Series Temporales

Introducción

Una serie temporal se define como una colección de observaciones de una variable recogidas secuencialmente en el tiempo. Estas observaciones se suelen recoger en instantes de tiempo equiespaciados. Si los datos se recogen en instantes temporales de forma continua, se debe o bien digitalizar la serie, es decir, recoger sólo los valores en instantes de tiempo equiespaciados, o bien acumular los valores sobre intervalos de tiempo.

Ejemplos

Aparecen en numerosos campos. Ejemplos:

Economía y Marketing

– Precio del alquiler de pisos durante una serie de meses.

– Evolución del índice del precio del trigo con mediciones anuales.

– Beneficios netos mensuales de cierta entidad bancaria.

– Indices del precio del petróleo.

Demografía

– Número de habitantes en cierto país por año.

– Tasa de mortalidad infantil por año.

Medioambiente

– Evolución horaria de niveles de óxido de azufre y de niveles de óxido de nitrógeno en una ciudad durante una serie de años.

– Lluvia recogida diariamente en una localidad.

– Temperatura media mensual.

– Medición diaria del contenido en residuos tóxicos en un río.

La característica fundamental de las series temporales es que las observaciones sucesivas no son independientes entre sí, y el análisis debe llevarse a cabo teniendo en cuenta el orden temporal de las observaciones. Los métodos estadísticos basados en la independencia de las observaciones no son válidos para el análisis de series temporales porque las observaciones en un instante de tiempo dependen de los valores de la serie en el pasado.

Clasificaciones de las series temporales

Una serie temporal puede ser discreta o continua dependiendo de cómo sean las observaciones.

Si se pueden predecir exactamente los valores, se dice que las series son *determinísticas*.

Si el futuro sólo se puede determinar de modo parcial por las observaciones pasadas y no se pueden determinar exactamente, se considera que los futuros valores tienen una distribución de probabilidad que está condicionada a los valores pasados. Las series son así *estocásticas*.

Objetivos del análisis de series temporales

Se pueden considerar varios posibles objetivos:

1. **Descripción**

   Cuando se estudia una serie temporal, lo primero que se tiene que hacer es dibujarla y considerar las medidas descriptivas básicas. Así, se tiene que considerar:

   a) Si los datos presentan forma creciente (*tendencia*).

   b) Si existe influencia de ciertos periodos de cualquier unidad de tiempo (*estacionalidad*).

   c) Si aparecen outliers (observaciones extrañas o discordantes).

2. **Predicción**

   Cuando se observan los valores de una serie, se pretende normalmente no sólo explicar el pasado, sino también predecir el futuro.

Componentes de una serie temporal

El estudio descriptivo de series temporales se basa en la idea de descomponer la variación de una serie en varias componentes básicas. Este enfoque no siempre resulta ser el más adecuado, pero es interesante cuando en la serie se observa cierta tendencia o cierta periodicidad. Hay que resaltar que esta descomposición no es en general única.

Este enfoque descriptivo consiste en encontrar componentes que correspondan a una tendencia a largo plazo, un comportamiento estacional y una parte aleatoria.

Las componentes o fuentes de variación que se consideran habitualmente son las siguientes:

1. **Tendencia**: Se puede definir como un cambio a largo plazo que se produce en relación al nivel medio, o el cambio a largo plazo de la media. La tendencia se identifica con un movimiento suave de la serie a largo plazo.

2. **Efecto Estacional**: Muchas series temporales presentan cierta periodicidad o dicho de otro modo, variación de cierto periodo (anual, mensual ...). Por ejemplo, el paro laboral aumenta en general en invierno y disminuye en verano. Estos tipos de efectos son fáciles de entender y se pueden medir explícitamente o incluso se pueden eliminar del conjunto de los datos, *desestacionalizando* la serie original.

3. **Componente Aleatoria**: Una vez identificados los componentes anteriores y después de haberlos eliminado, persisten unos valores que son aleatorios. Se pretende estudiar qué tipo de comportamiento aleatorio presentan estos residuos, utilizando algún tipo de modelo probabilístico que los describa.

