



Inteligencia Artificial Avanzada para Ciencia de Datos

Instituto Tecnológico de Estudios Superiores de Monterrey

Reporte final “Venta de televisores”

Rodrigo Montelongo Pinales, A00827757

Módulo 5: Estadística avanzada para ciencia de datos, Grupo 502

03 de diciembre del 2022

Resumen

Una empresa busca analizar y predecir la cantidad de televisores que venderán el quinto año. Para esto han recabado información de las ventas de televisores de los últimos 16 trimestres (4 años) y esperan analizar esta serie de tiempo.

Los datos de venta se almacenaron en un vector. Se utilizaron gráficos de dispersión para proporcionar apoyo visual. En cuanto a las técnicas de estadística, se realizó un análisis de tendencia y estacionalidad. Con los datos arrojados por el análisis se creó un modelo de regresión lineal y se validó el mismo.

Se encontró que el modelo de regresión explica en un 91% la variabilidad de los datos, tienen un P-value menor a 0.05, cumple con los supuestos del análisis de residuos, y cuenta con un CME muy bajo. Por lo tanto, se puede afirmar que predice con gran precisión las ventas de cada trimestre.

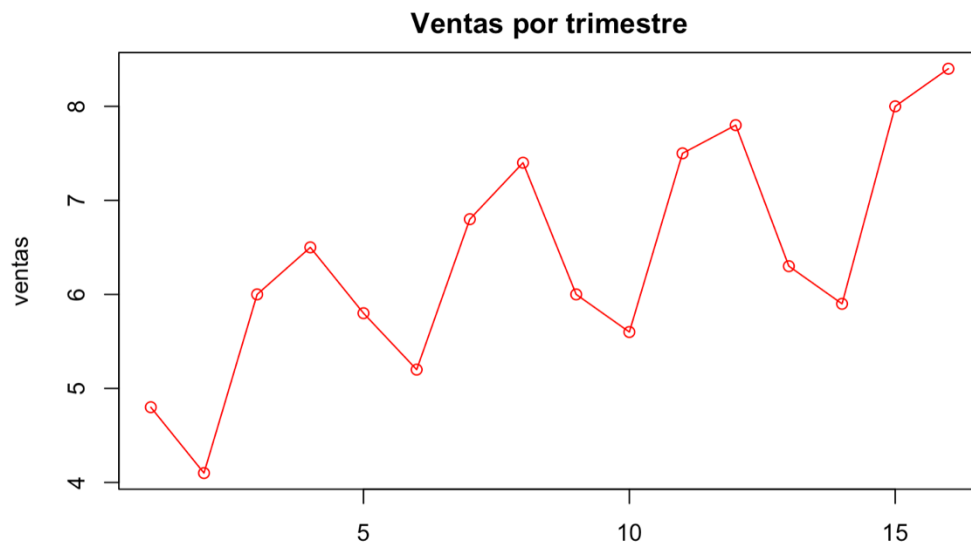
Introducción

Se busca crear un modelo que explique el comportamiento de las ventas de televisores a lo largo del año. Esto es importante ya que permite al negocio tomar decisiones estratégicas a futuro y estar mejor preparado para los próximos años.

Para esto es bueno preguntarse: ¿las ventas son iguales a lo largo de los años? ¿existe alguna tendencia o ciclo? ¿pudieran predecirse las ventas de un trimestre o de un año en base a un modelo estadístico?

