

Reporte Venta de Televisores

Felipe Yépez

2022-12-01

Módulo 5: Estadística Avanzada para ciencia de datos y nombre de la concentración Grupo 2

Resumen

Se analizó la venta de televisores de 4 años, cada uno con 4 trimestres de ventas totales. Se utilizó esta serie de tiempo para realizar predicciones de las ventas a futuro mediante el análisis de Serie de Tiempo No estacionaria.

Se analiza las ventas de televisores a lo largo del tiempo.

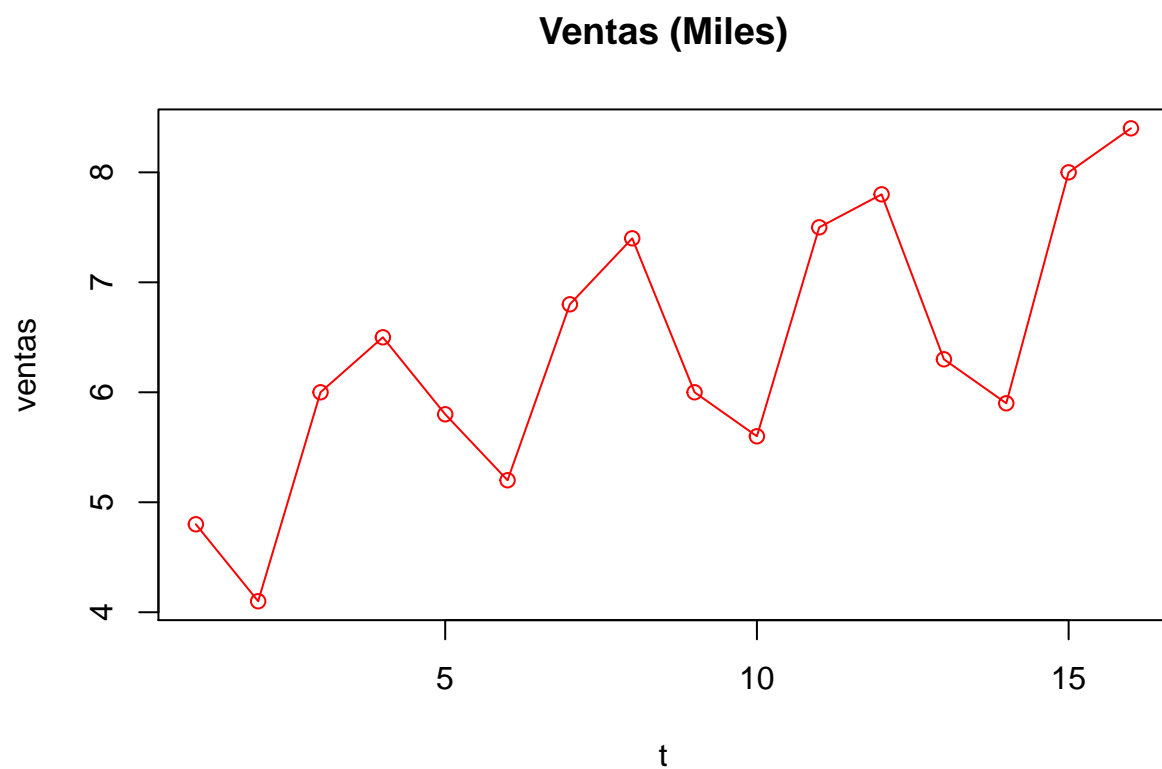
Introducción

Para un negocio es importante poder realizar predicciones informadas con la finalidad de que puedan tomar decisiones acertadas que favorezcan el bien y la perduración de su negocio a lo largo del tiempo.

Es importante poder realizar modelos lo más cercanos a lo que realmente pueda pasar para poder tomar decisiones informadas.

