Estadística III - 3009137
Introducción al Análisis de Series de Tiempo

Nelfi González Alvarez

Profesora Asociada Departamento de Estadística

e-mail: ngonzale@unal.edu.co

Facultad de Ciencias, Universidad Nacional de Colombia Sede Medellín

UNIVERSIDAD NACIONAL DE COLOMBIA

Departamento de Estadística
Semestre 01 de 2025

Contenido I

Contenido

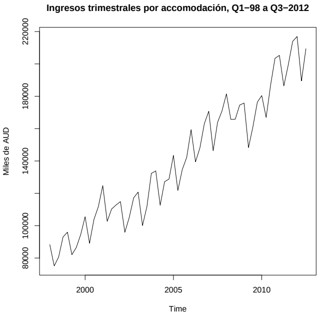
ANÁLISIS DE SERIES DE TIEMPO

Definición 1.1 (Serie de Tiempo)

Sucesión de valores al observar en el tiempo a una variable, con una frecuencia regular. Sea Y_t el valor de la variable observada en el tiempo t, $t \in \mathbf{Z}^+$ (índice de tiempo t).

Ingresos trimestrales por acomodación, en hoteles licenciados, con 15 o más habitaciones, en Victoria (Australia), trimestre 1 de 1998 al trimestre 3 de 2012 (miles de AUD)

Año	Trim. 1	Trim. 2	Trim. 3	Trim. 4
1998	88300.1	75110.4	80613.0	93020.3
1999	95972.2	82081.9	86481.3	94489.8
2000	105501.0	89134.9	103845.5	111762.6
2001	124810.7	102670.8	110291.7	112766.1
2002	114867.7	95953.7	105142.0	117245.5
2003	120720.1	100098.3	111805.9	132315.4
2004	133850.4	112594.7	127195.5	128819.4
2005	143454.0	121771.7	134677.0	142126.0
2006	159294.8	139485.9	148108.5	163054.3
2007	170744.8	146338.1	163909.5	170921.0
2008	181453.0	165696.9	165832.4	174547.8
2009	175799.8	148298.1	161120.9	176302.1
2010	180441.2	166905.0	186805.7	203461.7
2011	205270.1	186360.1	199326.4	213919.7
2012	216942.2	189515.3	209390.6	



Definición 1.2 (**Análisis de series de tiempo**)

Desarrollo de modelos estadísticos para explicar el comportamiento de una serie temporal, con el propósito de pronosticar, con fines de planeación y toma de decisiones.

Consideraremos modelos y métodos cuantitativos: Aplican cuando:

1. Se dispone de información histórica y cuantificable en forma de datos numéricos;
2. Puede asumirse que algunos aspectos de los patrones dados en el pasado continuarán en el futuro (***supuesto de continuidad***).

Los métodos cuantitativos de pronóstico se dividen en:

1. Métodos de series de tiempo (suavizamiento, descomposición, métodos Box - Jenkins);

Métodos de series de tiempo (suavizamiento, descomposición, Box - Jenkins).

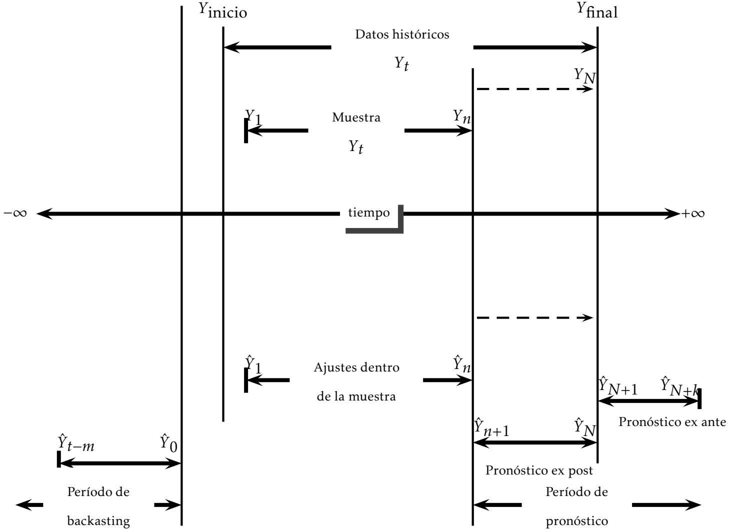
Modelos que explican Y_t vs. t: Analizan sucesión cronológica de observaciones de la variable de interés para la predicción del futuro bajo el ***supuesto de continuidad***.

2. Métodos causales o econométricos (regresión y correlación).

Métodos causales o econométricos (regresión y correlación)

Modelos que explican Y_t vs. $X_{1,t}\ldots,X_{p,t}$: Consideran una ***relación de causa-efecto*** entre la variable a ser pronosticada (variable dependiente) y una o más variables independientes o predictoras.