Análisis de los resultados

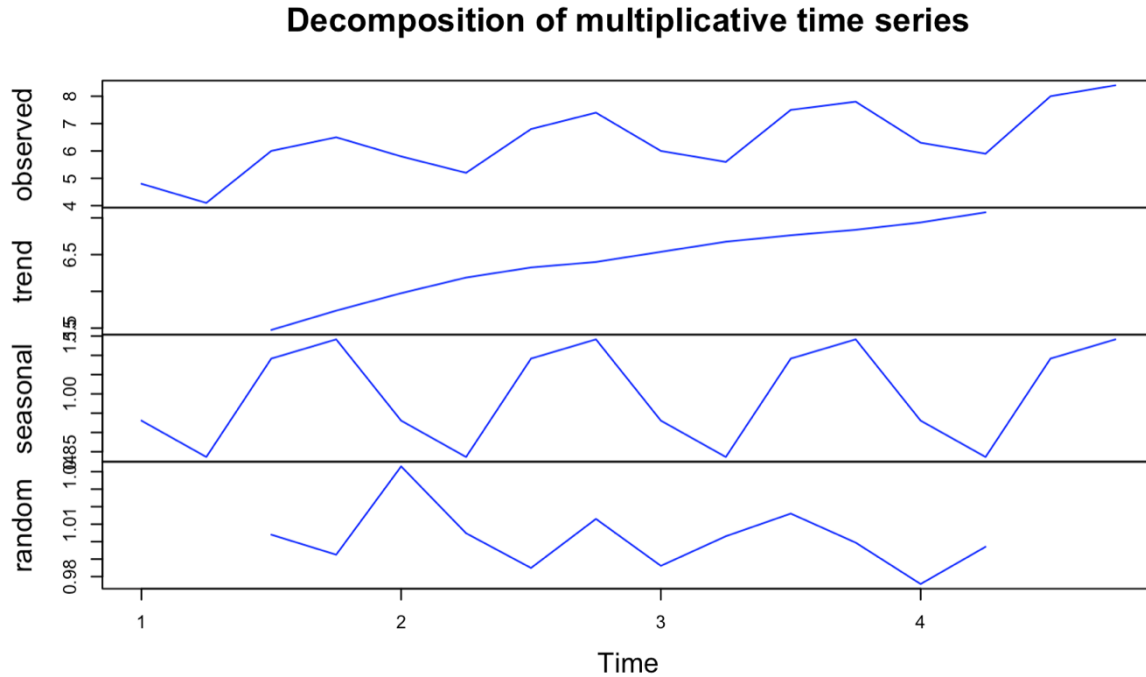
Después de cargar los datos lo primero que se hizo fue una gráfica de dispersión para poder analizarlos visualmente.



Se puede ver que existe una tendencia muy marcada siendo las ventas más bajas en el primer y segundo trimestre y las más altas en el tercero y cuarto.

Lo siguiente fue realizar el análisis de tendencia y estacionalidad. El procedimiento normal sería calcular los promedios móviles, después los promedios móviles centrados, luego los valores estacionales irregular y así obtener los índices estacionales.

