1.1 Bases generales

“La manera más confiable de predecir el futuro es tratando de entender el presente” John Naisbitt (como se cita en Frases go, s. f., https://goo.gl/Ymv5xU).

Pronosticar la variación de ciertas variables en el tiempo es una actividad que se ha desarrollado desde los inicios de la sociedad moderna. En esta parte se brindarán algunos casos de cómo evaluar esta evolución.

Bibliografía recomendada: Anderson, Sweeney y Williams (2011).

1.1.1 Introducción

En esta sección se abordarán conceptos que tratan sobre el análisis de series de datos temporales y cómo basarse en ellas para pronosticar o predecir situaciones. Estos tipos de estudios son muy comunes en las empresas, ya que muchas de sus decisiones sobre estrategias de ventas, mercadeo y promociones se fundamentan en tratar de entender, con base en lo conocido, cómo evolucionará el mercado. En este sentido, es necesario poseer intuición, precaución y un buen juicio sobre diversas variables que van desde lo político y lo social hasta lo natural (clima, estaciones, etc.).

En el trabajo de predecir, uno se puede apoyar sobre métodos cualitativos o cuantitativos. Los primeros son apropiados cuando los datos históricos de la variable que se debe predecir no están disponibles o no se pueden aplicar. Por otro lado, los métodos cuantitativos se pueden aplicar cuando sí se posee información pasada sobre la variable que se debe predecir, está cuantificada y se puede asumir que el comportamiento observado en el pasado continuará en el futuro. Para el caso de esta unidad, se tratará sobre algunos de los métodos cuantitativos más comunes utilizados para predicciones. Cuando la información sobre los datos se restringe a valores pasados que tomó la variable que se quiere predecir, el procedimiento predictivo es denominado método de series temporales, según el cual se tiene como objetivo encontrar un comportamiento de esta cantidad en función del tiempo de una manera que se pueda correlacionar matemáticamente para generar, luego, valores a tiempos futuros.

1.1.2 Patrones en datos de series temporales

Una secuencia temporal indica la evolución de una determinada variable a lo largo del tiempo, donde las medidas pueden ser realizadas a distintas escalas de tiempo: horas, días, semanas, etcétera. Para entender cómo dicha variable se ha comportado en el pasado, es necesario encontrar un patrón en los datos y si se espera que este se mantenga en el futuro. En general, lo primero que se debe realizar cuando se trata con series temporales es realizar una gráfica donde se representa el valor que van adquiriendo los datos (eje vertical y) en función del tiempo (eje horizontal x).

Patrones horizontales

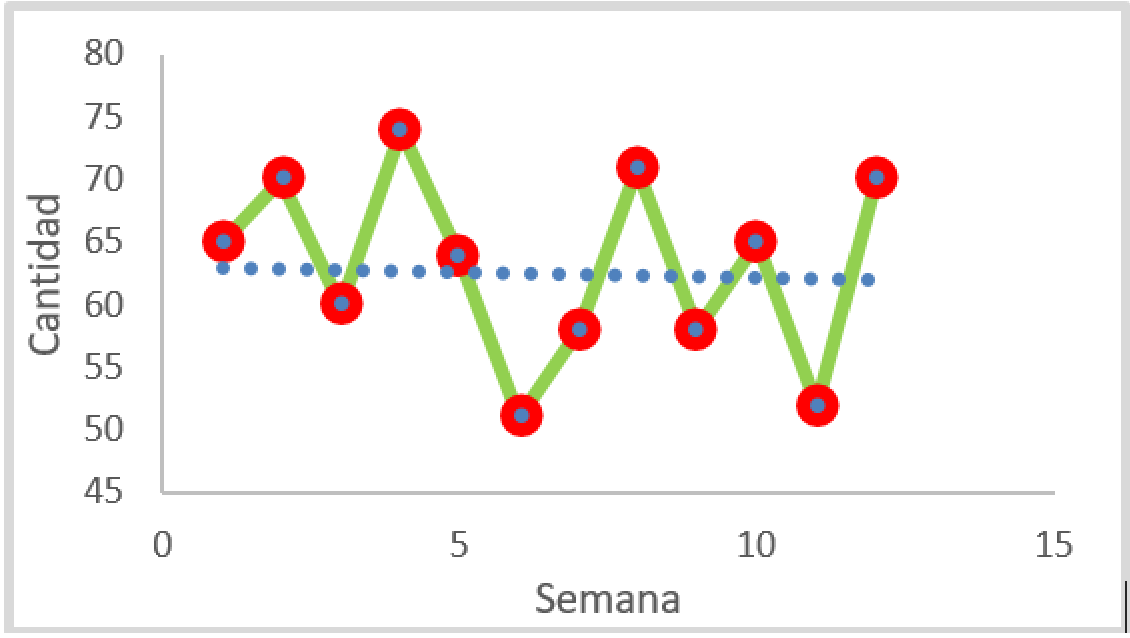
Un patrón horizontal, o estacionario, se puede deducir cuando los datos oscilan alrededor de una media constante. Por ejemplo: los datos resumidos en la Tabla 1, donde se representa gráficamente la cantidad vendida de automóviles por mes para una agencia de venta de vehículos.

