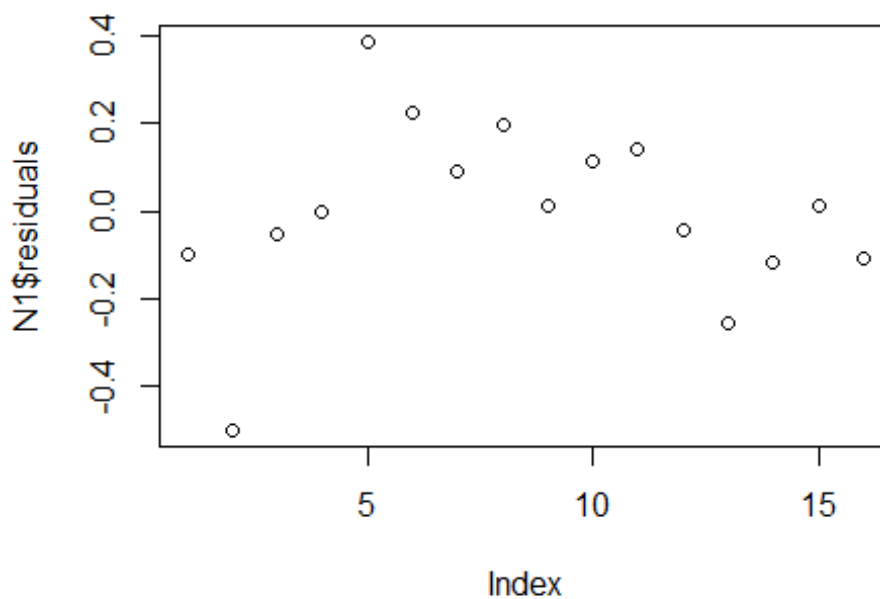
Aceptamos la hipótesis de que β_1 es significativo ya que contamos con un p -value es menor a 0.05 (p -value = 0.000000004248).

Variabilidad explicada por el modelo (c.d)

Multiple R-squared: 0.9208

De los mismos datos, podemos ver que el modelo explica una variabilidad del 92.07911% de las ventas.

Análisis de los residuos



Los residuos parecen ser aleatorios, no hay un patrón claro, no siguen linealidad y no siguen homocedasticidad.

Prueba de normalidad

Realizamos un Shapiro Test sobre los residuales del modelo y obtenemos:

Shapiro-Wilk normality test

```
data: N1$residuals
W = 0.96379, p-value = 0.7307
```

Como el p -value es mayor a 0.05, no podemos rechazar que los residuos sigan una distribución normal.

Cálculo del CME y el EPAM de la predicción de la serie de tiempo.

Realizamos este cálculo obteniendo la diferencia de las ventas y las ventas desestacionalizadas y obteniendo el promedio de los errores.

CME 0.6172467

EPAM 0.1194754

Contamos con un error de 0.6172 miles de pesos y un promedio de error porcentual de 11.95%

Pronóstico para el siguiente año.

Para realizar las predicciones calculamos los valores para Trimestre 17, 18, 19 y 20 en nuestro modelo de

$$Ventas = 0.1474 \text{ Trimestre} - 5.1080$$

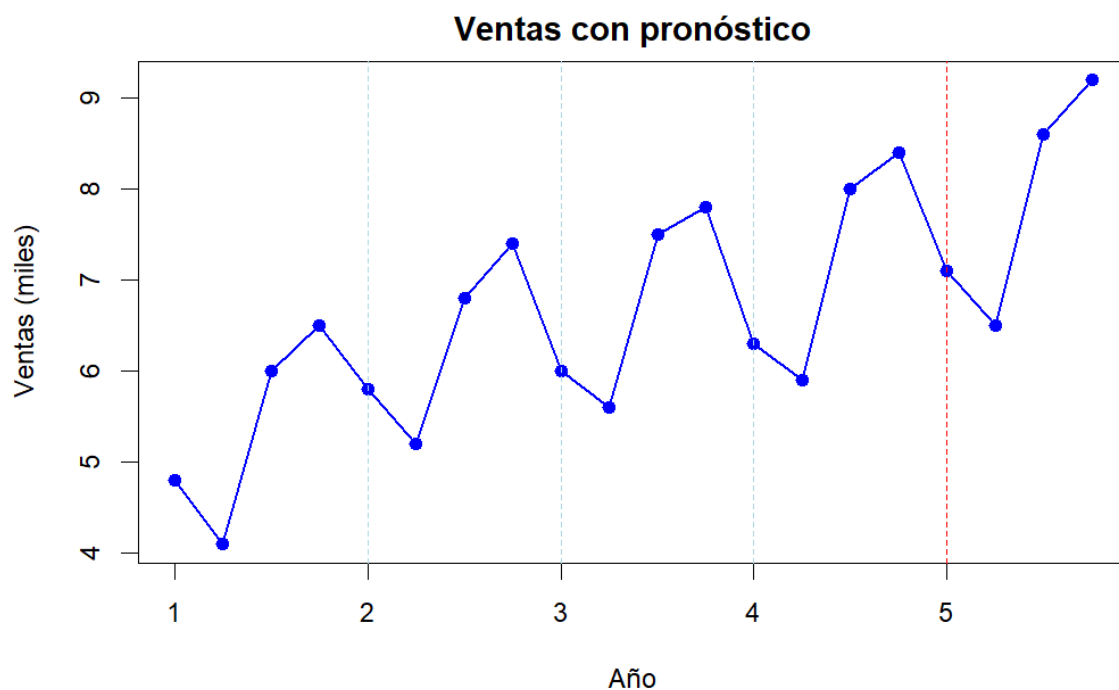
Obteniendo como resultado

7.613536 7.760918 7.9083 8.055682

A estos valores los multiplicamos por los 4 índices estacionales de nuestra serie descompuesta (\$seasonal[1] a \$seasonal[4]) y así obtenemos las predicciones finales de nuestro pronóstico:

Pronóstico para el siguiente año: 7.085626 6.491048 8.632258 9.194899

Gráfica de datos de ventas + pronóstico



Conclusiones

Al analizar los datos de la venta de televisores podemos ver como estos son una serie de tiempo no estacionaria, la cual siempre sigue un patrón anual con un aumento con el tiempo. Y al considerar estos elementos de la serie, al desestacionalizar los datos para crear una regresión lineal y al estacionar las predicciones de la misma, podemos obtener un pronóstico el cual, como vemos en la última gráfica de los resultados, si sigue la misma tendencia y patrón de los datos y si parece estar muy cerca de lo que serán los valores reales el siguiente año.

Siendo un buen modelo que explica el 92.08% de la variabilidad, este podría ser una importante y útil herramienta de análisis para las empresas y personas dentro del mercado de venta de teles, el cual les podrá ayudar a tomar decisiones más informadas, cumpliendo con las ofertas y demandas que más les convengan a lo largo del año.

Anexos

Link con proyecto en R:

<https://github.com/lcanoi/SeriesTiempoNoEstacionarias>