Series temporales

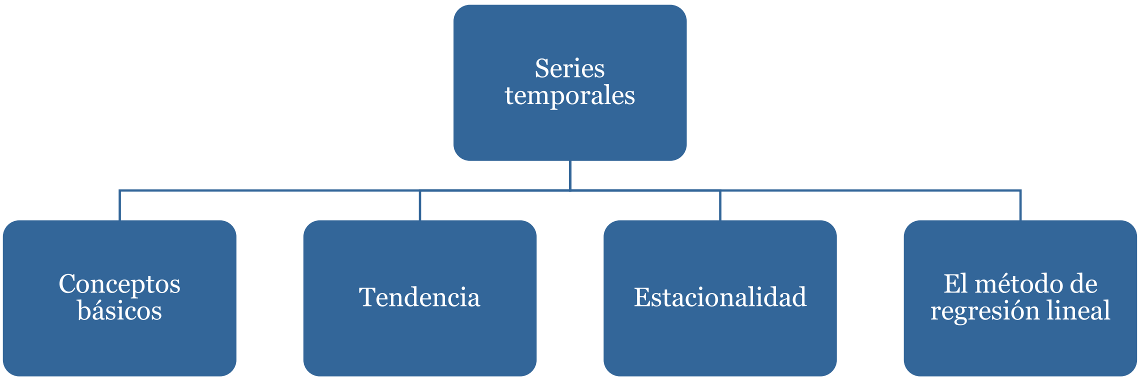
TEMA 11

Ideas clave

11.1. ¿Cómo estudiar este tema?

En este tema se van a introducir los conceptos claves de series temporales de forma muy breve para tener unas básicas nociones de las Ideas clave que rodean esta idea.

Para estudiar este tema **deberás comprender las Ideas clave** expuestas en este documento y que han sido elaboradas por el profesor de la asignatura. Estas ideas se van a complementar con lecturas y otros documentos para que puedas ampliar los conocimientos sobre el mismo.



11.2. Introducción a las series temporales

Hasta ahora todas las variables que hemos estudiado tenían en común que no estaban ligadas a un instante de tiempo concreto. Sin embargo, es muy frecuente que ciertas magnitudes estén ligadas a diferentes momentos temporales para ver su evolución a lo largo del tiempo, como pueden ser por ejemplo el IPC o la toma de la velocidad del viento en una determinada zona. Al tomar medidas de la misma variable en serie de momentos en el tiempo, da lugar a una serie temporal.

Una **serie temporal** o de tiempo es una sucesión de observaciones cuantitativas de un fenómeno ordenadas en el tiempo.

El análisis del comportamiento de las series temporales de los datos pasados o de los datos futuros requiere de una serie de técnicas diferentes de las vistas hasta este momento, ya que, aunque las técnicas empleadas hasta ahora son aplicables, no son óptimas para analizar la información que aportan este tipo de datos.

Las series temporales son variables **bidimensionales**, pues una variable son los instantes de tiempo y la otra variable es la propia de la serie temporal. La variable de tiempo pueden ser instantes de tiempo o pueden ser intervalos de tiempo. Cuando hablamos de instantes de tiempo se dice que tenemos magnitudes stockes, mientas que si hablamos de intervalos de tiempo se dice que tenemos flujo.

Por ejemplo, supongamos que recogemos el número de empleados de una empresa mensualmente. En este caso, estamos tratando una magnitud stock, mientras que si lo que estamos observando es el número de ventas de una empresa en un mes tenemos una variable flujo. La diferencia entre una y otra es que el número de empleados entre un mes y el sucesivo no se pueden sumar, ya que da lugar a duplicidad en los datos, mientras que en el segundo ejemplo, el volumen de ventas si es sumable. Para poder manejar los datos en este tipo de series es necesario que se dé una **homogeneidad en los datos**, esto es, que los intervalos de tiempo sean comparables, que tengan la misma dimensión para poder trabajar con los datos.

Cuando tenemos intervalos de tiempo de distinta dimensión, cambia la población de referencia o la metodología de las observaciones, el resultado es una serie de tiempo compuesta por una serie de valores no comparables porque son heterogéneos.

La falta de homogeneidad de una serie de tiempo puede perderse a lo largo del tiempo de forma natural, de manera que cuando las series son muy largas no se puede garantizar que los datos iniciales y finales sean comparables. Pero esto contradice el objetivo principal de la estadística que es detectar irregularidades y para eso se necesita una gran cantidad de datos, ya que de no ser así tales irregularidades pasan desapercibidas.