Tabla 1: Serie temporal de venta de automóviles por semana

Semana	Cantidad
1	65
2	70
3	60
4	74
5	64
6	51
7	58
8	71
9	58
10	65
11	52
12	70

Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

Figura 1: Serie temporal de los datos de la Tabla 1



Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

Como puede apreciarse, el promedio de venta entre semanas es de 63 automóviles. En este caso, aunque haya oscilaciones, se puede decir que los datos siguen un patrón horizontal y se lo puede denominar como series temporales estacionarias, las cuales se destacan, ya que presentan una media y una variabilidad constante en el tiempo.

Patrones de tendencia

Generalmente, los datos a lo largo del tiempo exhiben fluctuaciones aleatorias, pero también, tomando períodos largos, movimientos hacia valores mayores o menores. Cuando esto sucede se dice que la serie temporal refleja un patrón con una tendencia. Supongamos que la agencia de venta de vehículos está por cerrar sus puertas en 2 meses (8 semanas) y por ese motivo va bajando los precios discretamente a lo largo del tiempo. Tomando los nuevos valores de venta en el tiempo, se observa lo representado en la Tabla 2 y Figura 2. Un analista dirá que existe una tendencia de aumento en la venta de vehículos a lo largo del tiempo, pero, en realidad, esto deberá contextualizarse en la situación puntual de los eventos (disminución de precio de los productos a partir de la semana 12; línea de puntos en la Figura 2). En la Figura 2, en azul se muestra la línea de tendencia y con un trazo vertical el momento en que disminuyeron los precios.

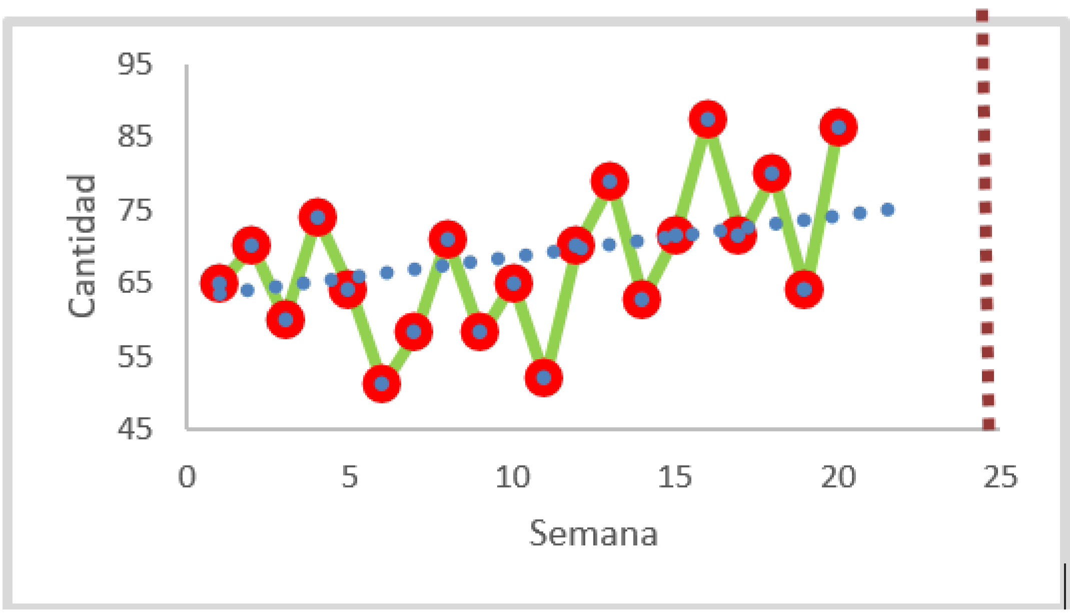
Tabla 2: Serie temporal de venta de automóviles por semana

Semana	Cantidad

1	65
2	70
3	60
4	74
5	64
6	51
7	58
8	71
9	58
10	65
11	52
12	70
13	79
14	63
15	71
16	87
17	71
18	80
19	64
20	86

Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

Figura 2: Serie temporal de los datos de la Tabla 2



Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

Patrones estacionales

La tendencia de una serie de datos se puede identificar mediante el análisis de los valores de la variable de interés en varios años. En este contexto, hay casos en los que se observan patrones repetidos de variaciones en las distintas estaciones del año. Un ejemplo claro sería la variación anual de la venta de suéteres, donde se puede predecir un aumento en las estaciones de otoño e invierno y una disminución en primavera y verano. Acá cabe mencionar que es clave la manera de representar los datos, ya que, si se observan los datos teniendo en cuenta todas las ventas en el año, la información de estacionalidad se perderá.