De las tres componentes reseñadas, las dos primeras son componentes determinísticas, mientras que la última es aleatoria. Así, se puede denotar que

$$X_t = T_t + E_t + I_t$$

donde T_t es la tendencia, E_t es la componente estacional, que constituyen la *señal* o parte determinística, e I_t es el *ruido* o parte aleatoria.

Es necesario aislar de alguna manera la componente aleatoria y estudiar qué modelo probabilístico es el más adecuado. Conocido éste, podremos conocer el comportamiento de la serie a largo plazo. Esto será motivo de estudio en Inferencia Estadística.

Este aislamiento de la componente aleatoria se suele abordar de dos maneras.

1. **Enfoque descriptivo**: Se estima T_t y E_t y se obtiene I_t como

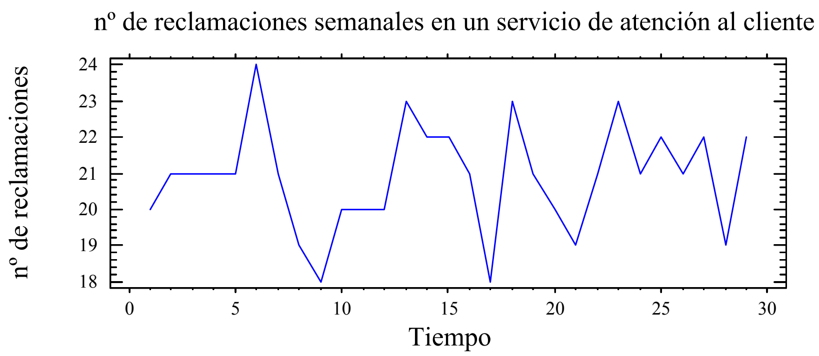
$$I_t = X_t - T_t - E_t$$

2. **Enfoque de Box-Jenkins**: Se elimina de X_t la tendencia y la parte estacional (mediante transformaciones o *filtros*) y queda sólo la parte probabilística. A esta última parte se le ajustan modelos paramétricos.

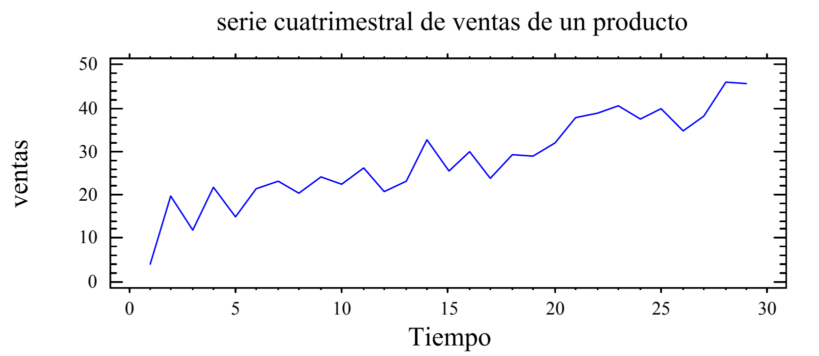
Análisis descriptivo de series temporales

La primera herramienta descriptiva básica es el gráfico temporal. Un gráfico temporal se construye situando los valores de la serie en el eje de ordenadas y los instantes temporales en el eje de abscisas. Construir este gráfico es de gran utilidad para observar el comportamiento de la serie temporal. Se presentan las siguientes series temporales.