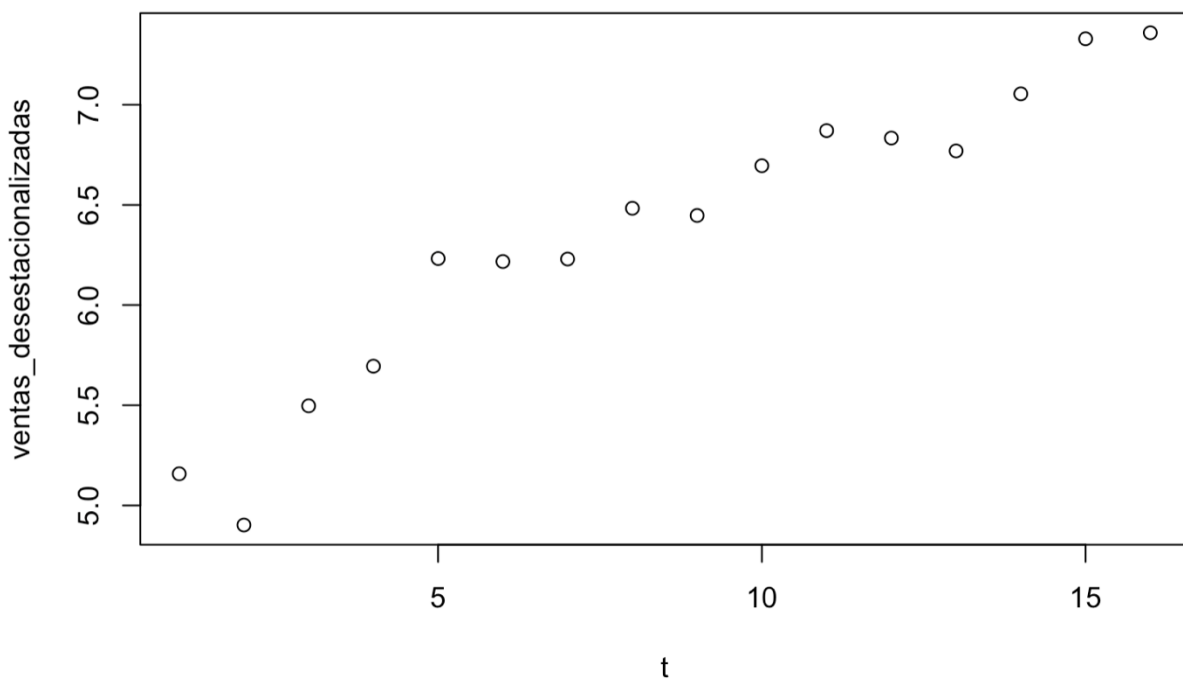
En nuestro caso utilizamos la función `ts()` de R y `decompose()` que nos permite obtener esta información con un comando. A ésta última le agregamos el parámetro para que sea una serie multiplicativa y no aditiva. Esto porque dio mejores resultados al hacer el modelo de regresión lineal.



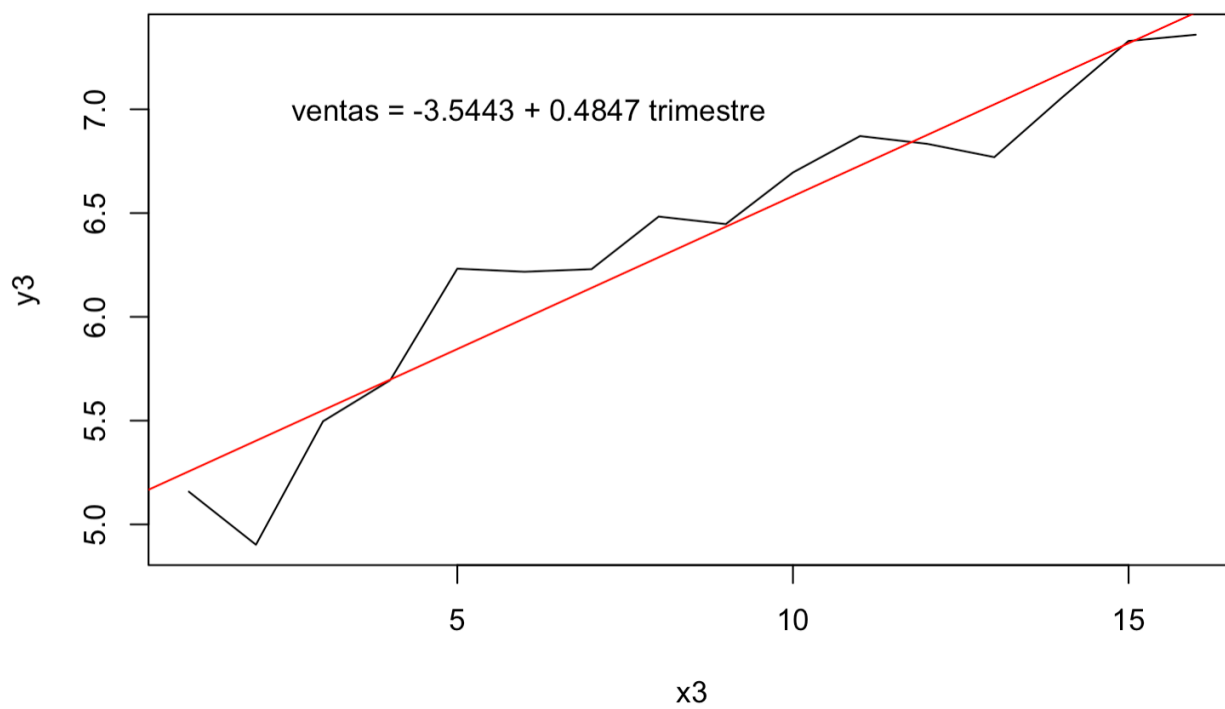
Por medio de T_{seasonal} se obtuvieron los índices estacionales, donde se puede ver cómo los primeros dos trimestres tienen valores menores a 1 y los últimos dos tienen valores superiores a 1.

