

Inteligencia Artificial para la Ciencia de Datos

# Reporte final de “Venta de televisores”

---

Yoceline Aralí Mata Ledezma

A01562116

Instituto Tecnológico y de Estudios Superiores de Monterrey

Módulo 5. Estadística avanzada para ciencia de datos

Profesora: Blanca Ruiz

Grupo: TC3007C.502

03 de diciembre de 2022

# Índice

<b>Índice</b>	<b>2</b>
<b>Resumen</b>	<b>2</b>
<b>Introducción</b>	<b>2</b>
<b>Análisis de los resultados</b>	<b>3</b>
Análisis de tendencia y estacionalidad	3
Regresión lineal	4
Análisis de la pertinencia del modelo lineal	5
<b>Conclusión</b>	<b>9</b>
<b>Anexo</b>	<b>10</b>

## Resumen

Los hábitos de consumo de las personas cambian a través del tiempo y también pueden seguir patrones que se repiten dependiendo del mes. Un ejemplo, es como las ventas de muchos artículos aumentan en las fechas cercanas a navidad.

A continuación se analiza la venta de televisores de un conjunto de datos con herramientas de visualización y se realiza una regresión lineal para la predicción de ventas futuras.

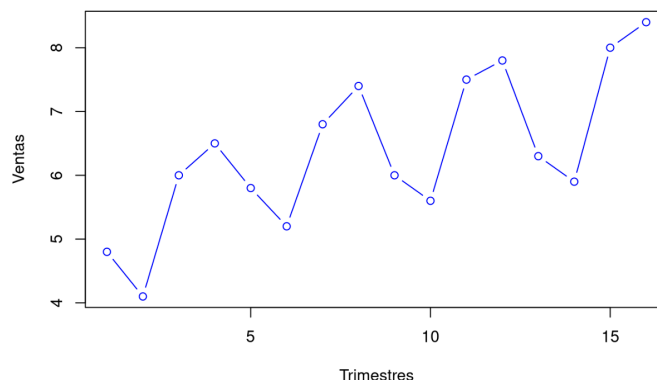
## Introducción

Actualmente, es importante el uso de herramientas estadísticas para la predicción de ventas, con el propósito de que cualquier empresa pueda contar con los artículos suficientes para satisfacer la demanda y de igual forma que estos artículos no sean más de los necesarios. Un ejemplo de este tipo de problemas es la venta de televisores, los cuales se analizan con herramientas para series de tiempo.

En este reporte se analiza un conjunto de datos sobre la venta de televisores y se realizan predicciones de venta para el siguiente año.

## Análisis de los resultados

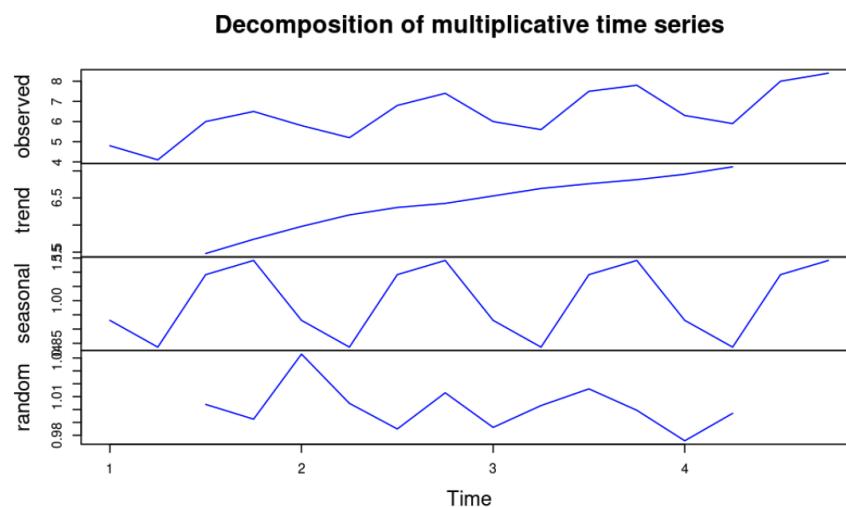
### Gráfico de dispersión



Se realizó un gráfico de dispersión para analizar el comportamiento de la serie de tiempo, se observó que hay una tendencia a través de los años, así como ciclos que se cumplen en cada uno de estos, por lo que se trata de una serie no estacionaria, lo que quiere decir que la venta de televisores va aumentando cada año.

