METODOS ELEMENTALES DE PROCESAMIENTO DE SERIES DE TIEMPO

Jorge Galbiati Riesco

En este apunte se da una visión general sobre algunos procedimientos en el análisis en series de tiempo. Inicialmente presentamos el problema general de predicción, luego presentamos métodos clásicos ó ingenuos de suavizamiento en el análisis de series de tiempo.

La aplicación de estas técnicas requieren de trabajo computacional que debe ser complementado en la ejercitación de este curso.

El material presentado fue escrito en base a apuntes del profesor Renato Allende, que tuvo como referencia notas de clase de los profesores Pilar Iglesias y Eduardo Engel.

INTRODUCCION

Una de las motivaciones para el estudio del tema surge desde tiempos remotos donde una de las principales inquietudes del hombre ha sido estimar el futuro utilizando información del presente y del pasado. Esto se llama *predecir*. Es evidente que las diversas instituciones requieren conocer el comportamiento futuro de ciertos fenómenos con el fin de planificar, preveer o prevenir.

La Estadística ha desarrollado teoría y métodos que apuntan a resolver el problema de predicción. Sin embargo, este no puede ser resuelto por argumentos puramente matemáticos, debe ser el resultado de una combinación matemático-especialista. La predicción es una Ciencia y es un Arte, y la mayor dificultad es la mala comunicación entre los analistas de información y de predicción y los usuarios de éstas.

DEFINICION BÁSICA DE SERIE DE TIEMPO

Una *serie de tiempo* es una colección o conjunto de mediciones de cierto fenómeno o experimento registrados secuencialmente en el tiempo, en forma *equiespaciada* (a intevalos de tiempo iguales) .

Las observaciones de una serie de tiempo serán denotadas por

$$Y(t_1),\ Y(t_2)\ ,\dots\ ,\ Y(t_n)$$

donde $Y(t_i)$ es el valor tomado por el proceso en el instante t_i.

Ejemplos de series de tiempo

1. Economía: Precios de un articulo, tasas de desempleo, tasa de inflación, índice de precios, precio del dólar, precio del cobre, precios de acciones, ingreso nacional bruto, etc.

2. Meteorología: Cantidad de agua caída, temperatura máxima diaria, Velocidad del viento (energía eólica), energía solar, etc.

3. Geofísica: Series sismológicas.

4. Química: Viscosidad de un proceso, temperatura de un proceso.

5. Demografía: Tasas de natalidad, tasas de mortalidad.

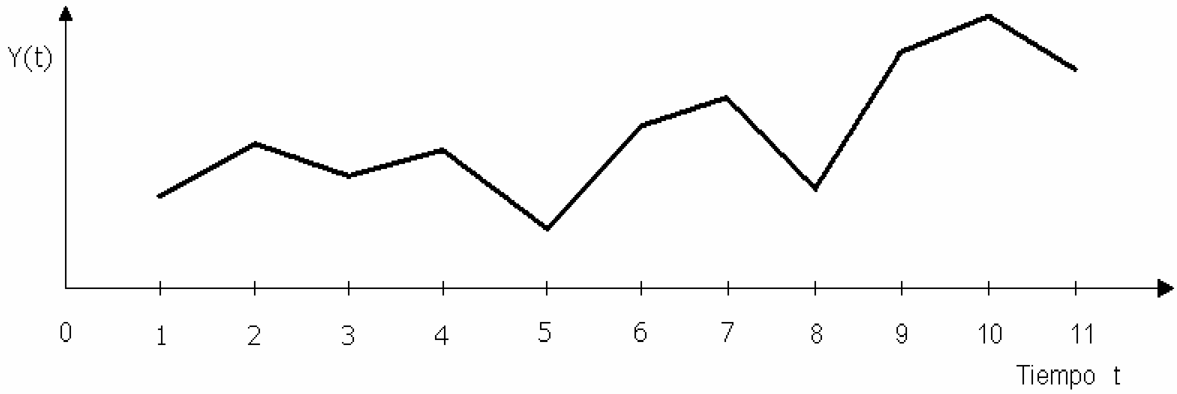
6. Medicina: Electrocardiograma, electroencéfalograma.

7. Marketing: Series de demanda, gastos, utilidades, ventas, ofertas.
8. Telecomunicaciones: Análisis de señales.

9. Transporte: Series de tráfico.

ANALISIS GRAFICO DE UNA SERIE DE TIEMPO

Por muy simple que parezca, el paso más importante en el análisis de series de tiempo consiste en graficar la serie.