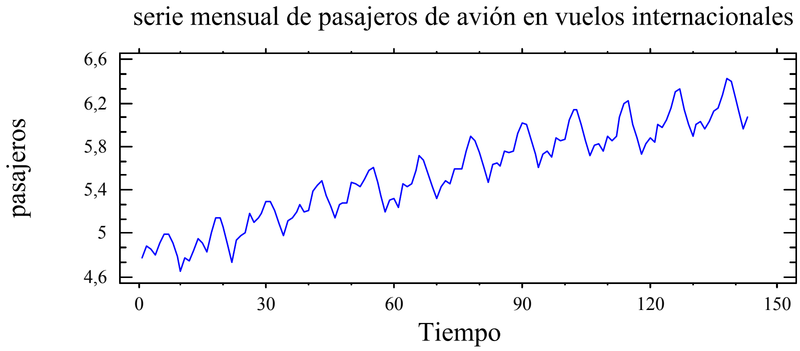
Ejemplo 1



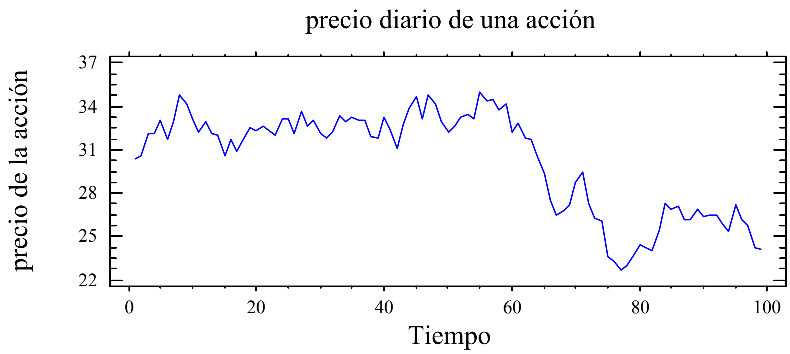
Ejemplo 2



Ejemplo 3



Ejemplo 4

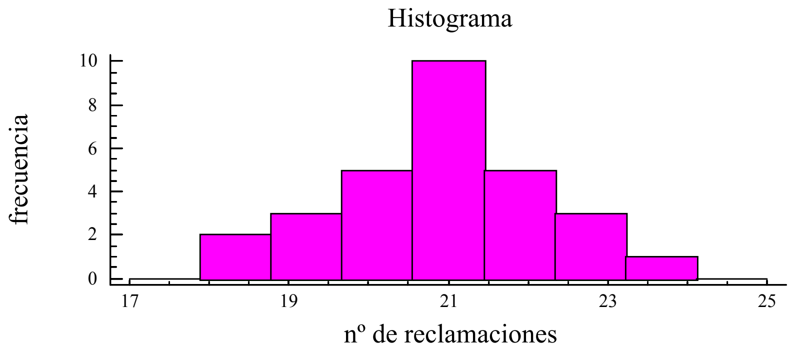


Clasificación descriptiva de las series temporales

Las series temporales se pueden clasificar en:

(i) **Estacionarias**: Una serie es estacionaria cuando es estable, es decir, cuando la media y la variabilidad son constantes a lo largo del tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo. Es una serie básicamente estable a lo largo del tiempo, sin que se aprecien aumentos o disminuciones sistemáticos de sus valores. Para este tipo de series tiene sentido conceptos como la media y la varianza. Sin embargo, también es posible aplicar los mismos métodos a series no estacionarias si se transforman previamente en estacionarias.

En el Ejemplo 1 se presenta una serie estacionaria discreta. La serie es estable alrededor de un valor central. Si representamos un histograma de esta serie, podemos describir adecuadamente la información: en promedio, se reciben unas 21 reclamaciones semanales. Este número es bastante estable y la distribución de la variable es aproximadamente simétrica. La mejor predicción para el próximo valor de la serie es la media, aunque lo ideal sería aplicar los modelos de Inferencia para series estacionarias que se presentarán más adelante.



(*ii*) **No Estacionarias**: Son series en las cuales la media y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

Por ejemplo, la serie del Ejemplo 2 presenta una fuerte tendencia creciente aunque existen importantes oscilaciones con relación a esa tendencia de crecimiento lineal. La serie del ejemplo 3 presenta además de de una tendencia creciente, una pauta estacional debido a que los pasajeros transportados en los meses de verano es mayor que en el resto del año. La serie del Ejemplo 4 muestra cambios de nivel y oscilaciones erráticas sin una pauta clara. La segunda mitad de la serie presenta una tendencia decreciente, estabilizándose en la parte final de la serie.