METODOLOGÍA DE LOS DATOS DE SERIES DE TIEMPO Y PRONÓSTICOS



Con esta metodología sobre los datos podemos comparar modelos en su calidad de ajuste y de pronósticos en los pronósticos ex- post **(ver Capítulo 4 de Notas de Clase: "Calidad de ajuste y de pronósticos")**:

En la muestra de ajuste

- Con los valores observados Y_t, $t = 1,\ldots,n$, en cada modelo se obtienen las estimaciones de la serie, $\widehat{Y}_t$;
- Luego con las diferencias $Y_t - \widehat{Y}_t$, $t = 1,\ldots,n$, se construyen medidas de ajuste como los criterios de información AIC y BIC (serán definidas luego).
- El mejor ajuste es con el modelo que minimice AIC y BIC.

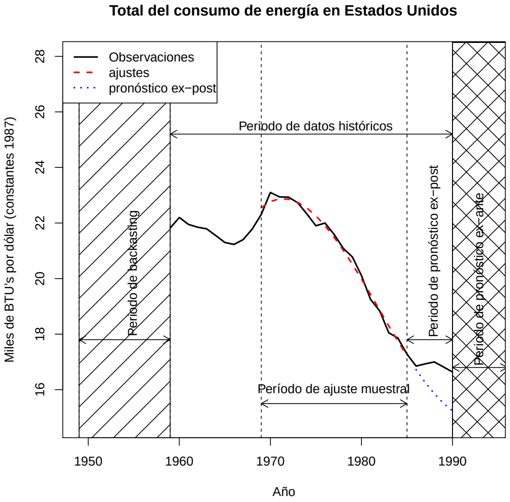
En la muestra de validación cruzada

- Con cada modelo estimado se obtienen los pronósticos de la serie para L periodos después de $t = n$, $\widehat{Y}_n(L) = \widehat{Y}_{n+L}$, con $L = 1,\ldots,m$;
- Con las diferencias $e_n(L) = Y_{n+L} - \widehat{Y}_n(L)$, $L = 1,\ldots,m$, conocidas como errores de pronóstico, donde Y_{n+L} es el valor observado de la serie en $t = n + L$, se construyen medidas de precisión promedio de los pronósticos puntuales, tales como el MAE, RMSE, MAPE de pronósticos (serán definidas luego).
- Dependiendo del modelo, se construirán intervalos de predicción (I.P).
- El mejor modelo en predicción minimiza MAE, RMSE, MAPE y con la mayor confiabilidad en la predicción por I.P (mayor cobertura y menor amplitud media en los I.P).

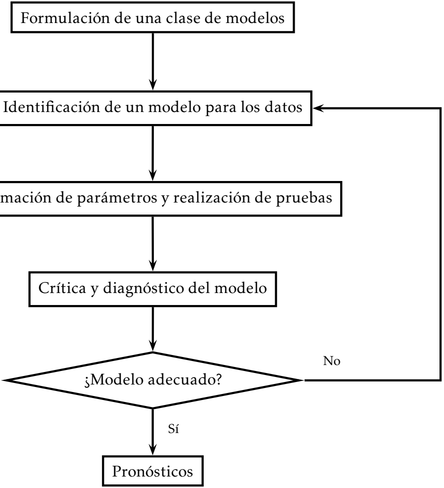
Nota 1.1

No siempre el mejor modelo en ajuste es el mejor modelo en la validación cruzada con los pronósticos ex-post.

EJEMPLO DEFINICIÓN DE LA LÍNEA DE TIEMPO



PASOS EN EL ANÁLISIS Y MODELACIÓN DE UNA SERIE TEMPORAL



El análisis descriptivo es el punto de partida para la formulación de los modelos.

Contenido

DESCRIPCIÓN DE UNA SERIE DE TIEMPO

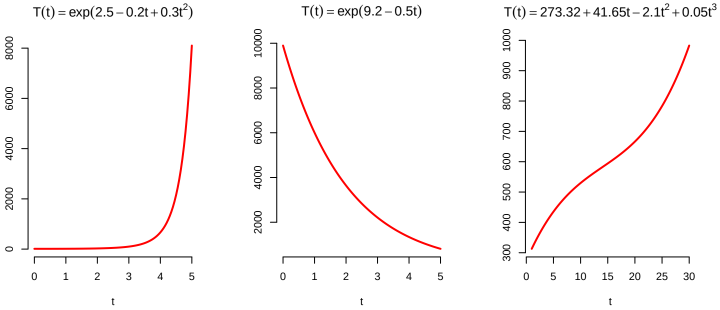
Componentes en la modelación de una serie:

1. Tendencia, T_t: *Patrón de largo plazo*
2. Estacionalidad, S_t: *Patrón dentro de un año calendario*
3. Ciclos, C_t: *Patrón de largo plazo*

Lo que no se logra explicar con las anteriores componentes, lo denominamos error E_t, *es la variación en el corto plazo.*

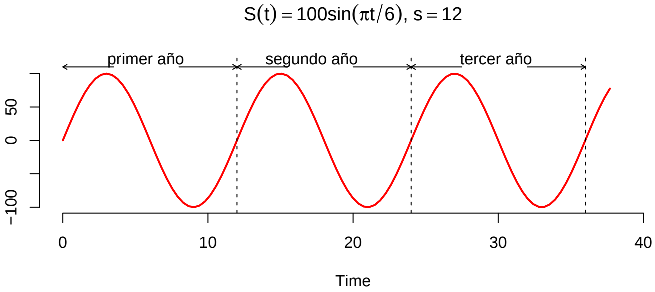
Componente de tendencia ($\mathbf{T_t}$)

Patrón de largo plazo caracterizado por la ***persistencia a un crecimiento o a un decrecimiento*** de los valores de la serie, por tanto refleja el crecimiento o declinación de la serie. ***Cuando es predecible*** se asocia a una ***función suave del tiempo,*** $T(t)$.