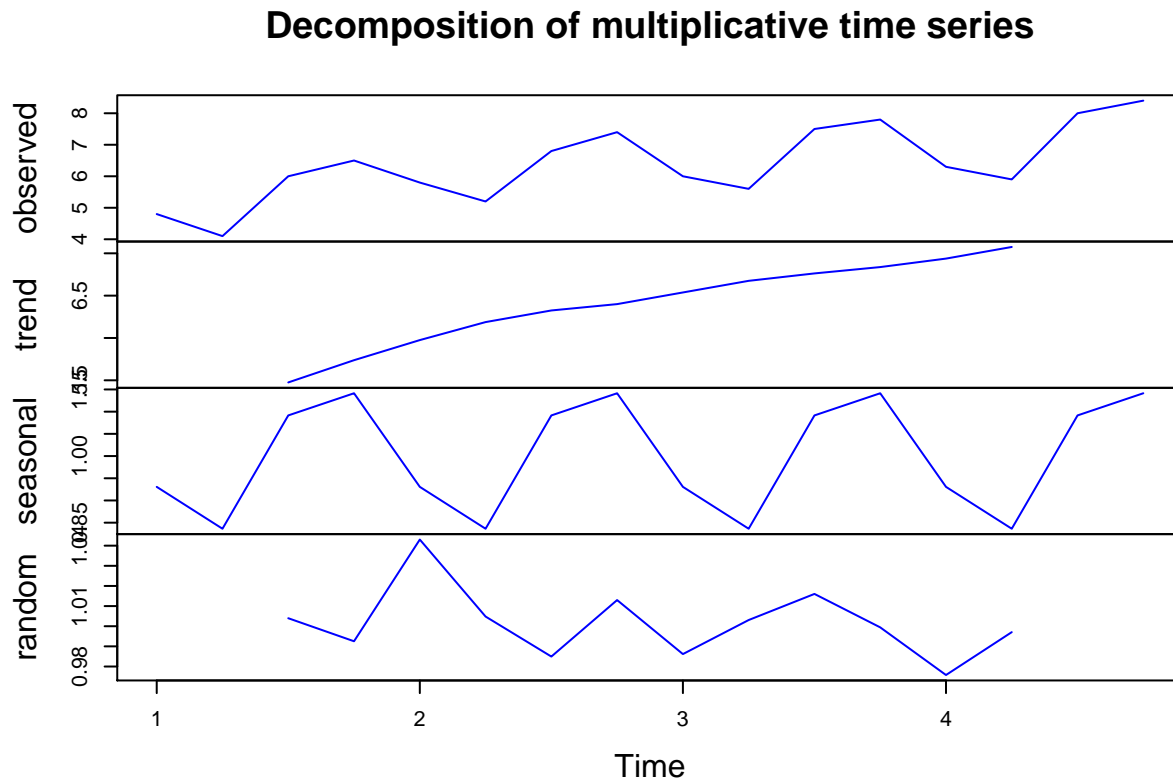
Explorando la base

Gráfica de dispersión de los datos:



Se puede observar en la gráfica anterior que existe tendencia alcista con ciclos de 4 trimestres. Cada trimestre las ventas van en aumento respecto al anterior. Se observa un canal alcista en la gráfica con límites inferiores en el 2do trimestre de cada año y con límites superiores en cada 4to trimestre.