| | Qtr1 | Qtr2 | Qtr3 | Qtr4 |
|---|-----------|-----------|-----------|-----------|
| 1 | 0.9306617 | 0.8363763 | 1.0915441 | 1.1414179 |
| 2 | 0.9306617 | 0.8363763 | 1.0915441 | 1.1414179 |
| 3 | 0.9306617 | 0.8363763 | 1.0915441 | 1.1414179 |
| 4 | 0.9306617 | 0.8363763 | 1.0915441 | 1.1414179 |

Se procedió a realizar el análisis del modelo lineal de la tendencia. Primero se graficaron las ventas desestacionalizadas, las cuales sirvieron como un método de suavización.



Ya con las ventas desestacionalizadas se creó el modelo de regresión lineal.



```
lm(formula = y3 ~ x3)
```

Coefficients:

```
(Intercept)      x3
      5.1080      0.1474
```

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

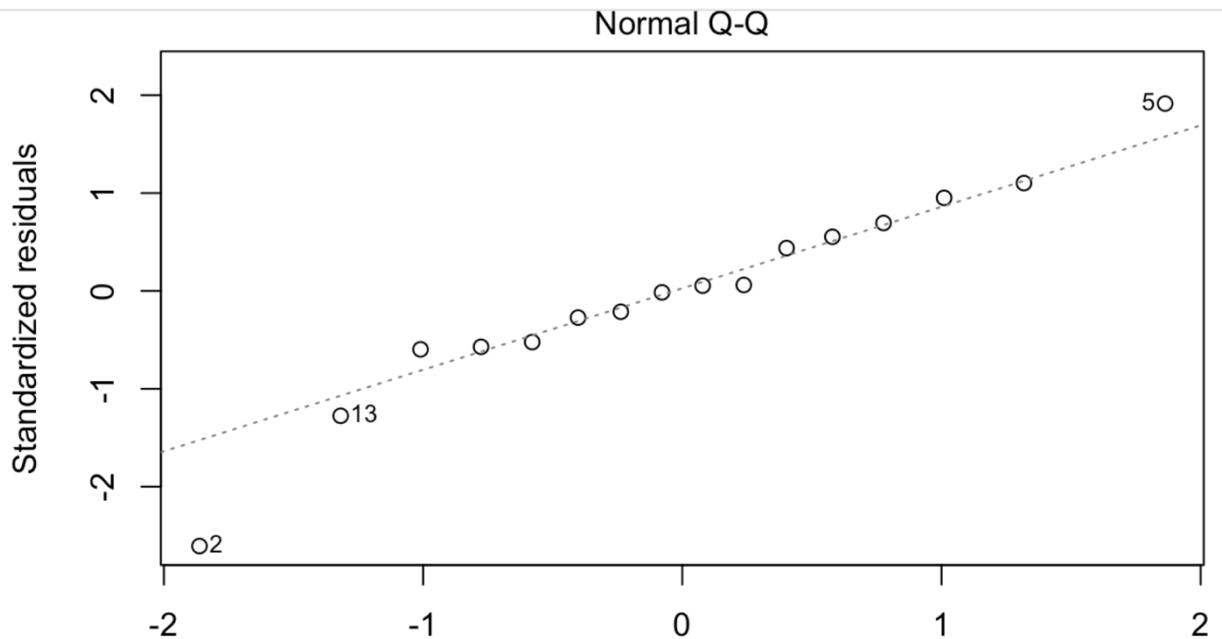
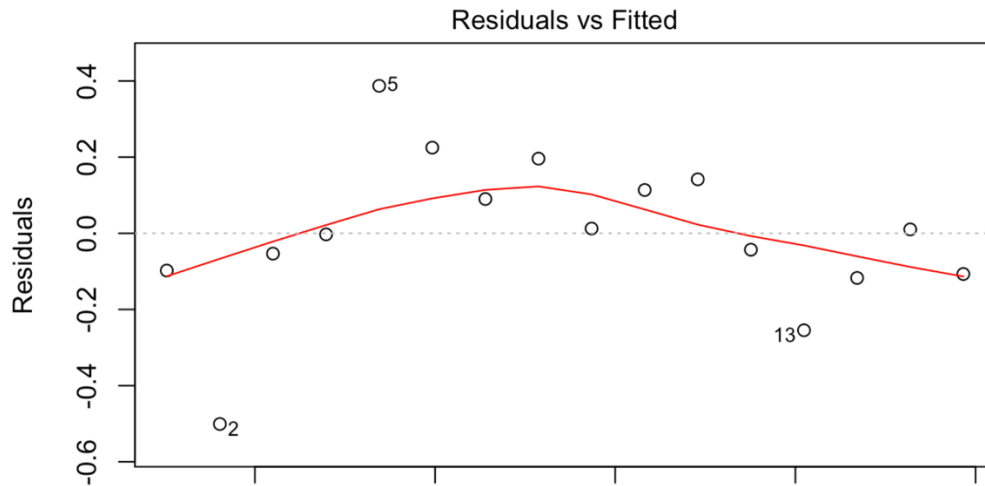
Multiple R-squared: 0.9208, Adjusted R-squared: 0.9151