Lo que se pretende con una serie de tiempo es describir y predecir el comportamiento de un fenómeno que cambia en el tiempo. Este comportamiento puede provocar variaciones de dos tipos:

» **Variaciones evolutivas**. Cuando el valor medio de la serie cambia, no permanece fijo en el tiempo, son series con un comportamiento creciente o decreciente.

- **Variaciones estacionarias**. Cuando el valor medio de la serie no cambia, puede sufrir pequeñas modificaciones que provocan oscilaciones en torno a un valor fijo medio.

11.3. Tendencia, estacionalidad y otras fluctuaciones

La forma más sencilla de empezar a analizar una serie temporal es mediante su representación gráfica. Para representar gráficamente los datos de una serie temporal, en el eje horizontal (x) aparecerán los valores propios de la serie y en el eje vertical (y) los valores de tiempo.

Con la representación gráfica se pueden detectar las características más importantes de la serie: su **variación** a largo plazo, **amplitud** de oscilaciones, existencia de **ciclos**, presencia de **valores anómalos**, etc. A continuación, en la gráfica se ha representado una serie que recoge la evolución del paro en España con datos recogidos mensualmente durante cinco años.

El primer paso es la representación gráfica de la serie temporal. Sin embargo, para analizarla no podemos quedarnos en este paso, ya que, aunque visualmente podemos sacar unas primeras conclusiones, para obtener conclusiones definitivas sobre su comportamiento pasado y futuro es necesario aplicar otras técnicas. Por tanto, el objetivo del análisis de las series temporales es doble, ya que por una parte nos interesan las observaciones pasadas, tratando de analizar si los datos reflejan algún tipo de patrón de comportamiento, y por otro lado, nos interesa analizar si es posible predecir valores futuros.

Para poder estudiar la serie técnicamente hay que tener en cuenta que las series temporales están formadas por unas componentes que son: tendencia, estacionalidad, variaciones cíclicas y variaciones residuales. Esta descomposición de las series facilita su estudio pero es cierto que no todas las series presentan siempre estas cuatro componentes. Por ejemplo, cuando se trabaja con series anuales la serie no tiene la componente estacional, o cuando trabajamos con series que no están ligadas a aspectos económicos la serie puede no tener variaciones cíclicas.

La tendencia es el componente de la serie que recoge el comportamiento de la serie a largo plazo. Para detectar este comportamiento en la serie es necesario tener un número de observaciones elevado, a lo largo de muchos años, para detectar si la serie sigue un determinado patrón de crecimiento o decrecimiento, o presenta un comportamiento de estabilidad. Esta tendencia de crecimiento o decrecimiento en la serie puede ser de distinto tipo: lineal, exponencial, parabólico, etc. Al considerar movimientos a largo plazo nos olvidaremos de las variaciones que sufre la a corto plazo. En el ejemplo visto anteriormente del paro registrado en España, vemos que la serie tiene una tendencia decreciente.

Las **variaciones estacionales** son movimientos de la serie que se repiten periódicamente. La periodicidad normalmente suele ser el año, pero puede tratarse de periodicidad mensual, semanal o incluso diaria. En el ejemplo que hemos visto del paro se ve que aunque el paro tiene tendencia a decrecer anualmente vemos que hay periodos en los que desciende y luego aumenta un poco y se mantiene en valores más elevados. Si analizamos la serie vemos que se debe a que en los meses de verano el paro se reduce, mientras que cuando acaba esta temporada el paro aumenta y se mantiene casi constante, hasta la nueva temporada estival.

Las **variaciones cíclicas** son un componente que aparecen en las series que están relacionadas con aspectos económicos y están relacionadas con la sucesión de fases expansivas o recesivas de la economía. Son movimientos a plazo medio, periodos superiores al año, que se repiten de forma casi periódica, pero que no son tan regulares como el componente estacional. Este componente es difícil de aislar, ya que puede ocurrir que se pueden superponer ciclos de distinto periodo o amplitud. La amplitud es el número de años que dura un ciclo completo. En el ejemplo anterior no vemos un componente cíclico. Esto puede ser debido a que es necesario mayor número de observaciones, ya que con un periodo mayor al año no vemos una repetición clara de ningún patrón.

Las variaciones accidentales no responden a ningún patrón de comportamiento. Son el resultado de observaciones fortuitas que aparecen de forma aislada. Pueden ser debidas a fenómenos meteorológicos, huelgas, accidentes, etc. Normalmente suelen deberse a factores externos a las observaciones pero que repercuten en los mismos de forma indirecta.