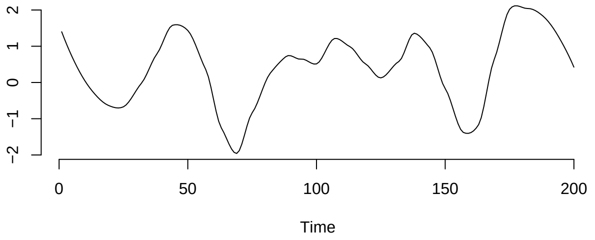
Componente estacional (S_t)

Patrón de ***cambio regular que se completa dentro de un año calendario y que se repite sobre una base anual.*** Es debido a factores tales como los hábitos de consumo o las estaciones climáticas. ***Cuando es predecible*** se asocia a una ***función periódica*** del tiempo de periodo s, $S(t) = S(t + m \times s)$, $m = \pm 1, \pm 2, \ldots$, con $s = 12$ en datos mensuales, $s = 4$ en datos trimestrales, etc.



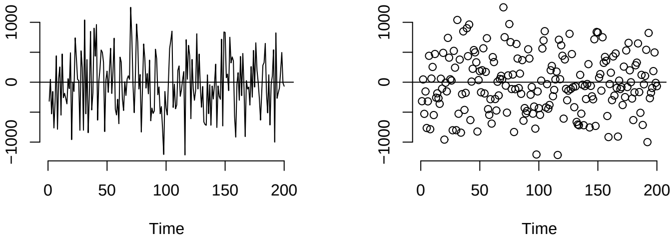
Componente cíclica (C_t)

Cambios o movimientos hacia arriba (prosperidad) y hacia abajo (recesión) con una duración de 2 o más años, debido a la influencia de fluctuaciones económicas de largo plazo como las asociadas a los ciclos de negocios. Los ciclos ***cambian la tendencia de los datos*** hacia arriba (**expansión**) a una tendencia hacia abajo (**contracción**). Los ciclos ***son las componentes más difíciles de pronosticar*** (no existe una función del tiempo para su modelación).



Componente irregular o error (E_t)

Fluctuaciones erráticas sin un patrón definido alrededor de una media constante. Muy a menudo estas fluctuaciones son debidas a eventos externos que sólo ocurren en un tiempo y de forma impredecible. Se dice de una serie que sólo exhibe tal comportamiento que es estacionaria en su media.



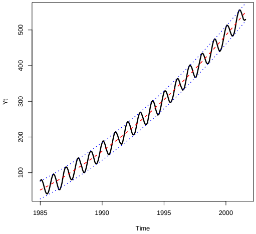
Contenido

TIPOS DE DESCOMPOSICIONES Y MODELOS

Una serie de tiempo puede ser una combinación de los cuatro patrones descritos, pero inicialmente modelamos ignorando los ciclos, así:

Aditiva

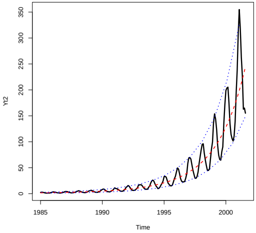
La serie es homocedástica alrededor de su tendencia



$$Y_t = T_t + S_t + E_t, \text{ con } E_t \overset{IID}{\sim} N(0, \sigma^2)$$

Multiplicativa

La serie es heterocedástica con varianza cambiando en la dirección de la tendencia:



- *Completamente multiplicativa*

$$Y_t = T_t \times S_t \times \exp(E_t), \text{ con } E_t \overset{IID}{\sim} N(0, \sigma^2).$$

- *Parcialmente multiplicativa*

$$Y_t = T_t \times S_t + E_t, \text{ con } E_t \overset{IID}{\sim} N(0, \sigma^2)$$

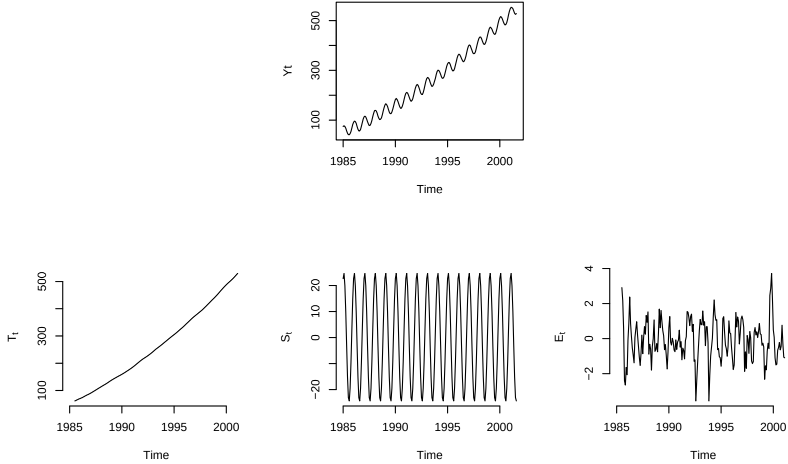
No es posible distinguir gráficamente cuál modelo multiplicativo considerar.