Patrones cilíndricos

Se dice que existe un comportamiento cíclico en una variable a lo largo del tiempo cuando reflejan valores por debajo y por encima de la línea de tendencia. En los negocios, y debido a diversos factores (inflación, deflación, política, etc.), se encuentra este tipo de patrones extremadamente difíciles de pronosticar.

1.1.3 Elección del método de pronóstico

Es importante determinar el patrón de evolución en una serie temporal para luego elegir un método determinado de pronóstico. Por este motivo es indispensable realizar una gráfica de evolución temporal (como se muestra en las figuras 1 y 2). Para cada tipo de patrón (horizontal, tendencia, etc.), existe un método apropiado para basar una predicción. Estos tópicos son la base de la siguiente unidad. Antes es necesario entender la precisión, o el error asociado, en series temporales.

1.1.4 Precisión de los pronósticos

¿Cuán exacta será una predicción basada en una serie temporal? Para esto, el concepto clave es conocer el error de pronóstico, el cual se define como:

$$Error\ de\ pronóstico = valor\ actual - valor\ predicho$$

Si volvemos al ejemplo de la Tabla 1 (venta de vehículos) podríamos utilizar un método naive de predicción, el cual propone que la venta de vehículos de la semana subsiguiente será igual a la de la anterior. Puntualmente, en la semana 1 se vendieron 65 vehículos, y en la semana 2, 70. Si la predicción es que en la semana 2 se deben vender 65, el error en el pronóstico es de 5 vehículos. El hecho de que el error tiene signo positivo implica que la predicción está subestimando el valor actual de ventas. En esta línea, si ahora se predice que en la semana 3 se venderán 70 vehículos y, actualmente, se vendieron 60, el error es de -10; en este caso, existe una sobrestimación. Una medida simple de la precisión es la media de los errores. La Tabla 3 ejemplifica este método de predicción: los errores totales y absolutos.

Tabla 3: Error de pronóstico. Método *naive*

Semana	Cantidad	Predicción	Error	Error absoluto	Error cuadrático	Error porcentual	Error porcentual absoluto
1	65						
2	70	65	5	5	25	7.1	7.1
3	60	70	-10	10	100	-16.7	16.7
4	74	60	14	14	196	18.9	18.9
5	64	74	-10	10	100	-15.6	15.6
6	51	64	-13	13	169	-25.5	25.5
7	58	51	7	7	49	12.1	12.1
8	71	58	13	13	169	18.3	18.3
9	58	71	-13	13	169	-22.4	22.4
10	65	58	7	7	49	10.8	10.8
11	52	65	-13	13	169	-25.0	25.0
12	70	52	18	18	324	25.7	25.7
		70					

Total	5	123	1519	-12.3	198.1

Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

De esto se puede decir que el error promedio en el pronóstico es igual al error total dividido por el número de errores calculado, o sea, 5/11=0,45. Al tener un valor positivo, el método de predicción está subestimando lo que implica que los valores observados son mayores que los pronosticados. Así, como las fluctuaciones se cancelan al considerar las sub y sobreestimaciones, una media más acorde es el *valor absoluto medio*. Este será el error absoluto total dividido por el número de errores calculados: 123/11=11,2. Otra medida que evita las desviaciones positivas o negativas es el *error medio cuadrático*, el cual se calcula sobre la suma de los cuadrados de los errores divididos por el número de errores. En este caso, será igual a 1519/11=138,1. Nuevamente, estos son errores totales del método y no son útiles para hacer comparaciones entre cada punto en el tiempo (semanas). Una medida que sí puede servir para este propósito es el *error porcentual medio absoluto*, que será el error total porcentual absoluto relativo al total de errores: 198,1/11=18 %.

Otro método para considerar en la predicción es el promedio de las series o valores históricos. Por ejemplo, para calcular el error en la semana, 3 se puede utilizar el promedio de ventas de la semana 1 y 2 y, de esta manera, calcular un error con base en esto. La Tabla 4 resume este nuevo método, con los errores asociados.

Tabla 4: Error de pronóstico. Método de promedios

Semana	Cantidad	Predicción	Error	Error absoluto	Error cuadrático	Error porcentual	Error porcentual absoluto
1	65						
2	70	65	5	5	25	7.1	7.1
3	60	68	-8	8	56	-12.5	12.5
4	74	65	9	9	81	12.2	12.2
5	64	67	-3	3	11	-5.1	5.1
6	51	67	-16	16	243	-30.6	30.6
7	58	64	-6	6	36	-10.3	10.3
8	71	63	8	8	62	11.1	11.1
9	58	64	-6	6	38	-10.6	10.6
10	65	63	2	2	2	2.4	2.4
11	52	64	-12	12	135	-22.3	22.3
12	70	63	7	7	56	10.6	10.6
		63					
Total			**-19**	**81**	**744**	**-48.0**	**134.8**

Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011

Para este caso, los errores medios absolutos y el error porcentual son de 7,4 % y 12,2 %, respectivamente. Si comparamos ambos métodos, se podría decir que el método de promedios tiene mayor precisión que el método naive.