Descomposición de la serie de tiempo



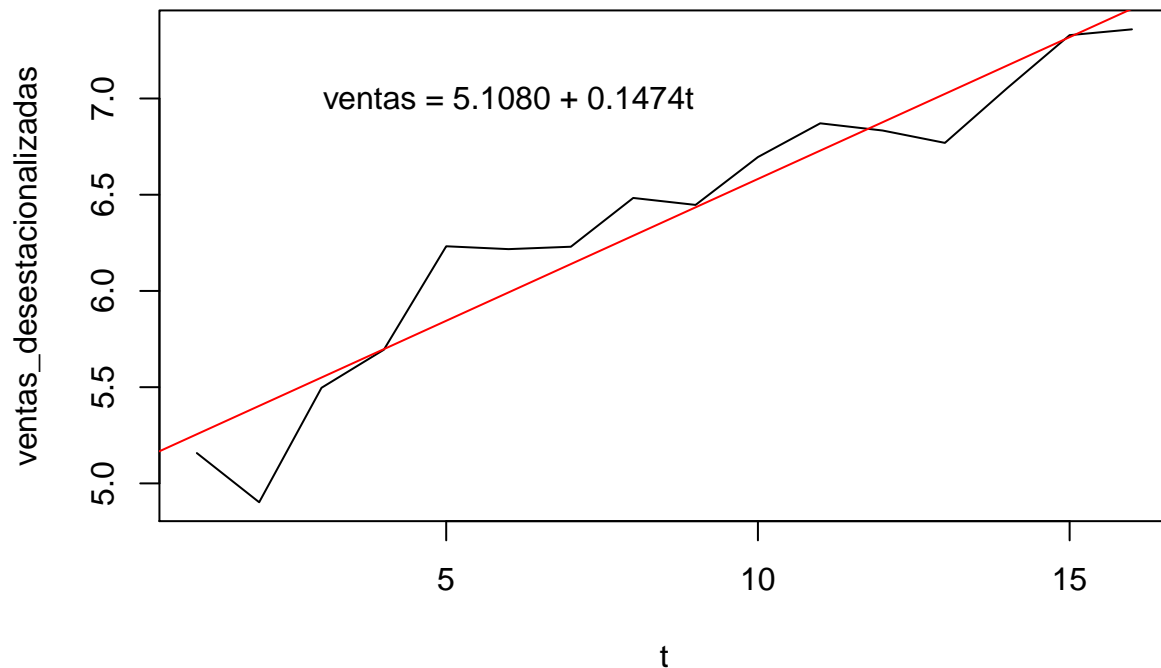
Se trata de una serie no estacionaria ya que existe tendencia en la gráfica que va en aumento.

En la gráfica de Observed se pueden visualizar los valores originales de la serie de tiempo.

Seasonal muestra la tendencia estacional mientras que random muestra la tendencia aleatoria, es decir los errores entre lo predicho en la tendencia estacional y lo observado.

Regresión lineal de la serie de tiempo desestacionalizada

```
##  
## Call:  
## lm(formula = ventas_desestacionalizadas ~ t)  
##  
## Coefficients:  
## (Intercept)          t  
##      5.1080      0.1474
```



Significancia de la regresión lineal

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

Se puede observar según los p values obtenidos que el modelo es significativo ya que se puede rechazar la hipótesis nula de que B1 es diferente de 0 y por lo tanto existe dependencia en el modelo. Los p values no superan el valor de significancia $\alpha = 0.05$

Se obtiene un valor R cuadrada de 0.9208 por lo que la variabilidad explicada por el modelo es alta, superando el 90%. Y una R cuadrada ajustada de 0.9151