Estimación de la tendencia

Para estimar la tendencia supondremos que tenemos una serie no estacionaria sin componente estacional, es decir, que la serie se puede descomponer en

$$X_t = T_t + I_t$$

Las series de los Ejemplos 2 y 4 son de este tipo. Para estimar T_t debemos realizar alguna hipótesis sobre su forma. Vamos a analizar varios casos.

Tendencia determinista

En este caso supondremos que la tendencia es una función determinística. La función más sencilla posible es una recta, es decir,

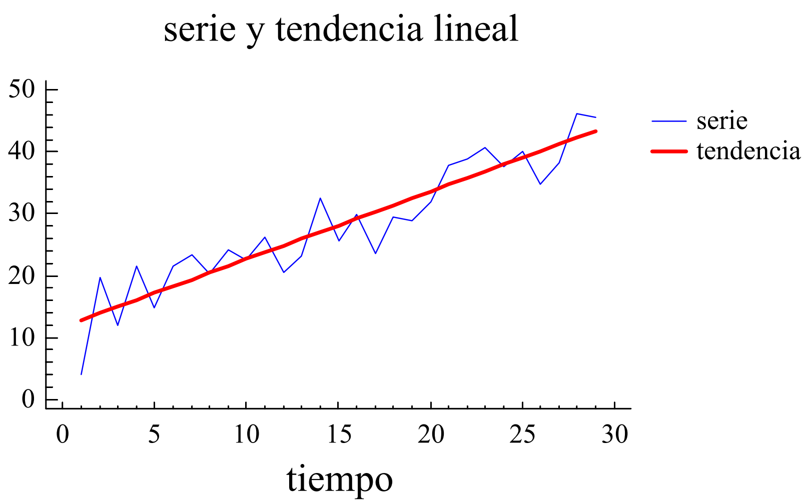
$$T_t = a + bt$$

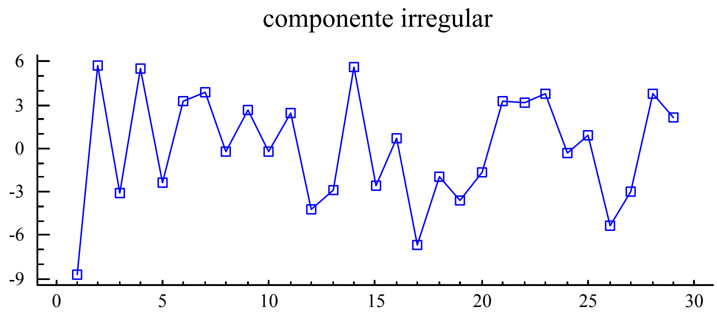
donde a y b son dos constantes a determinar. La forma de estimar estas constantes es mediante un modelo de regresión lineal entre las variables X_t y el tiempo $t = 1, 2, 3, \ldots$ De esta forma, si estimamos los parámetros $\widehat{a}$ y $\widehat{b}$, entonces la componente irregular será

$$I_t = X_t - \widehat{a} - \widehat{b}t$$

Ahora I_t sería una serie estacionaria que tendríamos que modelizar.

El Ejemplo 2 presenta una serie sin estacionalidad que presenta una tendencia que se podría expresar de forma lineal. La tendencia de la serie y la componente irregular serían, en este ejemplo,





Lo que nos queda es una serie estacionaria. En algunos casos, como en el Ejemplo 4, no es posible ajustar la tendencia mediante una recta. En estos casos, lo mejor sería ajustar la tendencia a un polinomio o a la curva que mejor se pueda ajustar. Para ello tendríamos que ajustar una regresión no lineal. Otra opción es describir la tendencia de manera evolutiva o *diferenciar* la serie.

Tendencia evolutiva (medias móviles)

Se supone que la tendencia es una función que evoluciona lentamente y que puede aproximarse en intervalos muy cortos (por ejemplo de 3 ó 5 datos) por una función simple del tiempo. En general se supone una recta, pero ahora sus coeficientes van cambiando suavemente en el tiempo.