- Mediante un procedimiento conocido como filtro de descomposición clásica, es posible obtener una estimación no paramétrica de las componentes en el modelo de la descomposición aditiva y completamente multiplicativa. El procedimiento está implementado en R en la función `decompose()`.
- *Ver en Moodle, Capítulo 3 Notas de Clase, Sección 3.9: "Guía para el análisis de las componentes de una serie", el Apéndice A de ese capítulo sobre la función `decompose()`, y además el código R en el siguiente Script,*

▸ ProgramaREjemplosdeguiaparaanalisiscomponentesdeunaserie.R

Modelo de componentes aditivas

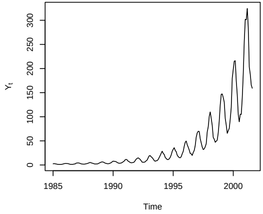
$$Y_t = T_t + S_t + E_t, \text{ con } E_t \overset{IID}{\sim} N(0,\sigma^2). \tag{1}$$

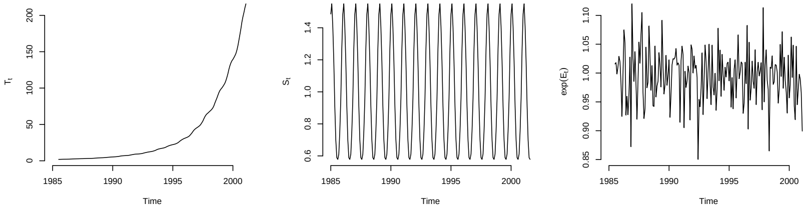


Modelo de componentes multiplicativas: Completamente multiplicativo

$$Y_t = T_t \times S_t \times \exp(E_t), \text{ con } E_t \overset{IID}{\sim} N(0, \sigma^2). \tag{2}$$

ir a ec. (4)

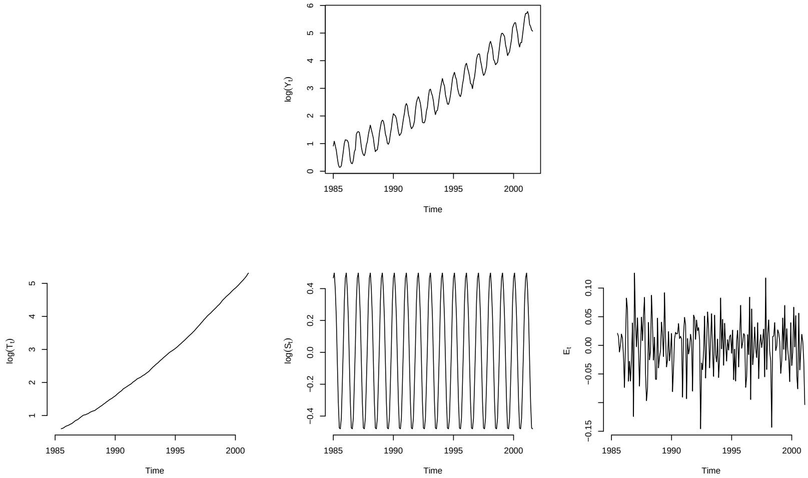




Modelo completamente multiplicativo transformado por logaritmo natural

$$\log(Y_t) = \log(T_t) + \log(S_t) + E_t, \text{ con } E_t \overset{IID}{\sim} N(0, \sigma^2) \tag{3a}$$

$$Y_t^* = T_t^* + S_t^* + E_t \quad (\text{decomp. aditiva para } \log(Y_t)) \tag{3b}$$



Modelo parcialmente multiplicativo

$$Y_t = T_t \times S_t + E_t, \text{ con } E_t \overset{IID}{\sim} N(0, \sigma^2) \tag{4}$$

La implementación de este modelo mediante regresión, implica:

1. un modelo no lineal en el vector de parámetros y no linealizable (por el contrario el caso completamente multiplicativo es linealizable usando transformación log());
2. su ajuste no es posible por mínimos cuadrados ordinarios, se requieren métodos de ajuste para modelos no lineales.

Tenga en cuenta que

Ambos modelos multiplicativos, (2) y (4), pueden ser considerados como candidatos para una misma serie de tiempo, pero observe y compare con cuidado la forma en que participa E_t en estos dos modelos. volver a ec. (2)

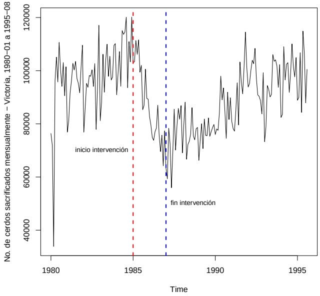
Nota

Cuando existen ciclos en una serie temporal pero son ignorados en un modelo de regresión, bien sea aditivo o multiplicativo, ***C_t se confunde con E_t***, razón por la cual ***el error del modelo no tendrá un patrón completamente al azar y mostrará algún tipo de dependencia*** (no habrá validez del supuesto de independencia entre los valores de esta componente).