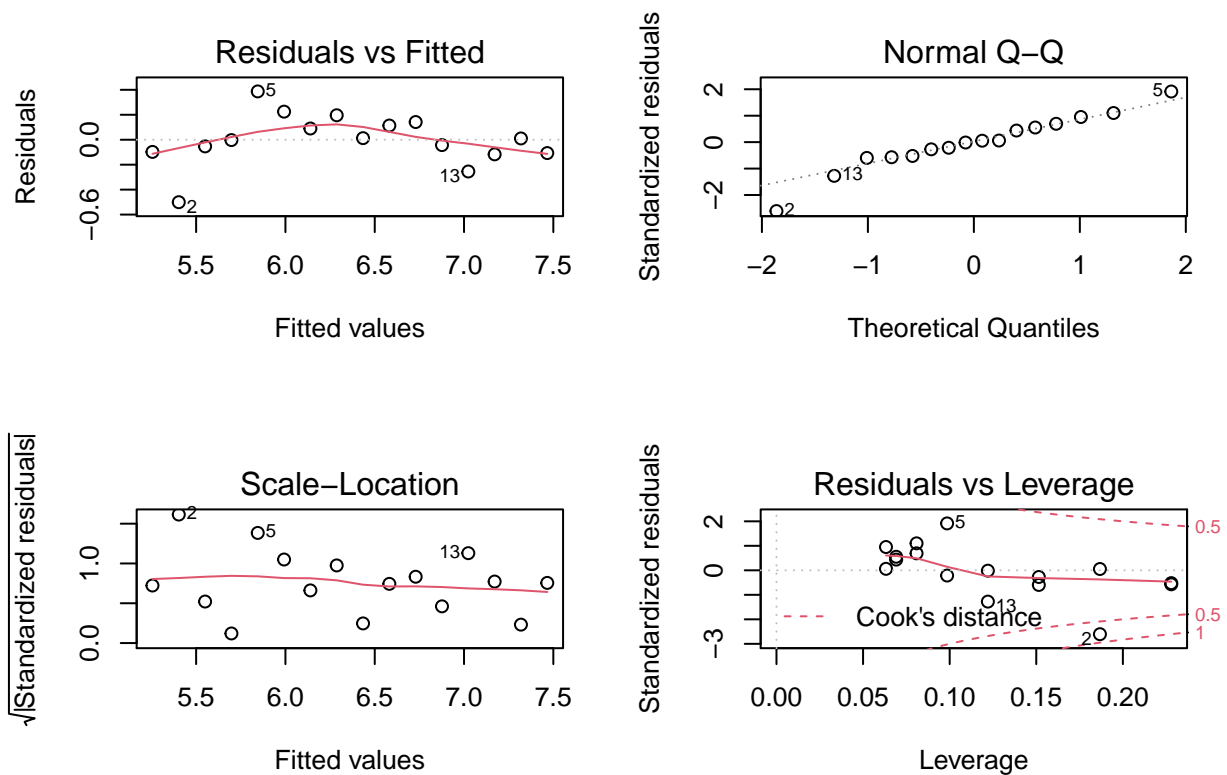
Normalidad de residuos

Mediante Shapiro Wilk se evaluará la normalidad con la finalidad de saber si el conjunto de datos proviene de una distribución normal.

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

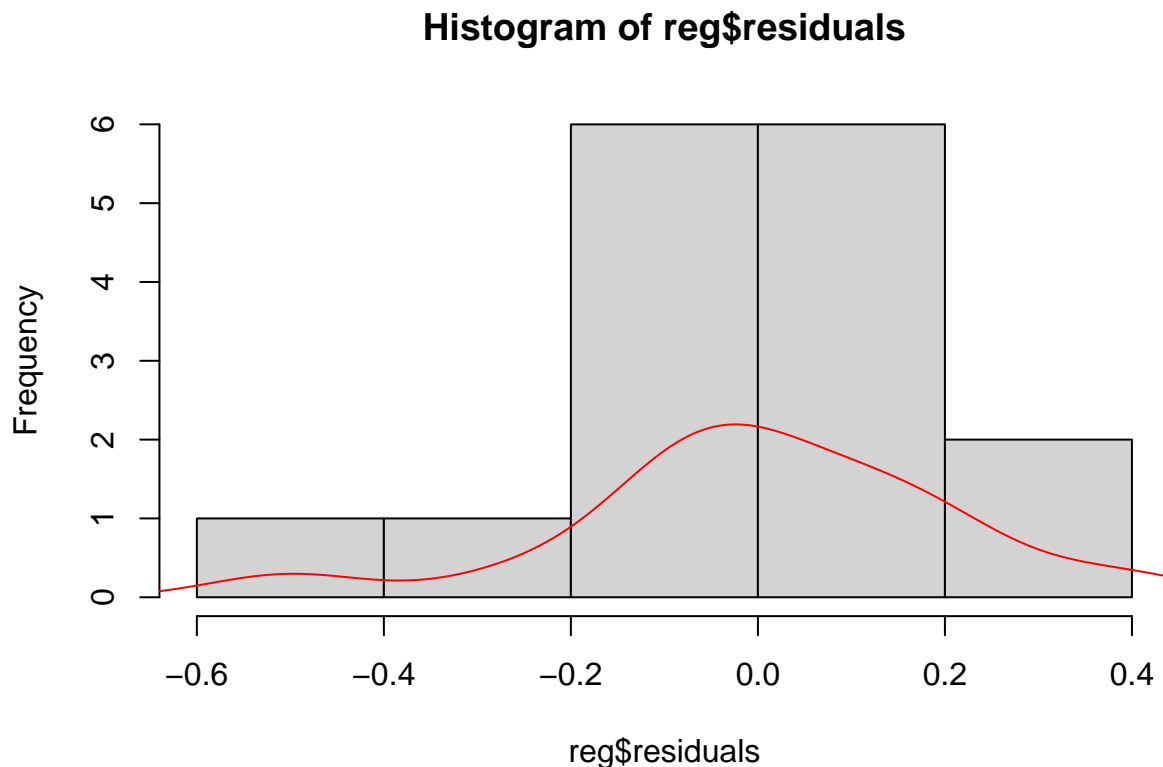
La hipótesis nula dice que los datos provienen de una población normal. Utilizando el nivel de significancia de $\alpha = 0.05$, no se puede rechazar H_0 por lo que se puede asumir que los residuos provienen de una distribución normal.