Suponemos que la representación de la tendencia por una recta es válida para tres períodos consecutivos de tiempo, $t-1$, t, $t+1$, y representamos las tendencias en los tres periodos consecutivos de la siguiente manera:

$$\begin{aligned} T_{t-1} &= T_t - crecimiento \\ & T_t \\ T_{t+1} &= T_t + crecimiento \end{aligned}$$

Si hacemos la media de tres observaciones consecutivas

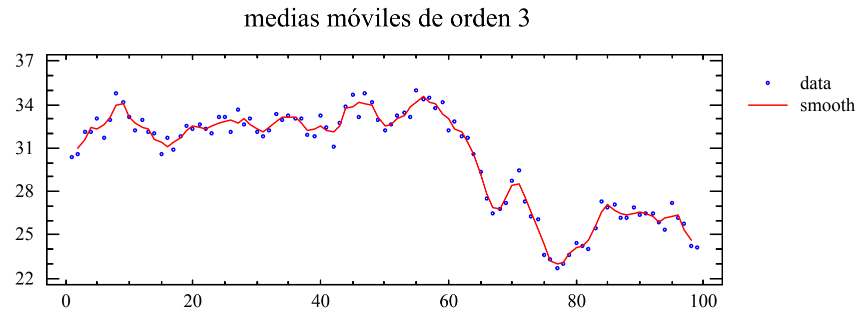
$$m_t = \frac{x_{t-1} + x_t + x_{t+1}}{3}$$

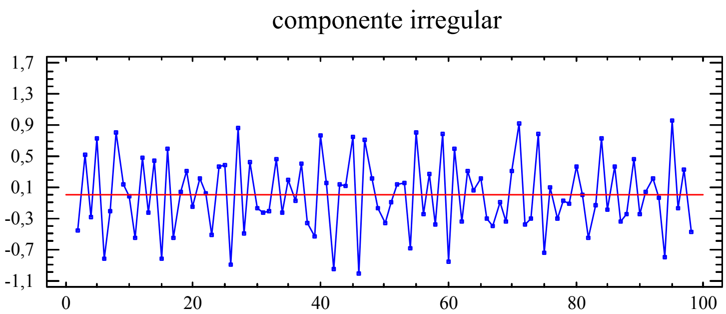
entonces

$$m_t = T_t + \frac{I_{t-1} + I_t + I_{t+1}}{3}$$

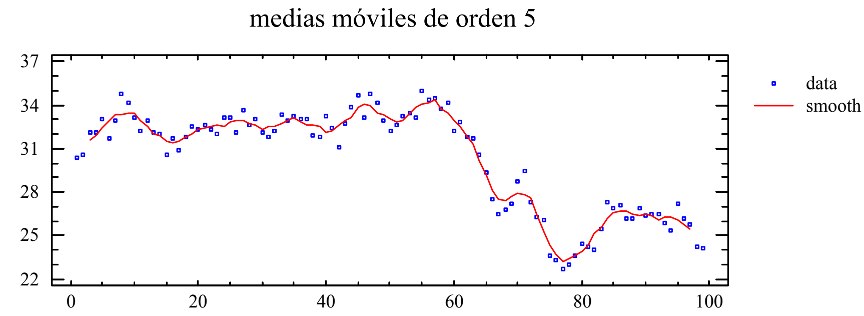
y como la componente irregular tiene media cero, la media de los tres valores del componente irregular se puede suponer que es despreciable frente a la tendencia, y m_t representa la tendencia en ese instante. Esta operación se denomina media móvil de orden tres. Se observa que realizando esta operación se pierde la primera observación y la última. Si calculamos las medias móviles de orden 5, perderemos las dos primeras observaciones y las dos últimas.