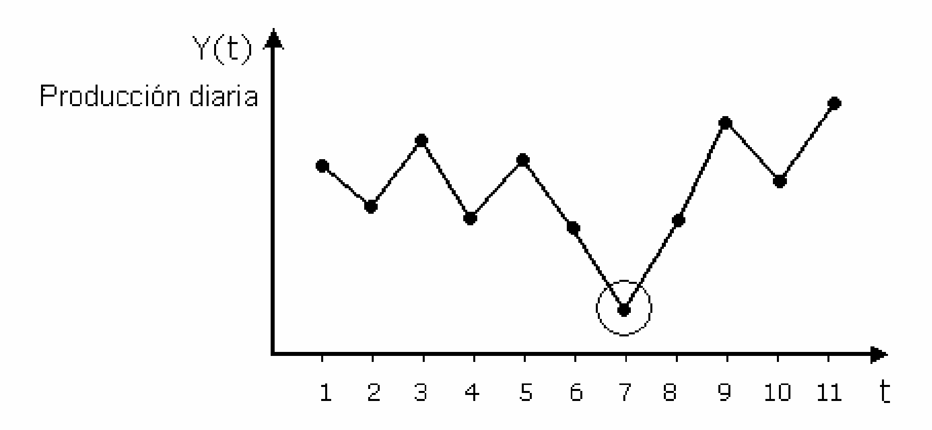


Esto debe hacerse siempre, independiente de cuán simples o complejos sean los procedimientos que se emplean posteriormente.

El gráfico de la serie permitirá detectar los siguientes elementos:

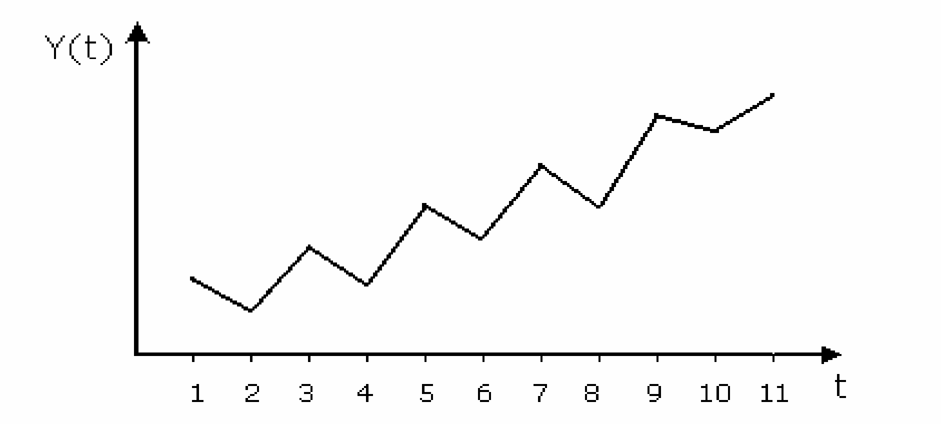
a) Outliers:

Se refiere a puntos de la serie que se escapan de lo normal.
Si se sospecha que una observación es un outliers, se debe reunir información adicional sobre posibles factores que afectaron el proceso. Por ejemplo, en un estudio de la producción diaria de cobre se presentó la siguiente situación:



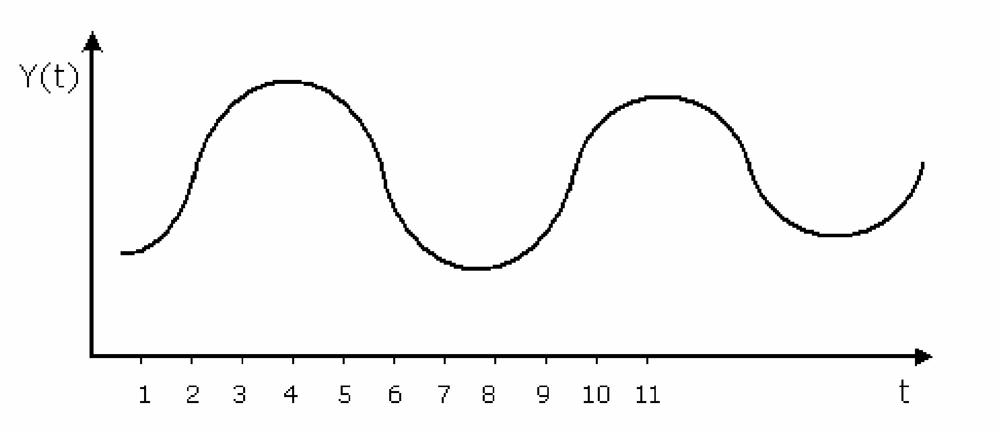
b) <u>Tendencias</u>

La tendencia representa el comportamiento predominante de la serie. Esta puede ser definida vagamente como el cambio de la media a lo largo de un extenso período de tiempo.