1.2 Ejemplos y características para pronósticos

En esta parte se mostrarán algunos de los métodos aplicados a los distintos patrones en series temporales. El principio de estos métodos se basa en tratar de minimizar las fluctuaciones observadas en una serie temporal, con una aplicabilidad simple y un alto nivel de precisión para períodos cortos.

1.2.1 Promedios y suavización de datos

Medias móviles

Este método usa el promedio de los datos k más recientes en la serie temporal como valor para predecir sobre el período siguiente. Esto se expresa como:

$$\boldsymbol{F}_{(t+1)} = \frac{\boldsymbol{Y}_t + \boldsymbol{Y}_{t-1} \ldots + \boldsymbol{Y}_{t-k+1}}{\boldsymbol{k}}$$

Donde $F_{(t+1)}$ es el valor pronosticado para la serie en el período $t+1$ e Y_t es el valor actual de la variable. Para aplicar este método, es necesario definir el número de valores de la serie temporal que serán promediados. Si los valores recientes son los importantes, generalmente se utiliza un valor bajo de k. Por el contrario, si los valores pasados tienen relevancia en el pronóstico, valores mayores de k son preferibles. Cambios en la serie temporal serán detectados a bajos valores de k, pero las fluctuaciones aleatorias son más marcadas, lo cual se refina al utilizar mayores valores de k. En esta línea, es necesario entender el problema que vaya a analizar para tener un criterio acertado sobre el valor de k que debamos definir.

Volviendo al ejemplo de la Tabla 1, si se utiliza un valor de k=3 (tres semanas), se puede obtener el siguiente pronóstico para la semana 4 y 5:

$$\boldsymbol{F}_4 = \frac{65+70+60}{3} = 65; \boldsymbol{F}_5 = \frac{70+60+74}{3} = 68.$$

Toda la serie hasta la semana 12 se resume en la Tabla 5.