F-statistic: 162.7 on 1 and 14 DF, p-value: 4.248e-09

El modelo explica un 91.5% de los datos aproximadamente y cuenta con un P-value menor a 0.05, por lo que podemos decir que es significativo.

Realizamos un análisis de residuos para validar el modelo. También vemos la significancia de beta1.

```
              2.5 %    97.5 %
(Intercept) 4.8684475 5.3476363
x3          0.1226037 0.1721603
```



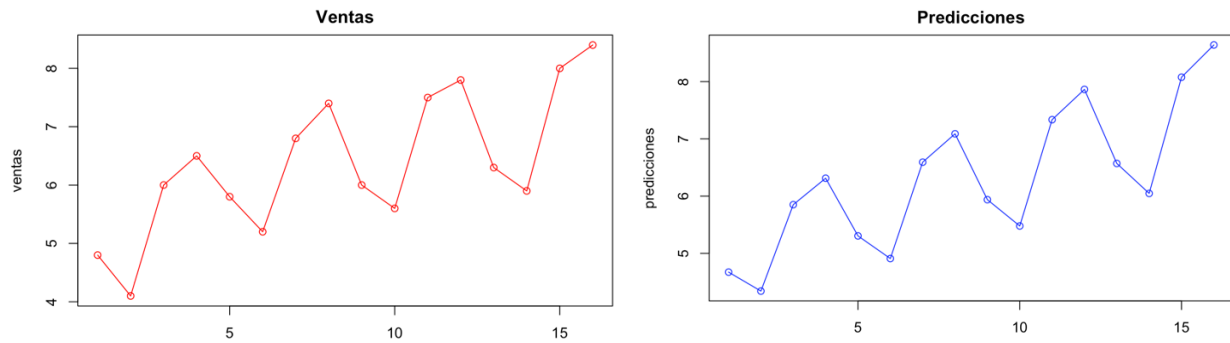
En la primera gráfica se puede ver como existe homocedasticidad en los residuos. En la segunda gráfica se puede ver que existe normalidad en los residuos. Asimismo, se puede ver que β_1 sí tiene significancia.