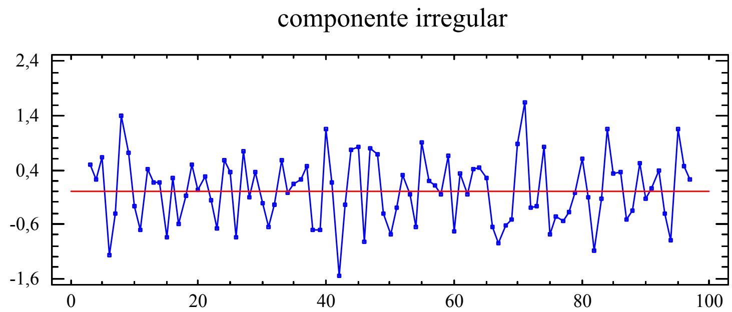
Aplicando este método a la serie del Ejemplo 4, las medias móviles de orden con la correspondiente componente irregular son:





mientras que para el ajuste a una media móvil de orden 5 queda:





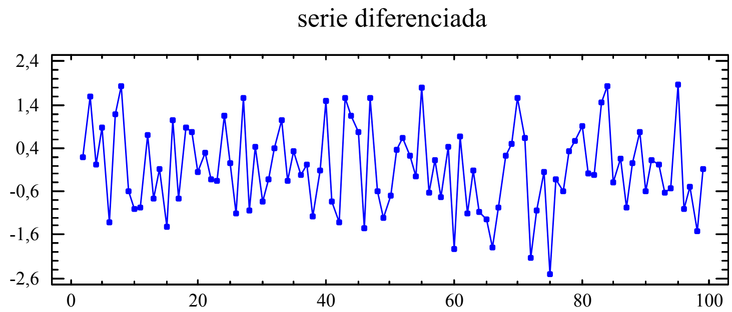
Diferenciación de la serie

Un tercer método más general para eliminar la tendencia consiste en suponer que la tendencia evoluciona lentamente en el tiempo, de manera que en el instante t la tendencia debe estar próxima a la tendencia en el instante $t-1$. De esta forma, si restamos a cada valor de la serie el valor anterior, la serie resultante estará aproximadamente libre de tendencia. Esta operación se denomina *diferenciación* de la serie y consiste en pasar de la serie original x_t a la serie y_t mediante:

$$y_t = x_t - x_{t-1}$$

De este modo, la serie diferenciada resulta ser estacionaria.

La serie del Ejemplo 4 diferenciada queda como



Estimación de la estacionalidad

Vamos a eliminar de la serie la componente estacional, es decir, se desestacionaliza la serie mediante los últimos 6 años de la serie del Ejemplo 3.

Esta serie presenta una estacionalidad mensual, de modo que se colocan los datos en una tabla de doble entrada:

	90	**91**	**92**	**93**	**94**	**95**	**medias**	**coef. est.**
Enero	5.49	5.65	5.75	5.83	5.89	6.03	5.77	-0.14
Febrero	5.45	5.62	5.71	5.76	5.83	5.97	5.72	-0.19
Marzo	5.59	5.76	5.87	5.89	6.01	6.04	5.86	-0.05
Abril	5.59	5.75	5.85	5.85	5.98	6.13	5.86	-0.05
Mayo	5.60	5.76	5.87	5.89	6.04	6.16	5.89	-0.02
Junio	5.75	5.92	6.05	6.08	6.16	6.28	6.04	0.13
Julio	5.90	6.02	6.14	6.20	6.31	6.43	6.17	0.26
Agosto	5.85	6.00	6.15	6.22	6.33	6.41	6.16	0.25
Septiembre	5.74	5.87	6.00	6.00	6.14	6.23	6.00	0.09
Octubre	5.61	5.72	5.85	5.88	6.01	6.13	5.87	-0.04
Noviembre	5.47	5.60	5.72	5.74	5.89	5.97	5.73	-0.18
Diciembre	5.63	5.72	5.82	5.82	6.00	6.07	5.84	-0.07

En este ejemplo hay efecto estacional mensual y, por tanto, existen 12 coeficientes estacionales, uno para cada mes del año. Para estimarlos, se calcula primero la media de las observaciones para cada mes, $M_1, \ldots, M_{12}$, y el coeficiente estacional resulta ser:

$$S_i = M_i - M \quad \text{para } 1 = 1, \ldots, 12.$$

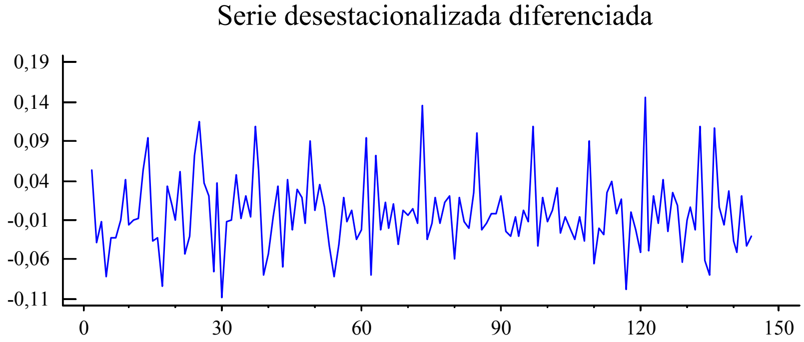
donde M es la media total de las observaciones. Así se puede observar que los meses con observaciones más pequeñas (por debajo de la media general) tienen coeficientes estacionales negativos, mientras que los meses con observaciones mayores tienen coeficientes estacionales positivos.

En el ejemplo anterior, la media total de las observaciones es $M = 5{,}91$, y así $S_1 = M_1 - M = 5{,}77 - 5{,}91 = -0{,}14$, $S_2 = M_2 - M = 5{,}72 - 5{,}91 = -0{,}19, \ldots$

Evidentemente, la suma de los coeficientes estacionales tiene que ser cero. Las temporadas más bajas son las correspondientes a los meses de febrero y noviembre, y las más altas las de los meses de julio y agosto.

Se denomina serie *desestacionalizada* a una serie donde se ha eliminado el efecto de cada mes y que se obtiene restando al valor de cada mes el coeficiente estacional de dicho mes. En nuestro ejemplo, con todos los datos, la serie desestacionalizada es:

Comprobamos que esta serie ya no tiene estacionalidad, pero todavía tiene tendencia. Para eliminar la tendencia procedemos a aplicar los procedimientos de la sección anterior a la serie desestacionalizada. Por ejemplo, diferenciando esta serie obtenemos:



Sin embargo esta serie vuelve a ser estacional, esto es, el problema es que al diferenciar ha aparecido otra vez un efecto estacional. Para explicar esto, vamos a analizar algunas observaciones de la serie $x_1, x_2, \cdots, x_{13,}x_{14}\cdots$

Al desestacionalizar la serie, hemos realizado la siguiente transformación

$$\begin{aligned} y_1 &= x_1 - E_1 \\ y_2 &= x_2 - E_2 \\ &\dots \\ y_{13} &= x_{13} - E_1 \\ y_{14} &= x_{14} - E_2 \\ &\dots \end{aligned}$$

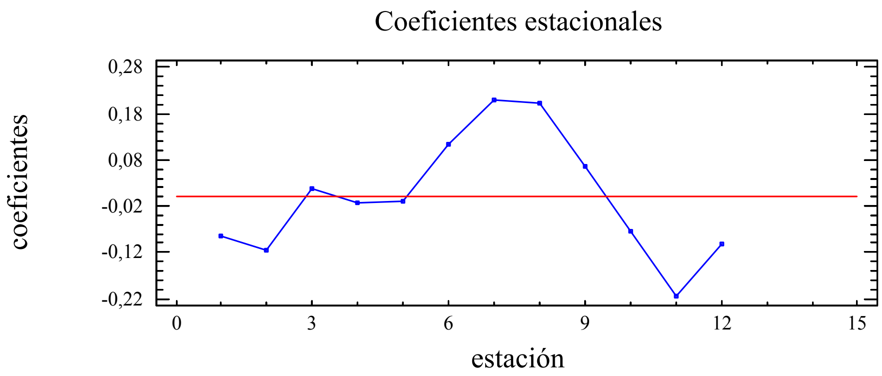
Ahora, al diferenciar la serie obtenemos

$$\begin{aligned} z_2 &= y_2 - y_1 = x_2 - x_1 - E_2 + E_1 = x_2 - x_1 - (E_2 - E_1) \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ z_{14} &= y_{14} - y_{13} = x_{14} - x_{13} - E_2 + E_1 = x_{14} - x_{13} - (E_2 - E_1) \end{aligned}$$