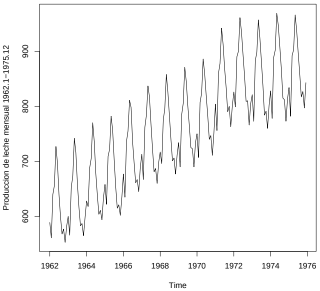
INTERVENCIONES Y CAMBIOS ESTRUCTURALES

Durante la evolución temporal, una serie puede ser afectada por ***intervenciones*** las cuáles causan ***cambios estructurales***. Observe las siguientes dos series:

Serie intervenida, tiene cambio estructural en la tendencia

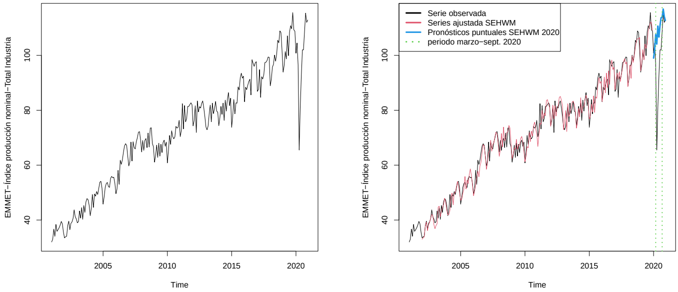


Serie con cierta estabilidad estructural



En presencia de cambios estructurales, los modelos que veremos pueden no tener un buen desempeño en el ajuste y/o en el pronóstico.

Vea por ejemplo el índice de producción nominal, Total industria, de enero de 2001 a diciembre 2020, año base 2018, Fuente DANE-EMMET:



(a) (b)

Figura 1: (a) la serie original; (b) la serie original, su ajuste de enero de 2001 a dic de 2019 y el pronóstico de enero a dic de 2020, mediante uno de los métodos que veremos en el curso: Suavizamiento Exponencial Holt-Winters Multiplicativo (SEHWM).

Si bien el modelo ajustado sigue relativamente bien los datos hasta 2019, los pronósticos para 2020 no son exitosos en vista de la intervención aplicada a raíz del COVID-19, como se puede ver mejor en la siguiente página.

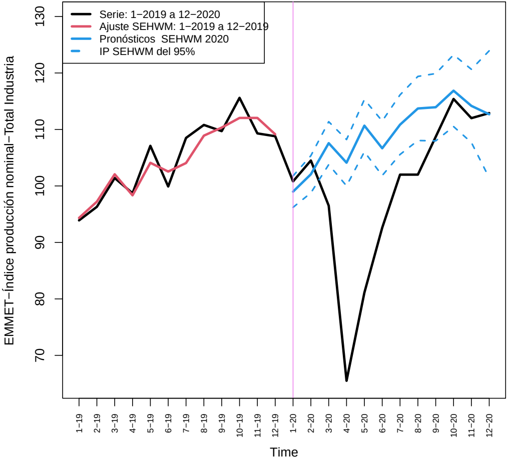
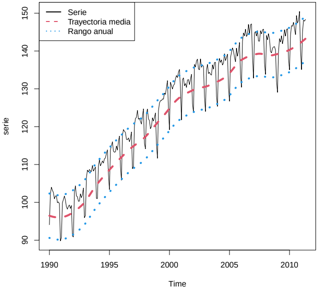
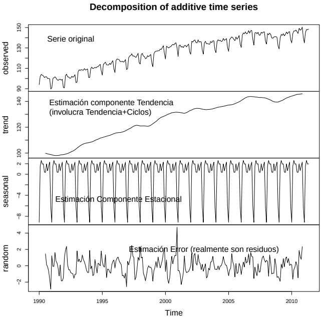


Figura 2: Últimos 24 datos (enero-2019 a dic. 2020) del índice de producción nominal, total industria. Serie ajustada: enero 2019 a dic. 2019 y pronósticos para el 2020, con base en el Suavizamiento Exponencial Holt-Winters Multiplicativo (SEHWM), aplicado sobre la serie de enero 2001 a dic. 2019.

Ejemplo 1: Índice mensual salario real sector manufactura



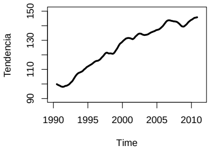
(a)



(b)

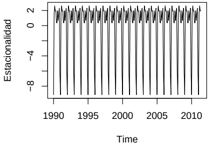
Figura 3: (a) Observaciones entre enero de 1990 a mayo de 2011; (b) Descomposición aditiva con función R decompose. Esta serie se considera aditiva pues su variabilidad alrededor de su tendencia no cambia en el tiempo. Observe además que tiene componente estacional.

Podemos extraer cada una de las componentes obtenidas en R con la función decompose(): Pero lo que el filtro `decompose()` estima como tendencia tiene combinada tendencia y ciclos.