## Análisis de tendencia y estacionalidad

### Descomposición de la serie de tiempo



Se realizó una descomposición de la serie de tiempo para analizar los componentes de esta por separado, se encontró que hay algo de estacionalidad, se confirma que hay una tendencia y también algo de aleatoriedad.

### Modelo lineal de la tendencia

	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

También se analizó el modelo lineal de la tendencia, en donde se observó que las ventas de televisores han aumentado en los últimos dos semestres de cada año.

## Regresión lineal

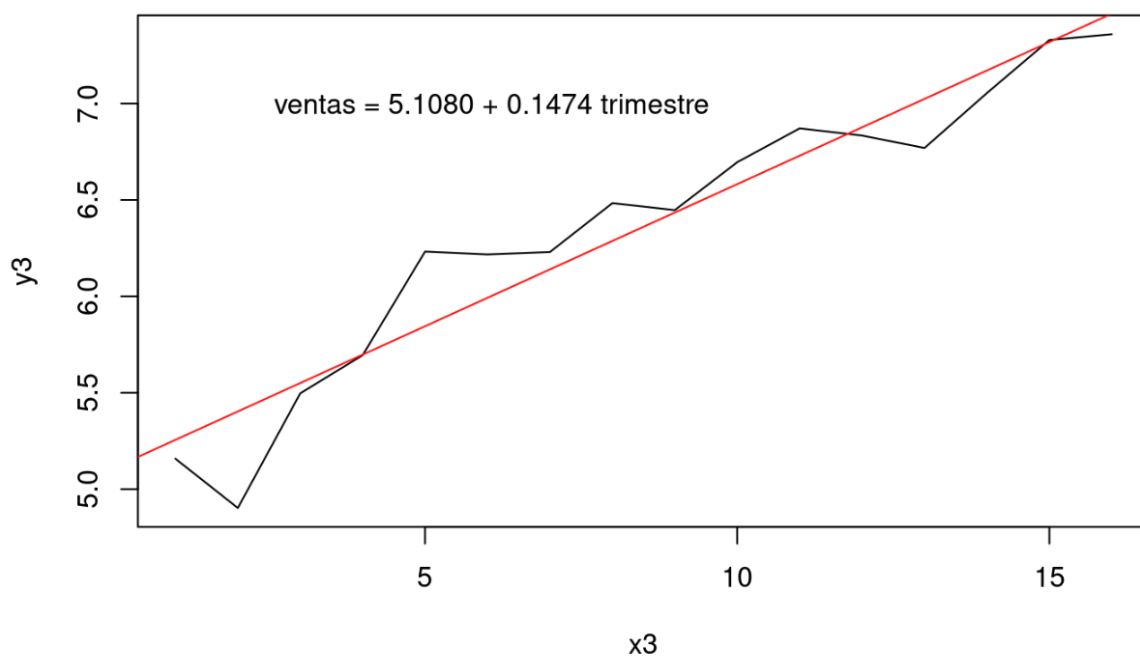
Se realizó una regresión lineal de la tendencia de ventas desestacionalizadas vs tiempo, para posteriormente predecir las ventas de televisores el próximo año.

### Resultados de regresión lineal

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

Coefficients:
(Intercept)          x3
      5.1080         0.1474
```

### Recta de regresión lineal



Se observó como la recta obtenida por la regresión lineal sigue la tendencia de las ventas.

Posteriormente para analizar la pertinencia de este modelo, se analizaron la significancia de  $\beta_1$ , el coeficiente de determinación y el análisis de residuos, esto tomando el nivel de significancia como 0.05.

## Análisis de la pertinencia del modelo lineal

### Significancia de $\beta_1$

- Hipótesis

$$H_0: \beta_1 = 0$$

$$H_0: \beta_1 \neq 0$$

- Regla de decisión:

Se rechaza la hipótesis nula si:

$$p < 0.05$$

- Análisis

