



## Análisis de series temporales

Preparado por Luis M. Molinero (Alce Ingeniería)

CorreoE: [estadistica@alceingenieria.net](mailto:estadistica@alceingenieria.net)

Enero 2004

 [Artículo en formato PDF](#)

 [www.seh-lelha.org/stat1.htm](http://www.seh-lelha.org/stat1.htm)

*"¿Será posible que la belleza fuera la clave para comprender este mundo? El escultor griego Policleteo de Argos, que vivió hacia el 450–420 a.C., estableció las bases de nuestra comprensión actual de las partículas fundamentales cuando en su Canon, su guía estética, dijo que la belleza aparece poco a poco, mediante muchos números. Policleteo escribió sobre simetría, el equilibrio dinámico de las partes relajadas y tensas del cuerpo humano y las orientaciones relativas de dichas partes, que tienen como resultado un todo armonioso. Dos mil quinientos años después recurrimos a los aspectos matemáticos de la simetría –y los aspectos simétricos de la matemática– para organizar nuestra comprensión de las entidades fundamentales con las que se esculpe la materia y del equilibrio dinámico de las fuerzas que las mantienen unidas. Siempre que por belleza entendamos simetría y la pérdida controlada de ésta, Mondrian se convierte en Monet; luego, sin duda, la belleza ocupa el centro del mundo. Parte de esta belleza está expuesta a la apreciación inmediata, como cuando contemplamos un diseño agradable; pero otra parte está muy escondida y no resulta evidente para el ojo inculto. Los miles de años transcurridos desde Policleteo se han empleado en desenterrar esta belleza oculta, vaciando su valoración en la forma matemática para después emplear las herramientas matemáticas en la investigación pormenorizada del paisaje de la realidad. A medida que ha avanzado la ciencia, ha aumentado su profundidad y alcance con la mayor abstracción de sus conceptos. En ningún otro lugar es más elaborada esta transición que en el descubrimiento de la simetría y su empleo como instrumento de comprensión.*

*Peter Atkins. El dedo de Galileo: Las diez grandes ideas de la ciencia. Ed. Espasa Calpe, 2003*

## Introducción

Cuando hablamos de una secuencia de valores observados a lo largo del tiempo, y por tanto ordenados cronológicamente, la denominamos, en un sentido amplio, serie temporal. Resulta difícil imaginar una rama de la ciencia en la que no aparezcan datos que puedan ser considerados como series temporales.

Si, conocidos los valores pasados de la serie, no fuera posible predecir con total certeza el próximo valor de la variable, decimos que la serie es no determinista o aleatoria, y lógicamente es de éstas de las que se ocupa el cuerpo de doctrina denominado "análisis de series temporales" y al que vamos a dedicar esta breve introducción.

A pesar de que en medicina no sea muy conocido, el análisis estadístico de series temporales se usa hoy día con profusión en muchas otras áreas de la ciencia, fundamentalmente en física, ingeniería y en economía.

Los objetivos del análisis de series temporales son diversos, pudiendo destacar la predicción, el control de un proceso, la simulación de procesos, y la generación de nuevas teorías físicas o biológicas.

Denominamos **predicción** a la estimación de valores futuros de la variable en función del comportamiento pasado de la serie. Este objetivo se emplea ampliamente en el campo de la ingeniería y de la economía, incluyendo en esta última rama también la sanidad pública y la vigilancia de la salud. Así por ejemplo, la predicción mediante modelos basados en la teoría de series temporales, puede servir para una buena planificación de recursos sanitarios, en función de la demanda que se espera en el futuro, prevista por el modelo. Otro de los campos en los que se aplica la predicción mediante series temporales es el de la meteorología o en la predicción de otros fenómenos naturales.

En la teoría de **control de procesos**, se trata de seguir la evolución de una variable determinada con el fin de regular su resultado. Esta teoría se utiliza en medicina en los Centros de Control de Enfermedades.