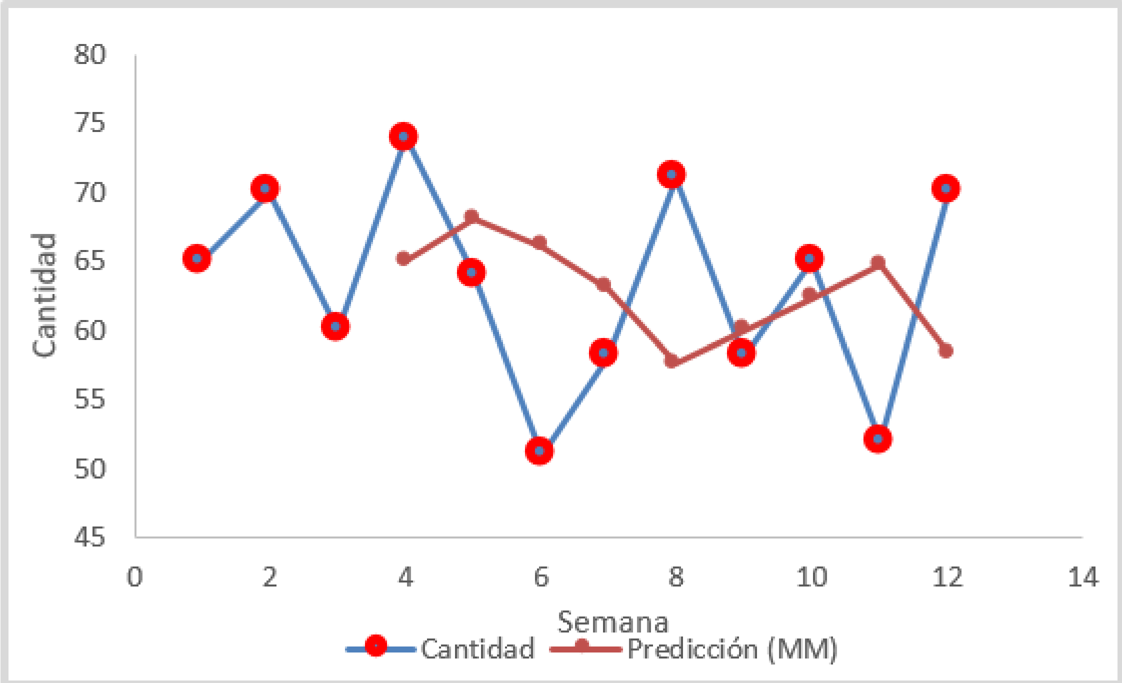
Tabla 5: Método de promedios móviles

Semana	Cantidad	Predicción	Error	Error absoluto	Error cuadrático	Error porcentual	Error porcentual absoluto
1	65						
2	70						
3	60						
4	74	65	9	9	81	12.2	12.2
5	64	68	4	4	16	-6.3	6.3
6	51	66	15	15	225	-29.4	29.4
7	58	63	5	5	25	-8.6	8.6
8	71	58	13	13	178	18.8	18.8
9	58	60	2	2	4	-3.4	3.4
10	65	62	3	3	7	4.1	4.1
11	52	65	13	13	160	-24.4	24.4
12	70	58	12	12	136	16.7	16.7
		63	62				
Total			**-2**	**75**	**832**	**-20.4**	**123.8**

Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

El gráfico que representa las *predicciones por media móvil* (predicción MM) comparadas con la serie temporal se resume en la Figura 5. Allí, como puede observarse, las fluctuaciones están suavizadas (menos marcadas).

Figura 5: Gráfica del método de promedios móviles



Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

Medias móviles ponderados

En este caso, a cada observación en el cálculo de la media móvil se le atribuye una ponderación (diferente peso) para el cálculo del promedio. Supongamos que se considera que el valor más reciente observado es el que más debería acercarse al pronosticado y, por ende, debería tener más peso en el cálculo del promedio. Esto se realiza al multiplicar cada observación por una fracción, donde la suma total de estas fracciones que multiplican a cada observación dé un total de 1. Esto se resume de la siguiente manera para los ejemplos anteriores, lo cual da una ponderación de 1/6, 2/6 y 3/6 (1/6+2/6+3/6=1):

$F_4 = 1/6 \times 65 + 2/6 \times 70 + 3/6 \times 60 = 64$

El valor de la ponderación va a depender del error o precisión del pronóstico, siempre tratando de ser minimizado.

1.2.2 Proyecciones de tendencias

Tendencias lineales (regresión lineal)

La regresión lineal es uno de los métodos más usados y simples para pronosticar la evolución de una variable, siempre y cuando la tendencia sea lineal (o aproximadamente lineal, dentro de las variaciones que presente). Un ejemplo clásico en los negocios es realizar un análisis de las ventas de un producto a lo largo del tiempo como en la Tabla 5, donde se resume la actividad de cierta empresa cuatrimestralmente a lo largo de tres años (12 cuatrimestres).

Una regresión es la relación funcional entre dos o más variables correlacionadas. Esto implica que el valor de una variable determina el valor de la otra y, por este motivo, si se entiende la correlación, se puede pronosticar el valor que adquirirá la cantidad en cuestión. Como ya se ha mencionado, es necesario realizar una gráfica de los datos para observar la tendencia. Una tendencia lineal es un caso particular de correlación donde la variable dependiente (Y) depende de una variable independiente (X), de la manera: $Y=a+bX$. Como toda proyección, la limitación que tiene que asumir esta tendencia es que se pronostica que todos los datos futuros recaerán sobre una línea.

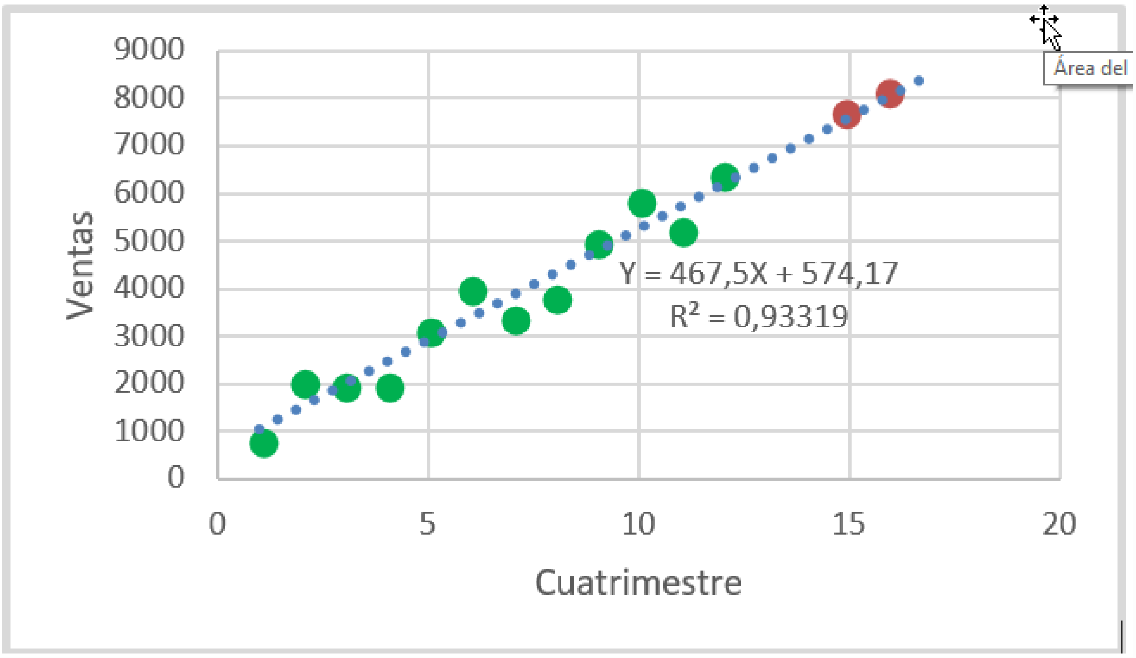
Excel presenta la posibilidad de realizar la regresión lineal o incorporar la línea de tendencia en una gráfica. El resultado se muestra en la Figura 6, donde se grafican los datos de la Tabla 6 y se ha incorporado la línea de tendencia, si asumimos un comportamiento lineal. En esta acción, Excel posibilita obtener la pendiente *b*, la ordenada al origen *a*, y cuánto se aproximan los datos a un comportamiento lineal, dado por el valor R^2, que es el coeficiente de determinación. Cuando R^2 es cercano a 1, los valores presentan un comportamiento lineal. Si R^2 tiende a 0, la serie temporal no se ajusta a una función lineal.