```
##
## 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
```

- Conclusión:

Ya que la p es menor a 0.05, se rechaza la hipótesis nula y se concluye que

$\beta_1$  es significativa.

### Variabilidad explicada por el modelo

El coeficiente de determinación es de 0.9208 lo cual indica que hay una alta variabilidad explicada por el modelo.

## Análisis de los residuos y prueba de normalidad

Para la prueba de normalidad se decidió utilizar la prueba de normalidad de Shapiro-Wilk.

- Hipótesis

$H_0$ : Los residuos son normales

$H_0$ : Los residuos no son normales

- Regla de decisión:

Se rechaza la hipótesis nula si:

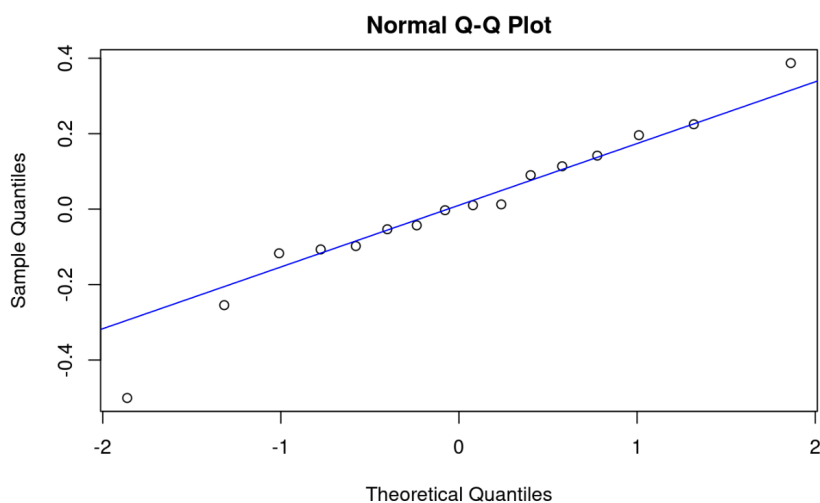
$$p < 0.05$$

- Análisis

Shapiro-Wilk normality test

data: residualsN3

W = 0.96379, p-value = 0.7307



- Conclusión:

Al observar la qqplot se puede decir que los datos son normales ya que se ajustan casi idealmente a la línea. También al hacer la prueba de normalidad Shapiro-Wilk se obtuvo que el valor p es mayor a 0.05, por lo cual no se rechaza la hipótesis nula y se concluye que los residuos son normales.

## Verificación de media cero

Para la verificación de media cero se utilizó la prueba

- Hipótesis

$$H_0: \mu = 0$$

$$H_0: \mu \neq 0$$

- Regla de decisión:

Se rechaza la hipótesis nula si:

$$p < 0.05$$

- Análisis

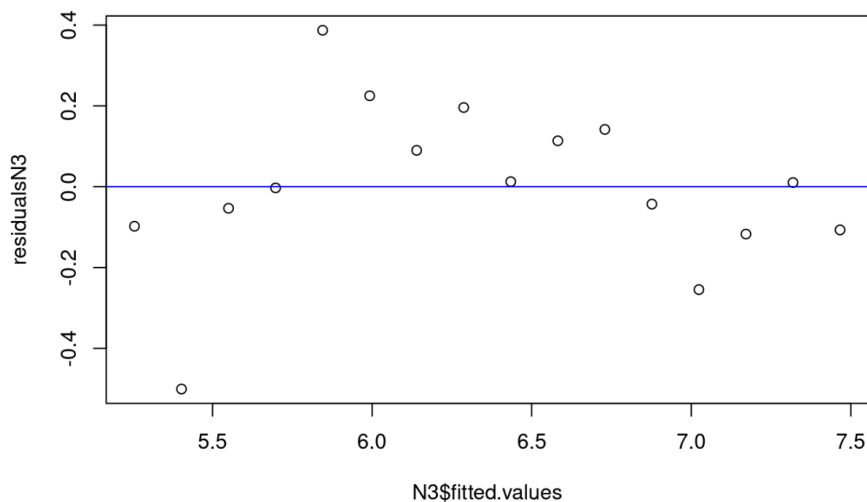
One Sample t-test