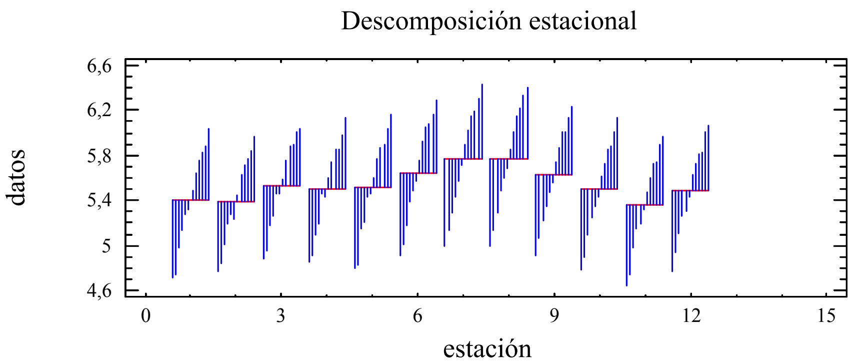
es decir, aparece en la serie un coeficiente estacional: $(E_2 - E_1)$.

Para solucionar este problema, se puede eliminar la tendencia de la serie desestacionalizada mediante los métodos descritos en las secciones anteriores. Otra solución, que es la más apropiada, es eliminar primero la tendencia de la serie y en segundo lugar desestacionalizar la serie.

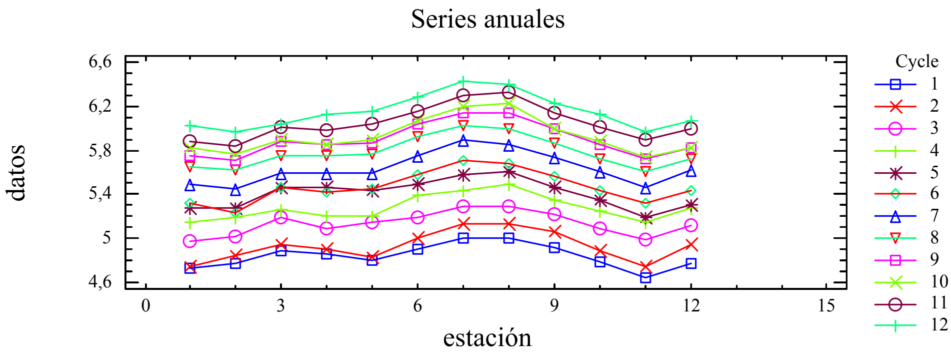
Para comprobar que la serie tiene componente estacional, se puede comprobar a partir del gráfico temporal de la serie. Statgraphics permite obtener gráficos como los que se muestran a continuación para la serie del Ejemplo 3. El siguiente gráfico muestra los coeficientes estacionales de la serie. Como se puede comprobar éstos son mayores para los meses de julio y agosto, y son menores para los meses de diciembre y febrero, lo cual indica que hay estacionalidad.



El siguiente gráfico muestra la descomposición estacional de la serie. Las líneas horizontales para cada mes muestran la media de la serie para cada uno de los meses (estaciones). Las líneas verticales que salen de cada línea horizontal indican, en cada mes, cómo varía la serie en los diferentes años. Por ejemplo, en el mes de enero, en los primeros años viajaron más personas en vuelos internacionales que en los últimos años. Además, este comportamiento es el mismo en cada uno de los meses, lo cual indica claramente que la serie tiene una tendencia creciente.



El siguiente gráfico muestra la serie año a año. La variación de la serie en los distintos meses es muy parecida todos los años, lo cual indica que hay estacionalidad. Además, las gráficas de los doce años están en orden creciente: la del primer año está por debajo de la del segundo año, la del segundo año por debajo de la del tercer año, y así sucesivamente. Esto indica que la serie tiene una tendencia creciente.



A veces, una simple inspección del gráfico temporal es suficiente para observar las principales características de la serie, pero en caso de duda, algunos de los gráficos anteriores pueden ser de gran utilidad.