Tabla 6: Ventas cuatrimestrales de un producto de una empresa

Cuatrimestre	Ventas	Cuatrimestre	Ventas
1	780	7	3380
2	2015	8	3770
3	1950	9	4940
4	1950	10	5850
5	3120	11	5200
6	4030	12	6370

Fuente: elaboración propia.

Figura 6: Gráfica temporal de los datos de la Tabla 6



Fuente: elaboración propia.

La tendencia lineal mostrada por la línea de puntos es el resultado del ajuste de los datos a la función lineal $Y = 467{,}5X + 574{,}17$. En este caso, $R^2 = 0{,}93319$, lo cual indica que hay una buena dependencia lineal de los datos. Con base en este resultado, ¿cómo realizar una predicción? Esto se hace al asignarle el valor de X al tiempo deseado para evaluar.

$$Y_{x=15} = 467{,}5 \times 15 + 574{,}17 = 7587$$

$$Y_{x=16} = 467{,}5 \times 16 + 574{,}17 = 8054$$

Estos nuevos valores ahora pueden incluirse en la gráfica (puntos rojos). Este pronóstico, por ejemplo, es útil para que la empresa se asegure la provisión adecuada del producto que vaya a vender o los insumos para producirlos y así poder llegar a esa meta de venta.

Tendencias no lineales (regresión no lineal)

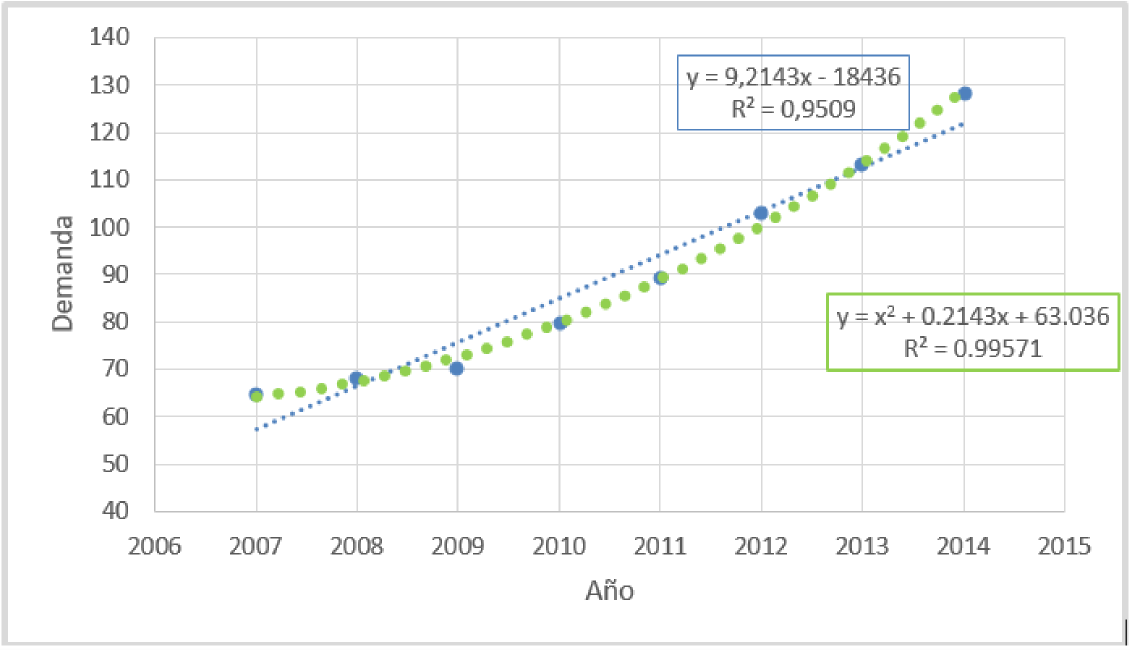
Muchas veces el patrón de comportamiento de los datos a lo largo del tiempo no es lineal. Existen otras funciones para interpretar variaciones no lineales. Un ejemplo es la dependencia cuadrática de los datos. De esta manera, se pueden obtener series temporales, como las mostradas en la Tabla 7, representadas gráficamente en la Figura 7. Se puede apreciar que, a simple vista, una tendencia lineal puede ser ajustada y obtenerse un valor de $R^2 = 0{,}95$. Pero, cuando se ajusta a una función cuadrática del tipo $Y = a + X^2$, ahora R^2 es más cercano a 1 (0,99). Cabe destacar que, si bien en el eje horizontal se muestran los años, el ajuste se realizó por períodos: el año 2007 es el período 1, y el 2017, el 11.