La serie temporal viene dada por la interacción de estos cuatro componentes y viene reflejada en dos modelos: modelo aditivo y modelo multiplicativo.

- En el modelo **aditivo**, los valores observados son el resultado de la suma de los cuatro componentes.
- En el modelo **multiplicativo**, los valores observados son el resultado de la multiplicación de las cuatro componentes.
- También existe el modelo mixto, el cual se emplea cuando la variación aleatoria es independiente de las demás componentes, pero no nos interesa especialmente.

La técnica más acertada para basarnos en un modelo aditivo o multiplicativo de forma gráfica consiste en:

- Calcular las medias y desviaciones típicas de cada año y representar estas dos variables estadísticas en unos ejes de coordenadas (media, desviación típica).
- Si al representar los puntos obtenemos una serie de puntos crecientes, entonces supondremos la hipótesis de modelo multiplicativo y en caso contrario modelo aditivo.

11.4. El método de ajuste lineal

El método de ajuste lineal, como su nombre indica, tratará de ajustar la tendencia de la serie mediante una función lineal tal que minimice la suma de los errores cuadráticos, es decir, ajustar la tendencia con una función lineal, imponiendo para conocer los valores de la pendiente y la ordenada en el origen la condición que la suma cuadrática de los errores sea mínimo.

Si recordamos la teoría de la recta de regresión, se trata justamente de esto, el método de ajuste de la recta de regresión por el método de los mínimos cuadrados. Se trata de aplicar la recta de regresión para describir la tendencia de la serie.

El método se basa en:

$$y_t = t_t^* + e_t = at + b + e_t$$

y_t es la serie originaria de datos observados, la cual se descompone en y_t^* (tendencia) y e_t (otros componentes que aparecen de forma residual).

1.

Es importante señalar que si empleamos este método estamos suponiendo que se trata de un modelo aditivo, por lo que si en la representación gráfica de la media y dispersión vemos que la serie no cumple el supuesto aditivo, no es recomendable aplicar este método para ajustar la tendencia.

Para obtener la recta de la tendencia, cuando los datos de la serie temporal son anuales es inmediato. Aplicando mínimos cuadrados obtenemos la recta de regresión y calculamos los valores de la tendencia.

Sin embargo, cuando los datos no son anuales, si no vienen dados en subperiodos (trimestrales, cuatrimestrales, mensuales, etc.) hay que hacer lo siguiente:

1. Primero: calculamos las medias de los datos anuales.
2. Segundo: obtenemos las tendencias medias anuales ajustando la recta de regresión para esos valores medios anuales obtenidos.
3. Tercero: calculamos la tendencia de los subperiodos aplicando la siguiente fórmula:

$$y_t^* = \bar{y}_t{}^* + \left[i - \frac{k+1}{2}\right]\frac{a}{k}$$

- » t es el año.
- » i es el subíndice que indica el subperiodo, es decir, si tenemos datos trimestrales $i = 1, 2, 3, 4$, para primer trimestre, segundo, tercero y cuarto respectivamente, y así para cada subperiodo que tengamos.
- » k es el número total de subperiodos, si los datos son cuatrimestrales $k = 4$.
- » a es la pendiente de la recta de regresión.

2.

La obtención de la tendencia con el método de ajuste lineal presenta una gran ventaja frente a los métodos de ajuste anteriores. Esto se debe a que es válido para predecir valores futuros de la serie puesto que en la forma de representar los datos de la serie de la tendencia aparece explícitamente la variable de tiempo, y esto es una gran ventaja, pues es uno de los objetivos, como hemos mencionado al principio de este tema, del análisis de las series temporales.

Ejemplo

En la siguiente tabla se muestran las ventas trimestrales de una empresa en millones de euros. Queremos obtener la serie de tendencia aplicando el método de ajuste lineal:

Trimestres/Años	2006	2007	2008	2009	2010
Primer	1	2	2	3	5
Segundo	2	3	4	4	7
Tercer	4	5	5	7	8
Cuarto	3	4	3	6	7

El primer paso para emplear el método de ajuste lineal es calcular las medias anuales:

Trimestres/Años	2006	2007	2008	2009	2010
Primer	1	2	2	3	5
Segundo	2	3	4	4	7
Tercer	4	5	5	7	8
Cuarto	3	4	3	6	7
Medias anuales	2.5	3.5	3.5	5	6.75