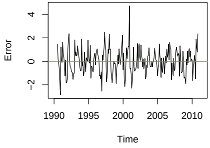
Tendencia+Ciclos

+

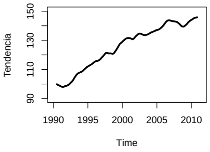


Estacionalidad

+

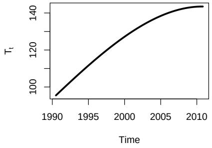


Error



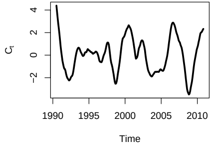
Tendencia+Ciclos

=



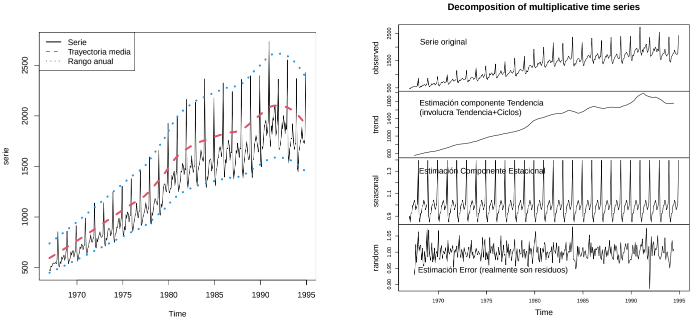
Tendencia pura

+



Ciclos

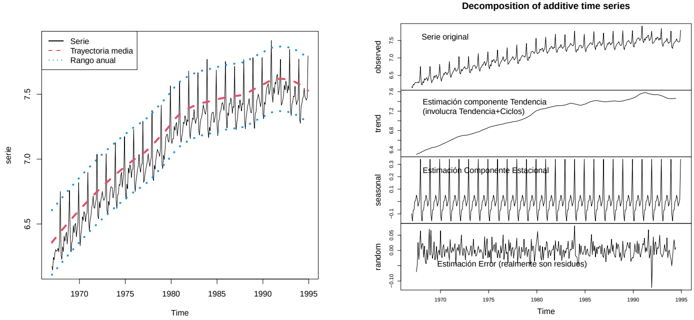
Ejemplo 2: Ventas mensuales de licor en USA



(a) (b)

Figura 4: (a) Observaciones entre enero de 1967 - diciembre de 1994 (miles de dólares); (b) Descomposición multiplicativa con función R decompose. Esta serie se considera multiplicativa pues su variabilidad alrededor de su tendencia cambia en el tiempo en la dirección de la tendencia. Observe que tiene componente estacional.

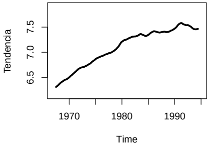
Para formular un modelo de regresión para la serie anterior, ***nos interesa estudiar su logaritmo natural.*** $\log(Y_t)$ es una serie de componentes aditivas:



(a) (b)

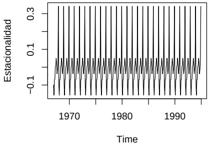
Figura 5: (a) Observaciones entre enero de 1967 - diciembre de 1994 (log. miles de dólares); (b) Descomposición aditiva con función R `decompose`. El logaritmo de los datos se considera una serie aditiva, pues su variabilidad alrededor de su tendencia es constante en el tiempo. Observe que el patrón estacional se conserva con la transformación.

Para $\log(Y_t)$ *podemos extraer cada una de las componentes obtenidas en R con la función* `decompose()`*:*



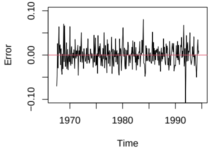
Tendencia+Ciclos

+

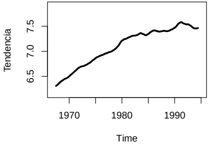


Estacionalidad

+

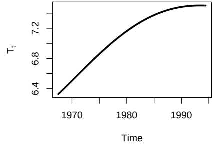


Error



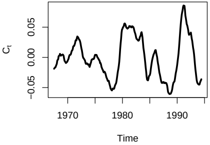
Tendencia+Ciclos

=



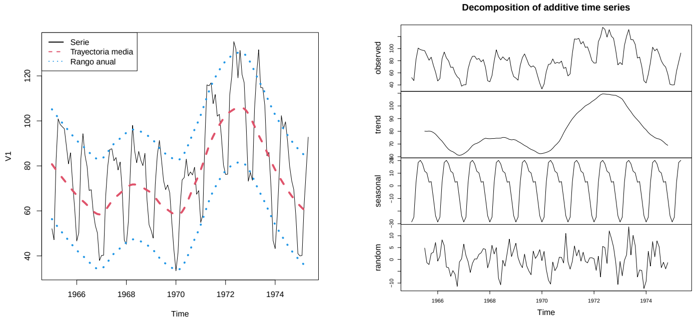
Tendencia pura

+



Ciclos

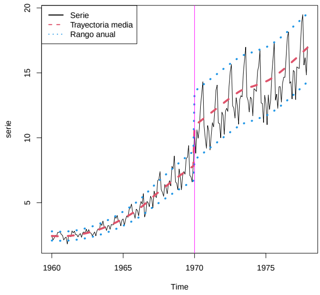
Ejemplo 3: Número mensual de casas unifamiliares iniciadas