Tabla 7: Demanda anual un producto de una empresa

Año	Demanda
2007	65
2008	68
2009	70
2010	80
2011	89
2012	103
2013	113
2014	128
2015	145
2016	161
2017	185

Fuente: elaboración propia.

Figura 7: Dependencia cuadrática



Fuente: elaboración propia.

1.2.3 Tendencia y las estaciones

Estacionalidad sin tendencia

En la unidad anterior, se adelantó el significado de la dependencia estacional. Por ejemplo, si se estudia cómo es la venta de camperas en una tienda de vestimentas a lo largo de 5 años, se pueden obtener datos y gráficas, como se muestra en la Tabla 8 y en la Figura 8. En ambas se ve claramente cómo las ventas de este producto dependen de la época del año, motivo por el cual es mayor en estaciones frías (invierno y otoño) y menor en estaciones cálidas (verano y primavera).

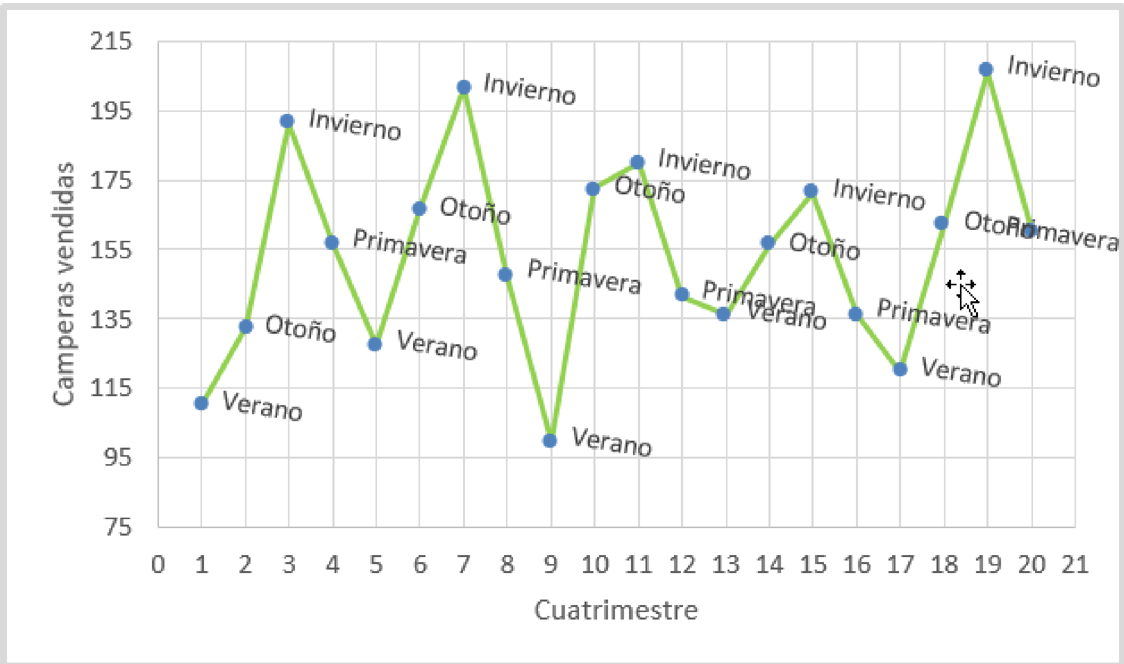
Tabla 8: Venta de camperas en una tienda de vestimentas. Estacionalidad

Año	período	Estación	Ventas
1	1	Verano	110
	2	Otoño	132
	3	Invierno	191
	4	Primavera	156
2	5	Verano	127
	6	Otoño	166
	7	Invierno	201

	8	Primavera	147
3	9	Verano	100
	10	Otoño	172
	11	Invierno	180
	12	Primavera	141
4	13	Verano	136
	14	Otoño	156
	15	Invierno	171
	16	Primavera	136
5	17	Verano	120
	18	Otoño	162
	19	Invierno	206
	20	Primavera	160

Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

Figura 8: Venta de camperas en una tienda de vestimentas. Estacionalidad



Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

Sin embargo, no se observa una tendencia clara de aumento o descenso de las ventas a lo largo del tiempo con un promedio de ventas que oscila entre los 154 abrigos.