La **simulación** se emplea en investigación aplicada, cuando el proceso es muy complejo para ser estudiado de forma analítica.

Evidentemente aunque el valor futuro de una serie temporal no sea predecible con total exactitud, para que tenga interés su estudio, el resultado tampoco puede ser completamente aleatorio, existiendo alguna regularidad en cuanto a su comportamiento en el tiempo, lo que hará posible su modelado y por ende, en su caso, la predicción. La búsqueda de regularidades y de patrones ha sido siempre una de las tareas básicas de la ciencia, y muchas veces se descubren simetrías que sirven de fundamento para la predicción del comportamiento de los fenómenos, incluso antes de que se entienda la razón o causa que justifica esa regularidad. Esto ocurrió, por ejemplo, con el sistema periódico de los elementos, descrito por Mendeleiev (1834–1907), quien organizó de forma muy correcta los elementos químicos en base a las simetrías observadas entre ellos, antes de que se comprendiese la razón de esas simetrías o periodicidad, razones que luego se fundamentaron sobre todo en trabajos de Schrödinger (1887–1961) y Pauli (1900–1958).

Por lo tanto, si podemos encontrar patrones de regularidad en diferentes secciones de una serie temporal, podremos también describirlas mediante modelos basados en distribuciones de probabilidad. La secuencia ordenada de variables aleatorias  $X(t)$  y su distribución de probabilidad asociada, se denomina **proceso estocástico**. Un proceso estocástico es por tanto el modelo matemático para una serie temporal.

Un concepto importante que encontramos en este ámbito, es el de procesos estacionarios. Si examinamos por ejemplo la temperatura para un determinado mes a lo largo de los años en una determinada zona geográfica, y se está produciendo un cambio climático, aunque haya fluctuaciones, habrá una tendencia creciente. De una manera informal, diremos que una serie es **estacionaria** cuando se encuentra en equilibrio estadístico, en el sentido de que sus propiedades no varían a lo largo del tiempo, y por lo tanto no pueden existir tendencias. Un proceso es **no-estacionario** si sus propiedades varían con el tiempo, como el clima.

Vamos ahora a presentar tres enfoques diferentes, aunque relacionados, para el análisis de series temporales.

### Modelado clásico de series temporales

El primer paso obligatorio para analizar una serie temporal es presentar un gráfico de la evolución de la variable a lo largo del tiempo, como puede ser el de la figura:

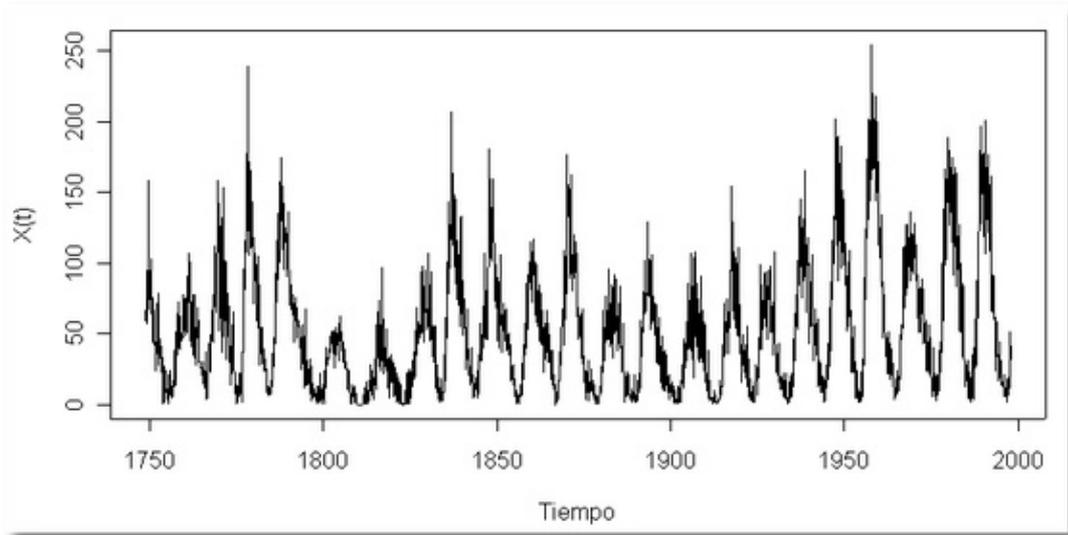


Fig. 1 Representación de una serie temporal

El siguiente paso consistirá en determinar si la secuencia de valores es completamente aleatoria o si, por el contrario, se puede encontrar algún patrón a lo largo del tiempo, pues sólo en este caso podremos seguir con el análisis.

La metodología tradicional para el estudio de series temporales es bastante sencilla de comprender, y fundamentalmente se basa en descomponer las series en varias partes: tendencia, variación estacional o periódica, y otras fluctuaciones irregulares.

- **Tendencia.** Es la dirección general de la variable en el periodo de observación, es decir el cambio a largo plazo de la media de la serie.
- **Estacionalidad.** Corresponde a fluctuaciones periódicas de la variable, en periodos relativamente cortos de tiempo.
- **Otras fluctuaciones irregulares.** Después de extraer de la serie la tendencia y variaciones cíclicas, nos quedará una serie de valores residuales, que pueden ser o no totalmente aleatorios. Volvemos a estar como en el punto de partida, pues ahora también nos interesa determinar si esa secuencia temporal de valores residuales puede o no ser considerada como aleatoria pura.

En la figura 2 vemos un ejemplo de una serie temporal en la que se aprecia la existencia de las distintas componentes comentadas

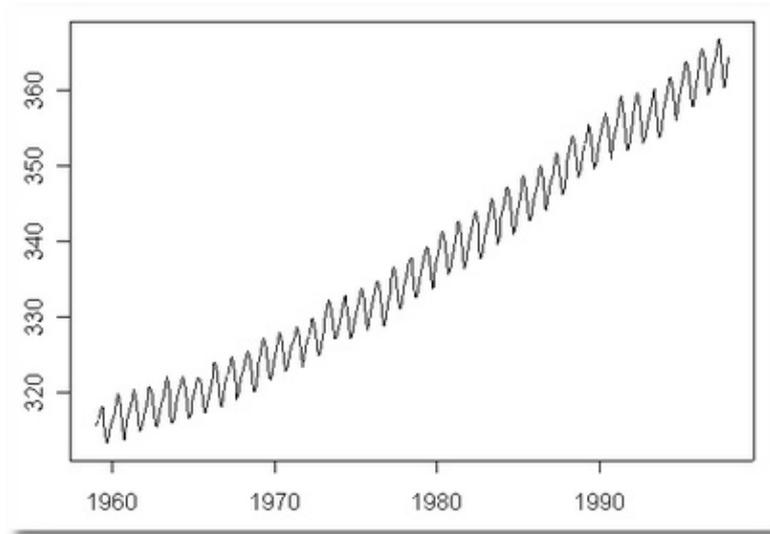


Fig. 2 Serie temporal con tendencia

### Análisis de la tendencia

Una primera idea sobre la presencia de tendencia en la serie la obtendremos en su representación gráfica. Pero no siempre estará tan clara como en la figura 2. Por ejemplo, en la siguiente imagen sigue habiendo tendencia pero ya no es tan marcada.