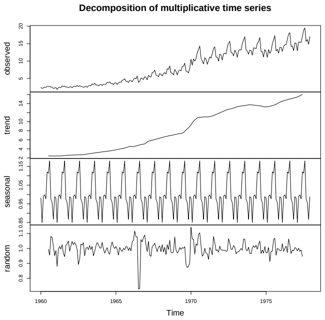
(a) (b)

Figura 6: (a) Observaciones entre enero 1965 a Mayo 1975. Fuente: Peña, Tiao and Tsay (2001); (b) Descomposición aditiva con función R decompose. Esta serie se considera aditiva pues su variabilidad alrededor de su trayectoria de largo plazo es aprox. constante. Observe que tiene componente estacional y que ***en el intervalo de tiempo observado, el patrón de largo plazo es dominado por la componente cíclica, por lo cual es difícil identificar una trayectoria suave para la tendencia.***

Ejemplo 4: Millas mensuales por pasajero, Estados Unidos



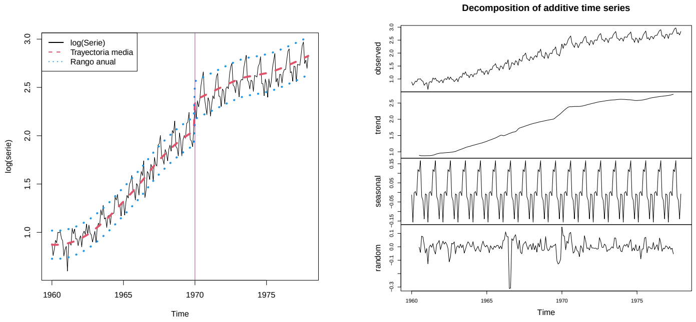
(a)



(b)

Figura 7: (a) Observaciones entre enero 1960 a Diciembre 1977; (b) Descomposición multiplicativa con función R `decompose`. ***Esta serie tiene varios cambios en sus patrones por un cambio estructural alrededor de 1970: Cambio en la tendencia*** pues hay un salto abrupto en su trayectoria, ***en la varianza*** porque hasta 1970 era no constante y luego se comporta con constancia, y ***en la forma del patrón estacional*** porque la forma del patrón dentro de un año calendario, comparando los primeros y los últimos años se observa con diferencias, ***aunque el filtro de la descomposición estima la estacionalidad como un patrón exacto que no cambia en el tiempo, eso no significa que sea consecuente con lo que realmente sucede en la serie.***

Para formular un modelo de regresión para la serie anterior, ***nos interesa estudiar su logaritmo natural.*** $\log(Y_t)$ es una serie de componentes aditivas:



(a) (b)

Figura 8: (a) Logaritmo natural de millas mensuales por pasajero, Estados Unidos, enero 1960 a diciembre 1977; (b) Descomposición aditiva con función R `decompose`. Esta serie se considera aditiva para su descomposición, pues a pesar que su variabilidad antes de enero de 1970 es menor que después de este tiempo, por lo menos dentro de cada tramo (antes y después del punto de cambio) ésta es constante. ***Observe además, que aún se mantiene el salto en la tendencia cerca a enero de 1970, y la forma del patrón estacional en los primeros años vs. los últimos años, muestra algunas diferencias (Ver Figura 9).***

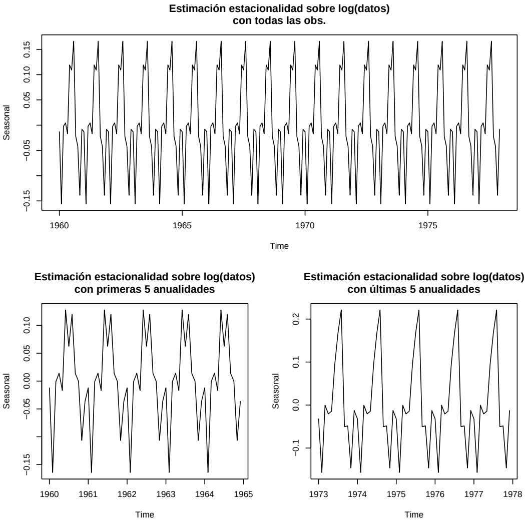
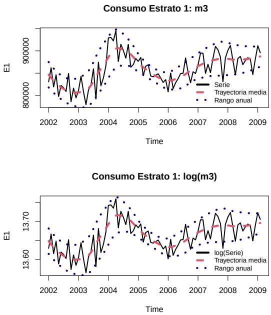
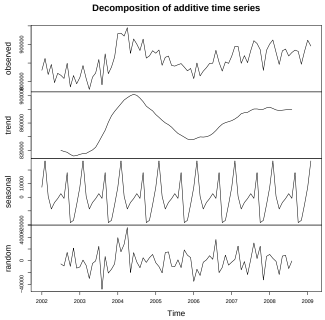


Figura 9: Estimaciones de la componente estacional con la descomposición aditiva sobre Logaritmo natural de millas mensuales por pasajero, Estados Unidos. Arriba: Con todos los datos (enero 1960 a diciembre 1977); Abajo-izq.: Con los primeros cinco años (enero 1960 a diciembre 1964); Abajo-der.: Con los últimos cinco años (enero 1973 a diciembre 1977).