Ahora bien, ¿cómo establecer un pronóstico fundamentado en los datos presentados? Una manera es contemplar, en este caso, el promedio de ventas para cada estación y luego realizar una combinación lineal que contemple, en proporción, cada una de las estaciones. De esta manera, las ventas promedio serían:

$$Verano: \frac{110 + 127 + 100 + 136 + 120}{4} = 119$$
$$Otoño: \frac{132 + 166 + 172 + 156 + 162}{4} = 158$$
$$Invierno: \frac{191 + 201 + 180 + 171 + 206}{4} = 190$$
$$Primavera: \frac{156 + 147 + 141 + 136 + 160}{4} = 148$$

Excel permite realizar regresiones múltiples mediante el complemento de análisis de datos. Para esto se puede generar una tabla como la siguiente:

Tabla 9: Promedio de ventas estacional

Estación	Promedio	Período	Período	Período
Verano	119	1	0	0
Otoño	158	0	2	0
Invierno	190	0	0	3
Primavera	148	0	0	0

Fuente: elaboración propia a base de Anderson, Sweeney & Williams, 2011.

Aquí se puede observar que se signó como valor de variable independiente un valor a cada estación: 1 para verano, 2 para otoño, 3 para invierno y 0 para primavera. Desde la tabla, se selecciona como valores de la variable dependiente (Y) a la columna promedio y como rango de valores de variable independiente (X) a las columnas período. Esto arroja una combinación lineal del tipo:

$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$, donde b_0 es la ordenada al origen.

Si aplicamos esta regresión, se obtiene la función: $Y = 148 - 29X_1 + 5X_2 + 14X_3$. Así corroboramos que se puede predecir, por ejemplo, para el período invierno (3), la cantidad de ventas. De este modo, se debe reemplazar en la ecuación por el número de período asignado:

$$Y = 148 - 29(0) + 5(0) + 14(3) = 190$$

1.2.4 Descomposición de series temporales

La *descomposición de series temporales* es un método clásico en pronósticos relacionados a los negocios y la economía. Este método asume que las series analizadas están compuestas por un atributo estacional (E), uno irregular (I) y otro de tendencia (T). La pregunta es en qué extensión cada uno de estos componentes impacta sobre el valor o variable estudiado en el tiempo. Cada uno de ellos puede contribuir de una forma aditiva o multiplicativa. En esta sección introduciremos la descomposición aditiva. De esta manera se puede representar como:

$$Y_t = E_t + I_t + T_t$$

Esto implica poder identificar y valorar cada uno de estos componentes para comprender en qué proporción se deben combinar para que determinen el valor que se deba predecir. En general, esto se realiza analizando cada componente por separado.

Impacto de la tendencia

Hay varias maneras de estimar en una serie de datos cuánto contribuye la componente tendencia. Una manera simple es calcular la media móvil de un período determinado. Si la serie temporal no presenta la componente estacional, se puede calcular como:

$$T_{(t+1)} = \frac{Y_{t-1} + Y_t \ldots + Y_{t-k+1}}{k}$$

Si la serie temporal sí muestra una componente estacional, el largo del promedio móvil (k) debe ser de la frecuencia de la estación. Por ejemplo, se pueden centrar dos promedios móviles en el año tomando una ventana de tiempo de 6 meses:

$$T_{(t)} = \frac{\frac{1}{2}Y_{t-5} + Y_{t-4} \dots + Y_{t+5} + \frac{1}{2}Y_{t+6}}{12}$$

Impacto de la estacionalidad

La contribución de la estacionalidad se puede calcular al restarle al valor observado la tendencia. De esta manera, se puede representar:

$$E_t = Y_t - T_t$$

Realizando esto, se obtiene una estimación del efecto en la observable para cada mes.

Referencias

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2011). *Statistics for Business and Economics* (11th ed.). Mason, OH: Cengage Learning.

Frases go. (s. f.). Frase de Futuro de John Naisbitt 48084. Recuperado de http://www.frasesgo.com/frase/frase-de-john_naisbitt-48084.html

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2013). *Introduction to Probability and Statistics*. Boston, MA, USA: Brooks/Cole, Cengage Learning.