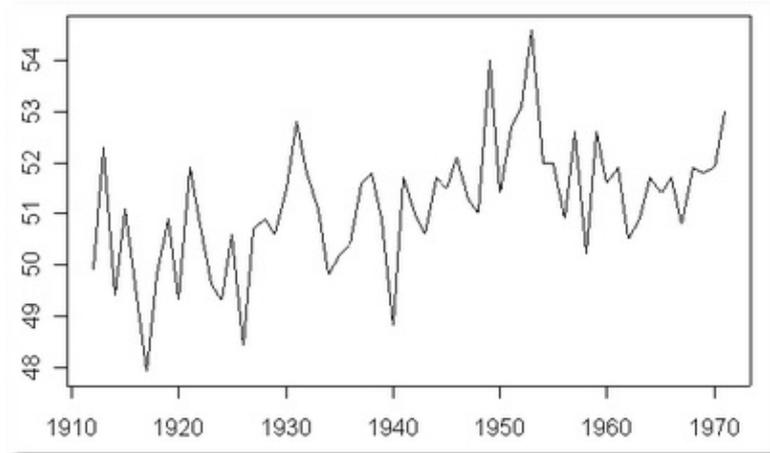


Fig. 3 Otra serie temporal con tendencia (menos pronunciada)

Los medios más utilizados para detectar y eliminar la **tendencia** de una serie se basan en la aplicación de **filtros** a los datos. Un filtro no es más que una función matemática que aplicada a los valores de la serie produce una nueva serie con unas características determinadas. Entre esos filtros encontramos las **medias móviles**.

Una media móvil se calcula, para cada punto, como un promedio del mismo número de valores a cada lado de ese punto. Así una media móvil de tres puntos se calcula como:

$$m(x_t) = \frac{x_{t-1} + x_t + x_{t+1}}{3}$$

Mientras que una media móvil de cuatro puntos viene dada por

$$m(x_t) = \frac{(x_{t-2} / 2) + x_{t-1} + x_t + x_{t+1} + (x_{t+2} / 2)}{4}$$

Cuando la cantidad de puntos de la media móvil es par, se toma la mitad de los valores extremos.

Existen otros procedimientos para extraer la tendencia, como **ajuste de polinomios**, **alisado mediante funciones exponenciales**, etc. Una clase de filtro, que es particularmente útil para eliminar la tendencia, se basa en aplicar **diferencias** a la serie hasta convertirla en estacionaria. Una diferencia de primer orden se obtiene restando dos valores contiguos:

$$\nabla x_{t+1} = x_{t+1} - x_t$$

Si volvemos a diferenciar esa serie, restando los nuevos valores consecutivos obtenemos una nueva serie más suavizada.

$$\nabla^2 x_{t+2} = \nabla x_{t+2} - \nabla x_{t+1}$$

Una vez que se aplica un proceso clásico de descomposición mediante un procedimiento de medias móviles a los datos de la figura 2, se obtiene las siguientes series:

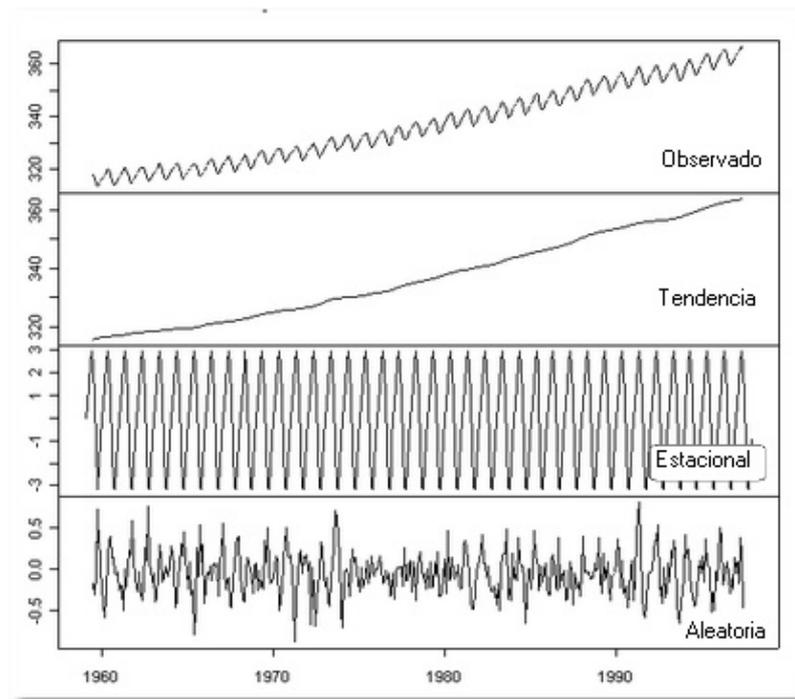


Fig. 4 Descomposición de una serie temporal en sus componentes

Para analizar la **estacionalidad** de una serie introduciremos un concepto de gran interés en el análisis de series temporales: la **función de autocorrelación**.