Ejemplo 5: Datos EPM consumo mensual de agua (m^3), estrato 1



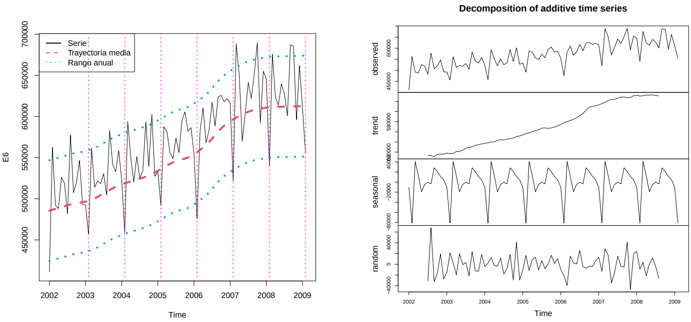
(a)



(b)

Figura 10: (a) Observaciones entre enero 2002 a Febrero 2009 (arriba en escala original, abajo en escala log natural); (b) Descomposición aditiva con función R `decompose`. Esta serie se considera aditiva para su descomposición pues su patrón de variabilidad alrededor de su trayectoria de largo plazo no muestra cambio al aplicar logaritmo natural. ***De otro lado, a pesar que el filtro de la descomposición le estima una componente estacional periódica exacta, tal patrón no es realmente observado en la serie. Puede ser que esta serie no sea estacional o bien, su estacionalidad es de tipo estocástica y para resolver esto se necesitan otros análisis que veremos al final del curso.***

Ejemplo 6: Datos EPM consumo mensual de agua (m^3), estrato 6



(a) (b)

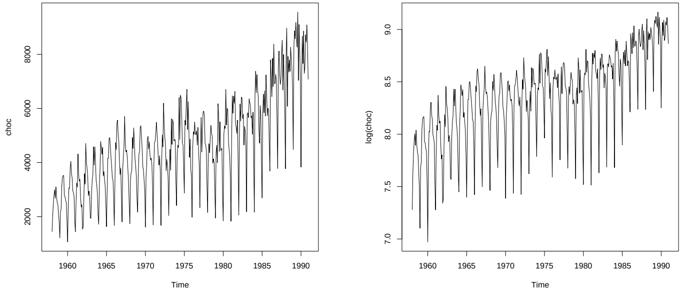
Figura 11: (a) Observaciones entre enero 2002 a Febrero 2009. (b) Descomposición aditiva con función R `decompose`. Esta serie se considera aditiva pues su variabilidad alrededor de su trayectoria de largo plazo es aprox. constante y ***al parecer es estacional: Las líneas verticales de referencia trazadas en (a) corresponden a los febreros, éstas nos permiten ver que en cada año en febrero el consumo es menor; aunque el patrón dentro del resto del año no mantiene exactamente su forma año tras año, contrario a lo que da como estimación el filtro de la descomposición.***

Contenido

ACTIVIDAD DE REPASO

Con el fin de afianzar y aplicar lo visto sobre análisis descriptivos en series de tiempo y familiarizarse con las rutinas R necesarias, realice lo siguiente.

1. Leer en el Capítulo 3 de Notas de Clase, la Sección 3.9: "**Guía para el análisis de componentes de una serie**", y el Apéndice A de este capítulo: "Función `decompose`: Descomposición estacional clásica mediante medias móviles".
2. Consultar en el "**Capítulo 1 de Notas de Clase: Breve introducción al R**" las funciones de lectura de datos, creación de objetos tipo series de tiempo, funciones para gráficos de dispersión, box plots y de descomposición.
3. Con la ayuda del paquete R, realizar análisis descriptivo sobre la serie mensual de la productividad de la industria del chocolate en Australia, cuyos datos se encuentran en el archivo cbe2.csv, columna 1. Use el programa R disponible en moodle "**Código R para análisis descriptivo serie choc en archivo cbe2.csv.R**". Resuelva las siguientes preguntas
   1. *De acuerdo al gráfico de la serie y de su logaritmo natural, es aditiva o multiplicativa esta serie?, en cuál escala es más factible asumir varianza constante o aproximadamente constante?*
   2. Determine únicamente con la gráfica de la serie o de su logaritmo natural por qué hay componente estacional.
   3. *Determine usando la gráfica de box plots comparativos de las distribuciones de la serie (o de su logaritmo) según meses del año, por qué hay patrón estacional.*
   4. La componente estacional es aproximadamente igual en su forma año tras año?
   5. *Puede caracterizar esta serie como creciente o decreciente? hay perturbación de la tendencia por componente cíclica?*



(a) (b)

Figura 12: (a) Serie de la productividad de la industria del chocolate en Australia, en escala original; (b) serie en escala logaritmo natural.