

Reporte final - Venta de televisores

Módulo 5: Estadística Avanzada para ciencia de datos Inteligencia artificial avanzada para la ciencia de datos

Daniel Salvador Cázares García A01197517

2022-12-04

Resumen

En este trabajo, por medio del análisis de datos de las ventas de televisores en los últimos años, se buscó entender el comportamiento de las ventas a lo largo de los años con el objetivo de poder estimar las ventas en los próximos trimestres e incluso años.

Para lo anterior, se emplearon métodos estadísticos como modelos lineales, regresión, tendencia y estacionalidad, los cuales permitieron explorar las ventas a lo largo de los años y realizar predicciones con respecto al tiempo.

Como resultado final, se realizó un modelo de regresión lineal para predecir las ventas a futuro.

Introducción

El análisis de datos cada vez cobra más importancia para las empresas y organizaciones. Un ejemplo muy común y de gran relevancia son las compañías que ofrecen productos y servicios, pues esto les permite analizar sus ventas y el comportamiento de los consumidores para mejorar sus ganancias.

A continuación, se utilizarán diferentes métodos estadísticos para analizar los datos de ventas de televisores en los últimos 4 años y tratar de predecir ¿Cuáles son las ventas estimadas a futuro?

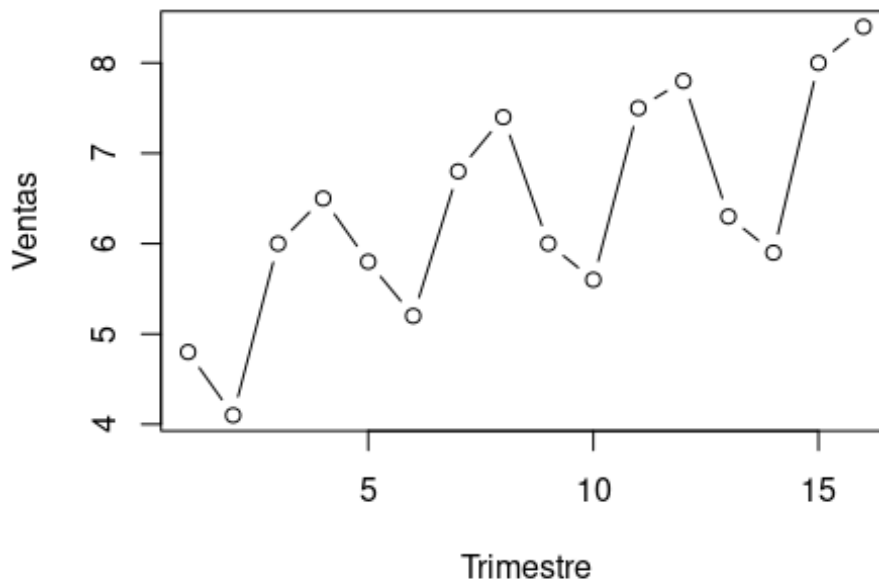
Datos

Año 1	2				3				4							
Trimestre	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Ventas (miles de \$)	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

Análisis de los resultados

Gráfico de dispersión para observar tendencia y ciclos

Inicialmente, se realiza un gráfico de dispersión para poder explorar de forma general los datos. Se puede observar que existe una tendencia positiva que aumenta con el paso de los años y además, existen ciclos en los que se presenta un patrón o estacionalidad.

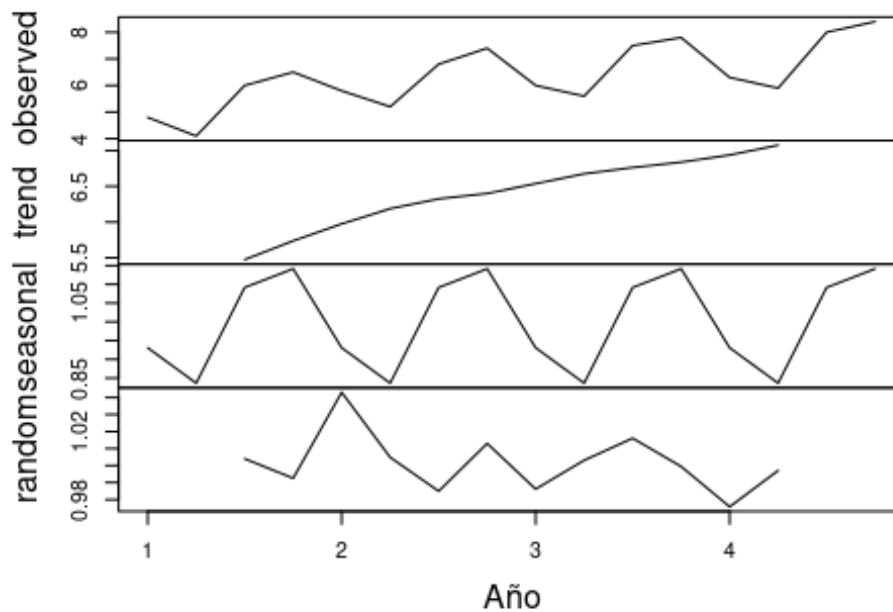


Análisis de tendencia y estacionalidad

Descomposición de la serie de tiempo

Posteriormente, se realizó una descomposición de la serie de tiempo en componentes de forma gráfica para poder analizar su comportamiento. Se puede ver que existe una tendencia positiva, así como estacionalidad y un cierto grado de aleatoriedad.