La **función de autocorrelación** mide la correlación entre los valores de la serie distanciados un lapso de tiempo  $k$ .

Recordemos la fórmula del coeficiente de correlación simple, dados  $N$  pares de observaciones  $y$ ,  $x$ :

$$r = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (x_i - \bar{x})^2}}$$

De igual forma, dada una secuencia temporal de  $N$  observaciones  $x_1 \dots x_N$ , podemos formar  $N-1$  parejas de observaciones contiguas  $(x_1, x_2)$ ,  $(x_2, x_3)$ , ...  $(x_{N-1}, x_N)$  y calcular el coeficiente de correlación de estas parejas.

A este coeficiente lo denominaremos **coeficiente de autocorrelación** de orden 1 y lo denotamos como  $r_1$ . Análogamente se pueden formar parejas con puntos separados por una distancia 2, es decir  $(x_1, x_3)$ ,  $(x_2, x_4)$ , etc. y calcular el nuevo coeficiente de autocorrelación de orden 2. De forma general, si preparamos parejas con puntos separados una distancia  $k$ , calcularemos el coeficiente de autocorrelación de orden  $k$ .

Al igual que para el coeficiente de correlación lineal simple, se puede calcular un error estándar y por tanto un intervalo de confianza para el coeficiente de autocorrelación.

La **función de autocorrelación** es el conjunto de coeficientes de autocorrelación  $r_k$  desde 1 hasta un máximo que no puede exceder la mitad de los valores observados, y es de gran importancia para estudiar la estacionalidad de la serie, ya que si ésta existe, los valores separados entre sí por intervalos iguales al periodo estacional deben estar correlacionados de alguna forma. Es decir que el coeficiente de autocorrelación para un retardo igual al periodo estacional debe ser significativamente diferente de 0.

Relacionada con la función de autocorrelación nos encontramos con la **función de autocorrelación parcial**. En el coeficiente de autocorrelación parcial de orden  $k$ , se calcula la correlación entre parejas de valores separados esa distancia pero eliminando el efecto debido a la correlación producida por retardos anteriores a  $k$ .

En la figura 5 vemos una gráfica típica de la función de autocorrelación parcial, en la que se marcan los intervalos de confianza para ayudar a detectar los valores significativos y cuya posición en el eje X nos indicará la probable presencia de un factor de estacionalidad para ese valor de retardo.