Por último, solo queda verificar que la media de los residuos sea 0. Al hacer el cálculo vemos que el número es prácticamente 0, por lo cual podemos validar el modelo.

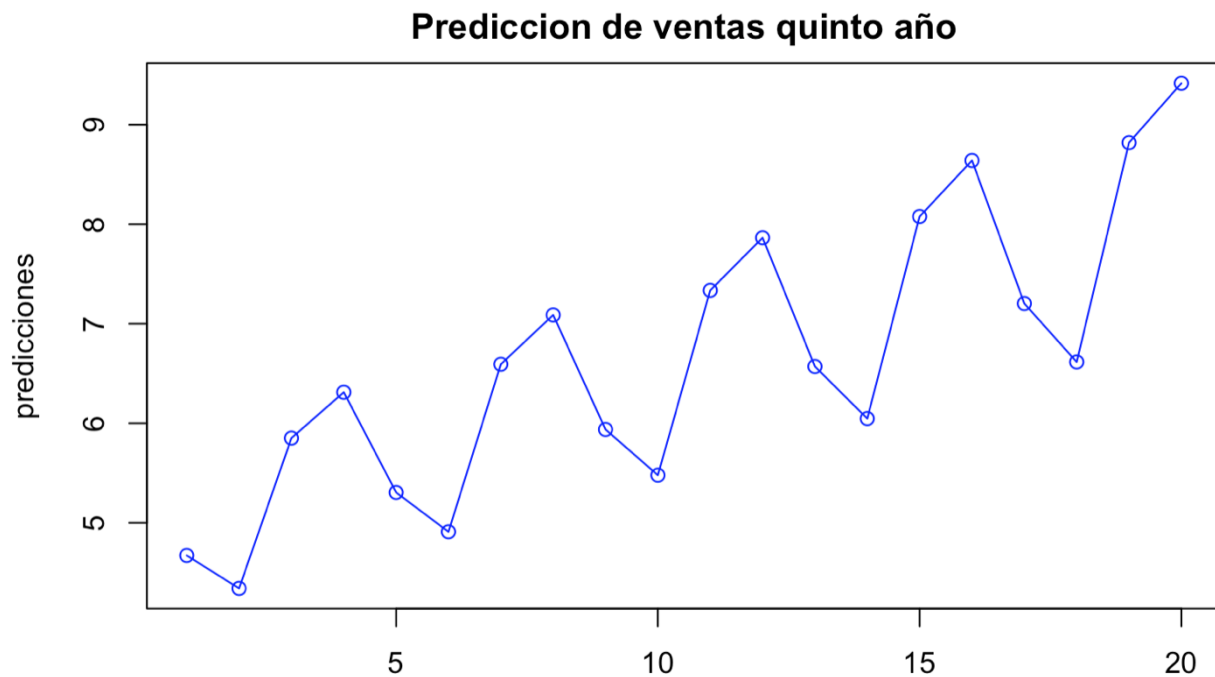
$-7.589415e-18$

Se calculó el cuadrado medio de los errores (CEM), el cual arrojó 0.05.

Después de validar el modelo se graficaron las ventas reales vs las predicciones.



Se puede observar que las predicciones son muy parecidas a los valores reales. Con el mismo modelo hicimos las predicciones del quinto año (trimestres 17,18,19 y 20).



Concluye

El modelo obtenido sí presenta predicciones acertadas tomando en cuenta las tendencias y ciclos de la serie de tiempo, teniendo un 91% de valor de r-ajustada cuadrada. Después de hacer el análisis de los residuos se demostró que había homostacidad, normalidad y media igual a 0. Podemos concluir que la predicción del quinto año será sumamente acertada si no existen eventos que afecten la tendencia actual.

Link a carpeta de drive:

https://drive.google.com/drive/folders/18YboF81-MMI1feH0YOaTYkiiHvsZ-RyQ?usp=share_link