Decomposition of multiplicative time series

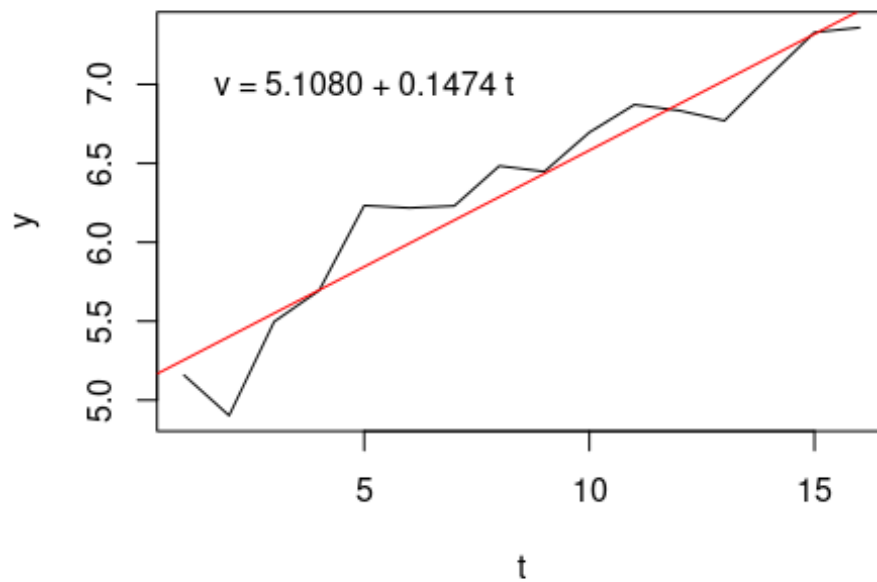


Análisis del modelo lineal de la tendencia

Regresión lineal

Después de analizar el comportamiento y la tendencia que han seguido las ventas en el pasado, ahora se busca realizar un modelo de regresión lineal que siga la tendencia de las ventas desestacionalizadas y que permita pronosticar las ventas de televisores con respecto al tiempo para el futuro.

Se puede observar que la ecuación obtenida que modela la tendencia de dicha situación es $tendenciaventas = 5.1080 + 0.1474 * trimestre$. Si se gráfica la recta se puede ver que se ajusta de forma muy cercana a las ventas.



Pertinencia del modelo lineal

Aunque la gráfica aparentemente si se ajusta, también es importante realizar algunas pruebas para verificar si el modelo es realmente adecuado. Para esto, se analizaron la significancia, la variabilidad, los residuos y la normalidad. En todas las pruebas se considera una significancia o valor p de 0.05.

```
##
## Call:
## lm(formula = y ~ 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
```

Significancia de β_1

$$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$$

Se rechaza H_0 si $p < 0.05$

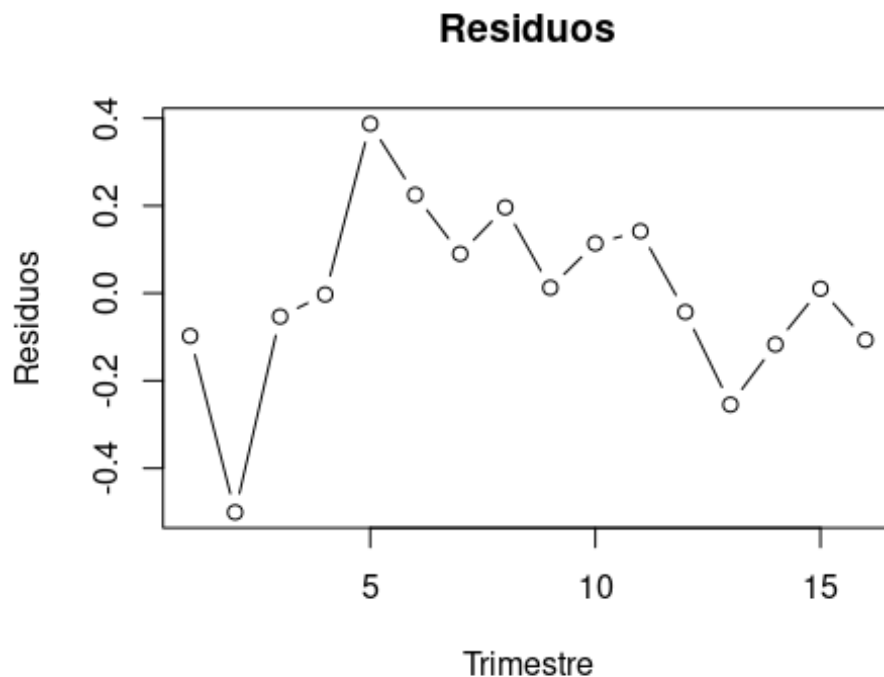
Se observa que la p obtenida es menor a 0.05, por lo que β_1 si es significativa.