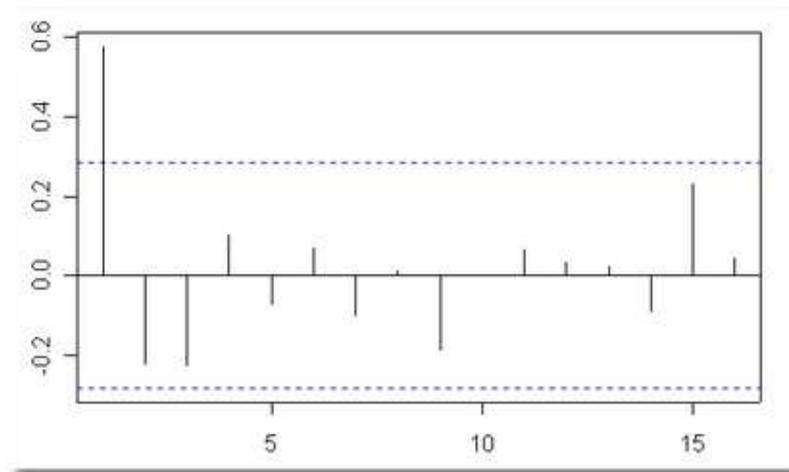


Fig. 5 Función de autocorrelación parcial

### **El enfoque moderno de series temporales: modelos ARIMA**

A comienzo de los años 70, G.E.P. Box, profesor de Estadística de la Universidad de Wisconsin, y G.M. Jenkins, profesor de Ingeniería de Sistemas de la Universidad de Lancaster, introdujeron una pequeña revolución en el enfoque del análisis de series temporales, en sus trabajos sobre el comportamiento de la contaminación en la bahía de San Francisco, con el propósito de establecer mejores mecanismos de pronóstico y control. El libro (1976) en el que describen la metodología, se convirtió rápidamente en un clásico, y sus procedimientos se utilizan ampliamente desde entonces en diferentes ramas de la ciencia, conociéndose como modelos ARIMA y también como **modelos Box-Jenkins**.

Para este tipo de modelos, el primer paso consiste en convertir nuestra serie de observaciones en una **serie estacionaria**, que es aquella en la que ni la media, ni la varianza, ni las autocorrelaciones dependen del tiempo. Una vez "estabilizada" la serie mediante las transformaciones adecuadas, se procede a estudiar la presencia de regularidades en la serie, para identificar un posible modelo matemático. Para ello se calculan la función de autocorrelación simple y parcial, y se compara su forma con un catálogo de patrones gráficos, que son típicos de los diferentes modelos propuestos, seleccionando el modelo que más se adecue a la forma de las

funciones de autocorrelación que hemos obtenido con nuestros datos.

Una vez elegida la forma del modelo, se estiman los coeficientes del mismo, y finalmente se procede a efectuar un análisis de los residuos (diferencia entre el valor realmente observado y el valor previsto por el modelo), con el fin de comprobar si el ajuste del modelo a nuestros datos es adecuado. Si no lo fuera repetimos el proceso buscando otros modelos.

Una vez determinado un modelo suficientemente válido, sobre la serie estacionaria, procedemos a deshacer la transformación inicialmente efectuada para estabilizar la serie, y ahora comprobamos si los pronósticos del modelo son adecuados con nuestros datos, volviendo a comenzar la búsqueda de otro modelo si no fuera el caso. Puede por tanto tratarse de un proceso iterativo de mejora del modelo.

En el modelo, cada valor tomado por la variable en un instante dado, está influido por los valores de la variable en momentos anteriores, y se expresa como una relación lineal, función de:

1. Valores recientes de la variable
2. Ruidos en valores recientes de la variable
3. Valores remotos de la variable
4. Ruidos en valores remotos de la variable

El esquema general del modelo es el siguiente:

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + Z_t + b_1 Z_{t-1} + \dots + b_q Z_{t-q}$$

que es la fórmula general de los modelos denominados **ARMA**. Está constituido por una combinación de  $p$  términos **AR** (proceso autorregresivo), y  $q$  términos **MA** (proceso de medias móviles). La parte AR modela la influencia de los valores anteriores de la serie ( $X_{t-1}$  hacia atrás), y la parte MA modela la influencia del ruido en valores anteriores de la serie ( $Z_{t-1}$  hacia atrás), junto con el término  $Z_t$  que corresponde al ruido esperado en el mismo momento  $t$  en el que se estima el nuevo valor de la variable  $X$ .

Una de las ventajas de estos modelos es su gran simplicidad (sumas de términos), frente a los modelos propuestos en la formulación clásica.

La letra **I** que aparece en el nombre del modelo completo –**ARIMA**–, corresponde al proceso último a realizar, una vez definido el tipo de modelo y estimados los coeficientes de éste, ya que entonces hay que restablecer las características originales de la serie de datos, que fue transformada para inducir estacionaridad. A ese proceso inverso se denomina en general **Integración** y aporta esa letra que completa el nombre.