Homocedasticidad e Independencia



Se puede observar en las anteriores gráficas de residuos que existe normalidad en el conjunto de datos. El QQPlot muestra normalidad, algo que ya se había logrado determinar con la prueba Shapiro.

Los residuos no muestran ninguna estructura evidente por lo que se puede asumir que el modelo de regresión es adecuado y cuenta con homocedasticidad e independencia.



Existe poco de sesgo a la izquierda en la distribución de los residuos, sin embargo, están bien distribuidos. Se prueba la media de los residuos.

```
##
## One Sample t-test
##
## data: reg$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
```

Sale un valor muy cercano a 0 para la media de residuos.

Promedio de los errores porcentuales de la predicción de la serie de tiempo

Mediante el CME se puede conocer la diferencia cuadrática media entre los valores estimados y los valores reales de la distribución.

```
## [1] 40.96092
```

Se obtiene un CME de 40.96 para el promedio móvil, un valor un tanto elevado que indica que existe diferencia entre la regresión y el cuadrado de los puntos.

Conclusión del modelo

En general el modelo es bueno por los anteriores hallazgos, se ajusta y explica la variabilidad de la venta de televisores y existe normalidad en los datos. El modelo tiene dependencia con sus variables. Existe normalidad, homocedasticidad e independencia.

Por lo anterior, se puede utilizar el modelo para pronosticar las ventas que se tendrá de televisores los siguientes 4 trimestres. El valor de R cuadrada indica que el modelo tiene buen ajuste al conjunto de datos.

Pronóstico para el siguiente año

Se obtienen las predicciones para los 4 trimestres del siguiente año en (miles).

```
## [1] 7.085872
```

```
## [1] 6.491284
```

```
## [1] 8.632585
```

```
## [1] 9.195263
```

Se pueden utilizar estas predicciones para tomar buenas decisiones en el negocio de venta de televisores.

Conclusión

Se puede concluir que realizar modelos de predicción mediante el conocimiento de series de tiempo puede llegar a ser beneficioso para negocios como en este caso de venta de televisores. Se puede llegar a realizar predicciones de lo que sucederá con mayor confianza al evaluar la pertinencia del modelo y su ajuste.

El modelo obtenido para pronosticar las ventas de los siguientes 4 trimestres, es bueno por los anteriores hallazgos.

Mediante series de tiempo se puede sintetizar la información de las ventas para poderlas ver de forma más sencilla y entender el comportamiento que las mismas tienen por trimestre.

Se puede conocer cuáles serán los trimestres con menos ventas y más ventas. Por ejemplo, se puede tomar decisiones de cuántas personas tener en un local de venta de televisores dependiendo de las ventas y la concurrencia que los locales tengan.

Anexos

Documento de análisis. <https://github.com/FelipeYepez/Inteligencia-Artificial-Avanzada2/tree/main/Estad%C3%ADstica%20Avanzada>