Variabilidad explicada por el modelo

El coeficiente de determinación del modelo es de 0.9208, lo que significa que logra explicar el 92.08% de la variabilidad de los datos.

Análisis de los residuos y normalidad

Para analizar la distribución de los residuos, se puede realizar una graficación de los mismos. Se puede ver que los residuos no siguen alguna tendencia.



A simple vista se podría decir que los residuos tienen una distribución aleatoria. Sin embargo, además de graficar, es recomendable realizar pruebas estadísticas que permitan verificar y confirmar dicha hipótesis. Para esto, se puede utilizar la prueba de normalidad de Saphiro-Wilk, donde:

$$H_0: \text{Residuos normales} \quad H_1: \text{Residuos NO normales}$$

Se rechaza H_0 si $p < 0.05$

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

De este modo, se obtiene un valor p mayor que 0.05, por lo que no se rechaza la hipótesis nula y se puede confirmar que los residuos se distribuyen de forma normal.

Errores de la predicción de la serie de tiempo

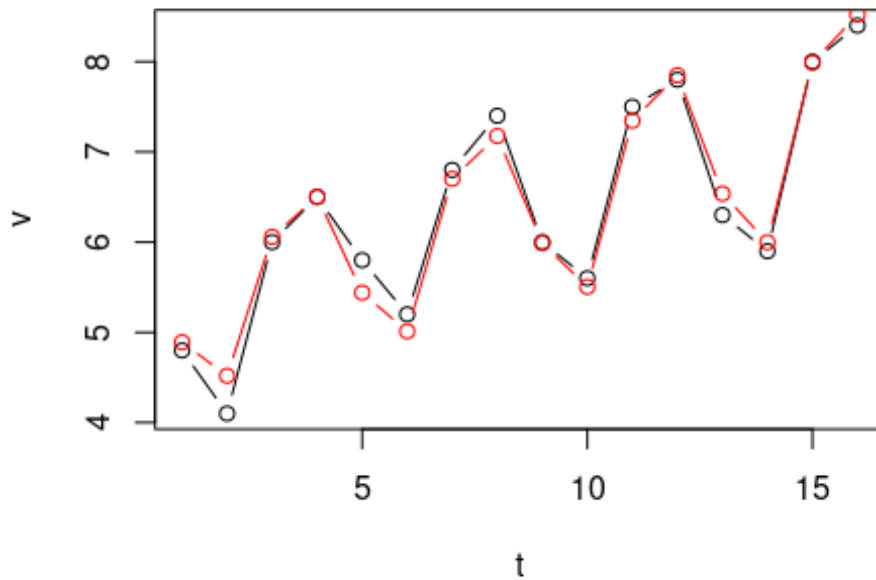
Además de verificar si el modelo es adecuado, también es importante observar la cantidad de error que puede existir en la predicción de la serie de tiempo. Para esto, se puede calcular el error Medio (CME) y el promedio de los errores porcentuales (EPAM).

```
## [1] 0.6957218  
## [1] 0.125894
```

Se puede observar, que el error medio porcentual no es tan elevado, lo cual indica que el modelo es relativamente bueno.

Predicciones vs el tiempo

Adicionalmente, podemos graficar las predicciones con respecto al tiempo para poder compararlas con los valores reales de las ventas. Se puede observar que las predicciones son muy cercanas a los valores reales. Además, estas también siguen el mismo patrón de estacionalidad, donde se reducen y aumentan las ventas cada 2 trimestres.



Pronóstico para el siguiente año

Finalmente, ya con el modelo realizado y verificado, se puede utilizar el modelo para obtener pronósticos (en miles de dólares) a futuro. En este caso, se realizará el cálculo para el siguiente año.

##	trimestre	ventas (miles)
## 1	17	7.085626
## 2	18	6.491048
## 3	19	8.632258
## 4	20	9.194899

Conclusiones del modelo

De forma general, se puede concluir que el modelo es adecuado para realizar pronósticos de las ventas, pues además de ajustarse muy cercanamente a los valores reales, también cumplió con las pruebas realizadas.

Conclusión

La utilización de herramientas estadísticas como la estacionalidad y los modelos de regresión pueden ser de gran utilidad para modelar situaciones de la vida real.

En este caso, se pudo realizar un modelo lineal para una serie no estacionaria que permite predecir las ventas de televisores. Al tratarse de un modelo sencillo con poco error, puede

utilizarse para hacer predicciones con cierto grado de certeza a futuro. Con este, se estiman ventas para el siguiente año que rondan entre 6.5 y 9.20 mil dólares por trimestre.

Cabe mencionar que, en las situaciones reales de ventas, por lo general existen otras variables que también influyen además de la estacionalidad y que el registro de datos se tiene con una mayor frecuencia (cada mes o semana), por lo que normalmente sería necesario un análisis e implementación más complejos.

Anexos

Repositorio de GitHub: <https://github.com/dscazares/Portafolio-TC3007C>