## **Análisis del espectro de frecuencias**

Se demuestra que cualquier proceso periódico se puede modelar, con la precisión deseada, mediante series de términos de funciones senoidales (seno y coseno), lo que se conoce como **series de Fourier**, y se denomina **espectro** a la representación de las amplitudes, en el eje de las Y, que constituyen los diferentes términos de la serie para toda la gama de frecuencias (eje de las X).

El espectro es una herramienta fundamental para detectar estacionalidad en una serie y determinar su periodo. Como es de esperar, el espectro está íntimamente relacionado con la función de autocorrelación.

En la siguiente figura se muestra una imagen típica del espectro de frecuencias de una serie, en el que se representa en el eje de las Y la amplitud y en el de las X la frecuencia, y partiendo de la estimación directa del espectro a partir de los datos (esquina superior izquierda), se va refinando mediante procedimientos de alisado y nos permite en este caso detectar la presencia de un factor de periodicidad para la frecuencia en torno del valor 1.

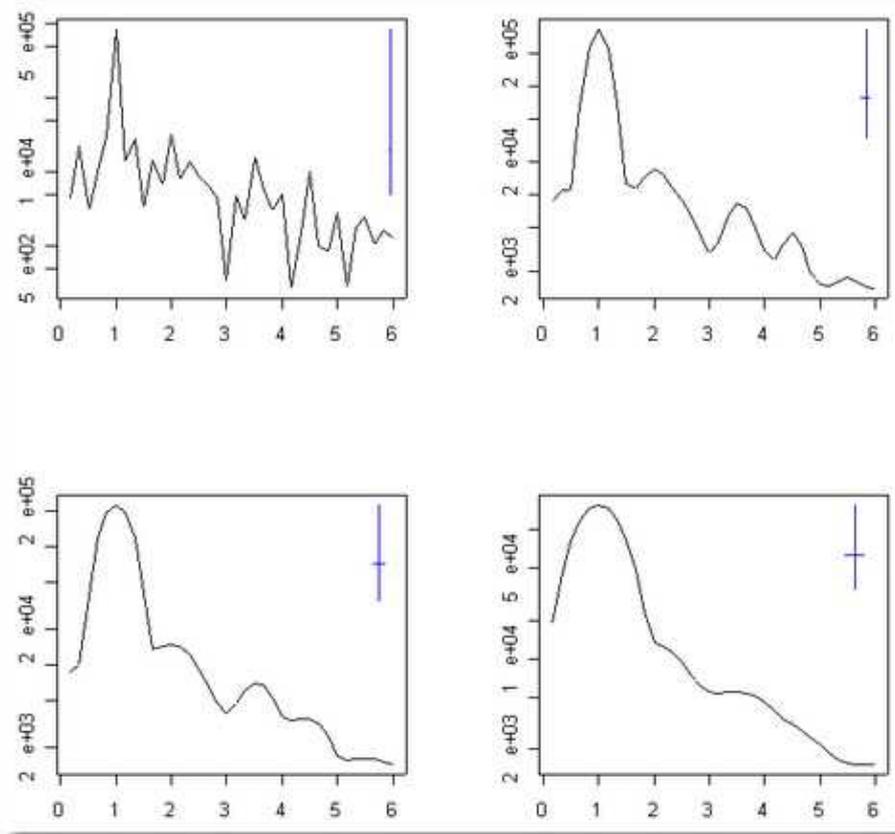


Fig. 6

Puesto que  $Periodo = 1/Frecuencia$ , obtenida la frecuencia, o frecuencias (picos en el espectro), a partir de ésta se calcula de forma sencilla el periodo de las oscilaciones en la serie de datos.

## Referencias

- Aguirre Jaime, Armando. **Introducción al tratamiento de series temporales. Aplicación a las Ciencias de la Salud**. Ed. Díaz de Santos, Madrid 1994
- Chatfield, C. **The Analysis of Time Series: An Introduction**. Ed. Chapman and Hall. London, 2003.

## Enlaces

- [Búsqueda en BMJ de artículos con el texto "time series" en el título o en el abstract](#)

 [Índice de artículos](#)

[Principio de la página](#) 