```
data: residualsN3
t = -2.9502e-17, 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
-1.517883e-18
```

- Conclusión

Se obtuvo que el valor p es mayor a 0.05, por lo que no se rechaza la hipótesis nula y se concluye que la media de los residuos es igual a 0.

## Homocedasticidad



No se observa homocedasticidad lo cual podría indicar que hay dependencia entre los residuos.



## CME y el EPAM de la predicción de la serie de tiempo

Otro paso en el análisis de la pertinencia del modelo fue observar el error de sus predicciones, para esto se utilizaron las métricas CME y EPAM.

### Predicciones del modelo

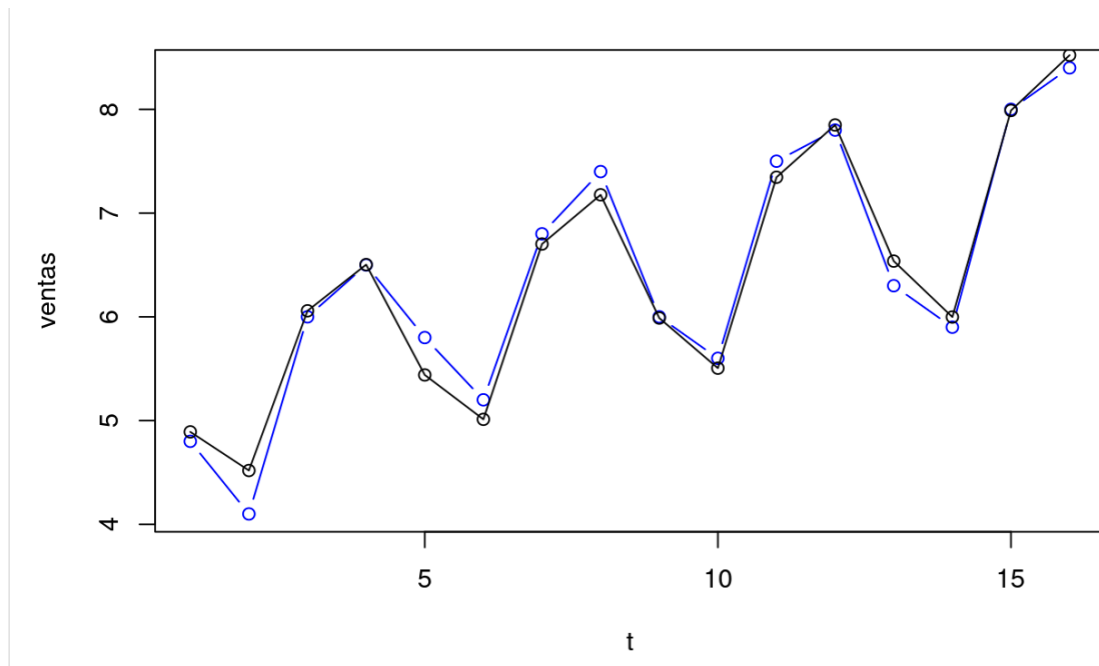
```
[1] 4891.000 4518.774 6058.288 6503.343 5439.718 5011.901 6701.862 7176.323 5988.436  
[10] 5505.029 7345.437 7849.303 6537.154 5998.156 7989.011 8522.283
```

### CME y EMAP

```
[1] 0.03302078
```

```
[1] 0.02439396
```

### Gráfico de los valores de las ventas y las predicciones vs el tiempo



Con este modelo se realizó el pronóstico de ventas para el siguiente año y se obtuvo el siguiente resultado:

```
[1] 7085.872  
[1] 6491.284  
[1] 8632.585  
[1] 9195.263
```

### **Conclusión sobre el modelo**

Los resultados en el análisis de la pertinencia del modelo, así como las métricas y el gráfico de los valores de las ventas y las predicciones vs el tiempo, nos indica que se trata de una buena herramienta para realizar el pronóstico de las ventas de estos televisores.

## **Conclusión**

Con este reporte se concluye que herramientas como la regresión lineal, pueden ser muy útiles al tratar con series de tiempo no estacionarias. De la misma forma, es importante conocer cómo se comportan estos datos, para así saber tratarlos de la forma más óptima. En este caso, se cayó en cuenta que se trataba de una serie de tiempo no estacionaria, sin embargo pudo haberse tratado de una serie de tiempo estacionaria y habría sido necesaria la utilización de otros métodos, como promedios móviles.

## **Anexo**

[Link a notebook](#)