Ahora la tendencia media anual, $\overline{y_t}$, se obtiene ajustando la recta de regresión: $\overline{y_t} = at + b$ y aplicando el método de los mínimos cuadrados resulta:

$$a = 1;\ b = -2003.75.$$

Recordemos como se calculaba la recta de regresión:

	t	$\bar{y}$	$(\bar{y}_t - \bar{\bar{y}}_t)$	$(\bar{y}_t - \bar{\bar{y}}_t)^2$	$(t - \bar{\bar{t}})$	$(t - \bar{t})^2$	$t\,\bar{y}_t$
	2006	2.5	-1.75	3.0625	-2	4	5015
	2007	3.5	-0.75	0.5625	-1	1	7024.5
	2008	3.5	-0.75	0.5625	0	0	7028
	2009	5	0.75	0.5625	1	1	10045
	2010	6.75	2.5	6.25	2	4	13567
Media	2008	4.25		11		10	42680

Recordemos que el coeficiente de regresión se calcula con esta fórmula:

$$\sigma_{xy} = \sum_{i,j} \frac{n_{ij} t_i y_j}{n} - \bar{t}\,\bar{\bar{y}} = \frac{42680}{5} - 2008 * 4.25 = 8536 - 8534 = 2$$

Las varianzas de t e y:

$$\sigma_t^2 = \sum_{i=1}^{5} \frac{\left(t - \bar{\bar{t}}\right)^2}{5} = \frac{10}{5} = 2$$

$$\sigma_y^2 = \sum_{i=1}^{5} \frac{(\bar{y}_i - \bar{\bar{y}})^2}{5} = \frac{10}{5} = 2$$

Por tanto los valores a y b de la recta de regresión son:

$$a = \frac{\sigma_{ty}}{\sigma_t^2} \quad y \quad b = \bar{\bar{y}} - \frac{\sigma_{ty}}{\sigma_t^2}\,\bar{t}$$

Que en nuestro ejemplo:

$$a = \frac{\sigma_{ty}}{\sigma_t^2} = \frac{2}{2} = 1 \qquad b = \bar{\bar{y}} - \frac{\sigma_{ty}}{\sigma_t^2}\,\bar{t} = 4.25 - \frac{2}{2}\,2008 = -2003.75$$

De este modo, con la recta $\overline{y_t^*} = -2003.75 + t$ obtenemos los valores de la tendencia anual media:

Años	2006	2007	2008	2009	2010
Medias anuales	2.5	3.5	3.5	5	6.75
Tendencia anual media	2.25	3.25	4.25	5.25	6.26

Ahora lo que queremos es calcular el valor de la tendencia en cada subperiodo, para obtener esto aplicamos la fórmula:

$$y_t^* = \bar{y_t}^* + \left[i - \frac{k+1}{2}\right]\frac{a}{k}$$

- a es la pendiente de la recta de regresión a=1.
- t es el año (2006, 2007, 2008, 2009, 2010).
- i es el subperiodo indicando por 1, 2, 3, 4 a los trimestres primero, segundo, tercero y cuarto respectivamente.
- k, es el número total de subperiodos en cada año (al ser trimestrales k=4).

3.

Por tanto aplicando esta fórmula para cada subperiodo tenemos la serie tendencia:

Tiempo	i	Serie	Tendencia
I 2006	1	1	1.875
II 2006	2	2	2.125
III 2006	3	4	2.375
IV 2006	4	3	2.625
I 2007	1	2	2.875
II 2007	2	3	3.125
III 2007	3	5	3.375
IV 2007	4	4	3.625
I 2008	1	2	3.875
II 2008	2	4	4.125
III 2008	3	5	4.375
IV 2008	4	3	4.625
I 2009	1	3	4.875
II 2009	2	4	5.125
III 2009	3	7	5.375
IV 2009	4	6	5.625
I 2010	1	5	5.875
II 2010	2	7	6.125
III 2010	3	8	6.375
IV 2010	4	7	6.625

El gráfico sería el siguiente:

Lo + recomendado

No dejes de leer…

Estadística descriptiva e inferencial. Contraste de hipótesis

Mauricio, J. A. (2007). Introducción al análisis de series temporales. Manuscrito inédito. Universidad Complutense de Madrid.

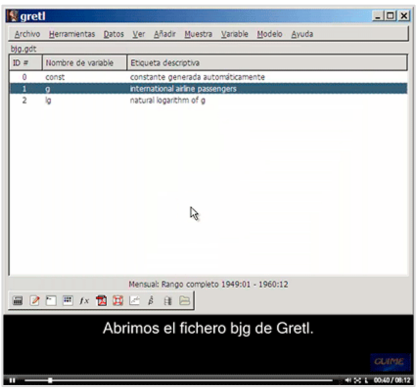
En este libro encontrarás toda la teoría relacionada con las series temporales y su análisis.

Accede al artículo a través del aula virtual o desde la siguiente dirección web:
https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf

No dejes de ver…

Series temporales ARIMA

Este vídeo supone un acercamiento al alumno hacia las series temporales.



Accede al vídeo a través del aula virtual o desde la siguiente dirección web:
http://www.ugr.es/~jchica/Pagina2/GUIME/Series%20temporales/aerolineas/aerolineas.html

+ Información

Webgrafía

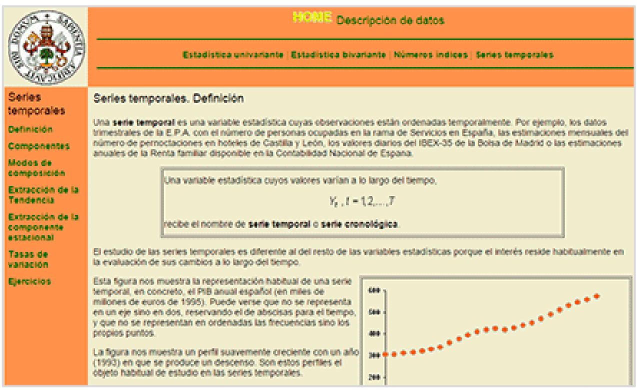
Análisis de series temporales

Esta página web tiene explicaciones y ejemplos sobre series temporales aplicadas.

Accede a la página web a través del aula virtual o desde la siguiente dirección:
http://www.seh-lelha.org/tseries.htm

Series temporales

Esta página web tiene explicaciones y ejemplos sobre series temporales aplicadas.



Accede a la página web a través del aula virtual o desde la siguiente dirección:
http://www.eco.uva.es/estadmed/datos/series/series.htm

Bibliografía

Aguirre, A. (1994). *Introducción al tratamiento series temporales: aplicaciones a las ciencias de la salud*. Madrid: Ediciones Díaz de Santos.

Test

1. El componente tendencia de una serie temporal:

A. Siempre es estacionaria.
B. Es el movimiento general a largo plazo de la serie.
C. Refleja la tendencia de los últimos datos de la serie.
D. Siempre es muy variable.

2. Las oscilaciones regulares de una serie temporal, de periodicidad igual o inferior al año, es el componente:

A. Variación estacional.
B. Variación cíclica.
C. Variación aleatoria.
D. Ninguna de las anteriores.

3. Las oscilaciones de las variaciones estacionales son causadas principalmente:

A. Por el momento del ciclo económico, principalmente.
B. Por las observaciones aleatorias.
C. Por el tiempo, por el clima o efectos del calendario.
D. Puede deberse a muchas causas adversas.

4. Generalmente, cuando se mantiene la amplitud de las oscilaciones a lo largo de la tendencia tenemos un esquema:

A. Aditivo.
B. Multiplicativo.
C. Mixto.
D. Solo podemos decir que la tendencia es estable.

5. Para predecir los valores futuros de una serie temporal:

A. Da igual la técnica.
B. No podemos utilizar ninguna técnica.
C. Con el ajuste lineal.
D. Ninguna de las respuestas anteriores es correcta.

6. El método de ajuste lineal trata de ajustar mediante:
   A. Una función cuadrática.
   B. Una función muy complicada.
   C. Una función lineal.
   D. Todas son ciertas.

7. Las oscilaciones periódicas, pero no regulares, en medio plazo, es la componente:
   A. Tendencia.
   B. Variación estacional.
   C. Variación cíclica.
   D. Variación aleatoria.

8. Generalmente, cuando la amplitud de las oscilaciones varía con la tendencia, aplicaríamos un modelo:
   A. Aditivo.
   B. Multiplicativo.
   C. Mixto.
   D. Mínimos cuadrados.

9. De la componente cíclica es cierto que:
   A. Sean variaciones que se aprecian en situaciones relacionadas con la economía.
   B. Se aprecian con situaciones de cambios climatológicos.
   C. Son variaciones a corto plazo.
   D. Siempre aparecen.

10. Un método que nos pueden servir mucho a la hora de las series temporales:
   A. La gráfica.
   B. La comparación de medias.
   C. La comparación de varianzas.
   D. Todas son correctas.