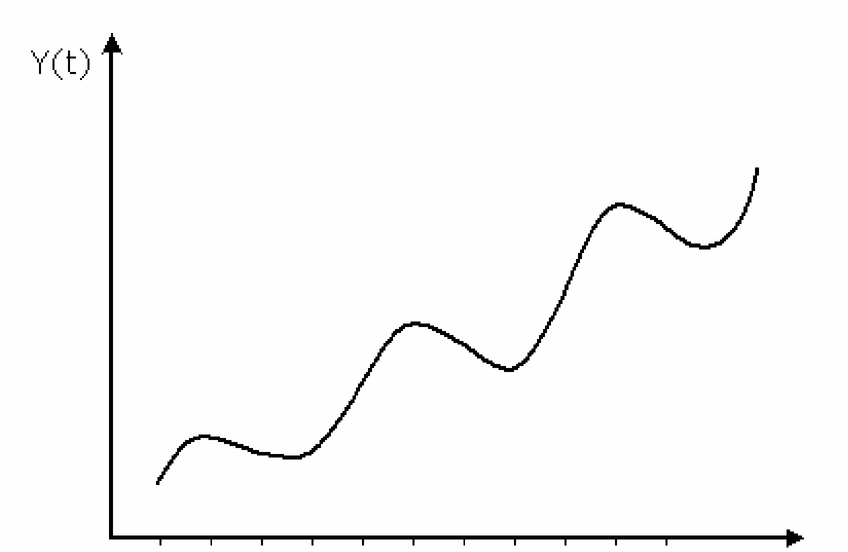
c) <u>Variaciones cíclicas o estacionales</u>

La variación estacional representa un movimiento *periódico* de la serie de tiempo. La duración del período puede ser un año, un trimestre, un mes, un día, etc.



Se suele hacer una distinción entre cíclicas y estacionarias. Estas últimas ocurren con períodos identificables, como la estacionalidad del empleo, o de la venta de ciertos productos, cuyo período es un año. El término variación cíclica se suele referir a ciclos grandes, cuyo período no es atribuible a alguna causa. Por ejemplo, fenómenos climáticos, que tienen ciclos que duran varios años.

Las tendencias y estacionalidades pueden darse simultáneamente.



d) Variaciones aleatorias.

Los movimientos irregulares (al azar) representan todos los tipos de movimientos de una serie de tiempo que no sea tendencia, variaciones estacionales y fluctuaciones cíclicas.

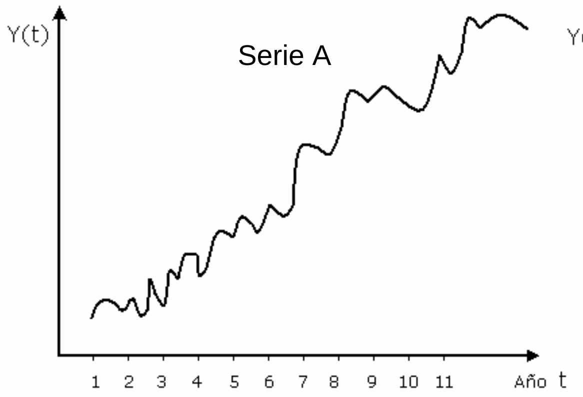
El análisis gráfico de los datos se acostumbra a resumir en una tabla como la que siguiente:

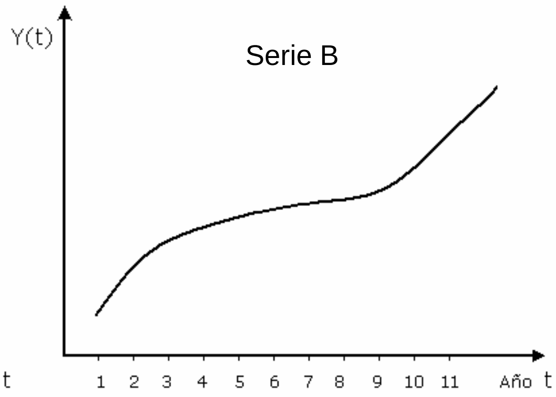
Tabla de familiarización

	Estacionalidad	**Tendencia**	**Aleatoria**
Alta			
Media			
Baja			

Ejercicios:

1) Para cada una de las series graficadas a continuación realizar al análisis gráfico completando la tabla de familiarización.





2) Construya la grafica y la tabla de familiarización para la siguiente serie.

Planificación de un casino

Objetivos: Planificación de compra de alimentos y necesidades de servicio para satisfacer la demanda de almuerzo en un gran casino.

Descripción de la serie.
Serie: Número de almuerzos servidos por mes en el casino II de la Universidad de Campinas-Brasil para el período de enero 1977 a marzo de 1980, de acuerdo a la administración general del restaurante.

Tabla de datos

t	Y(t)	t	Y(t)	t	Y(t)
1	20636	14	28183	27	63167
2	18708	15	56632	28	42520
3	62944	16	56641	29	50572
4	50272	17	56555	30	53875
5	69375	18	57185	31	27233
6	50056	19	33906	32	57942
7	20604	20	67261	33	47610
8	54947	21	52232	34	61738
9	50576	22	58232	35	51168
10	50425	23	45726	36	26370
11	44202	24	24550	37	42964
12	27604	25	30954	38	42748
13	28791	26	34295	39	62390

Modelos Clásicos

Un modelo clásico de series de tiempo, supone que la serie Y(1), ..., Y(n) puede ser expresada como suma o producto de tres componentes: tendencia, componente estacional y un término de error aleatorio:

1. $Y(t) = T(t)+E(t)+A(t)$ Modelo aditivo
2. $Y(t) = T(t)\,E(t)\,A(t)$ Modelo multiplicativo

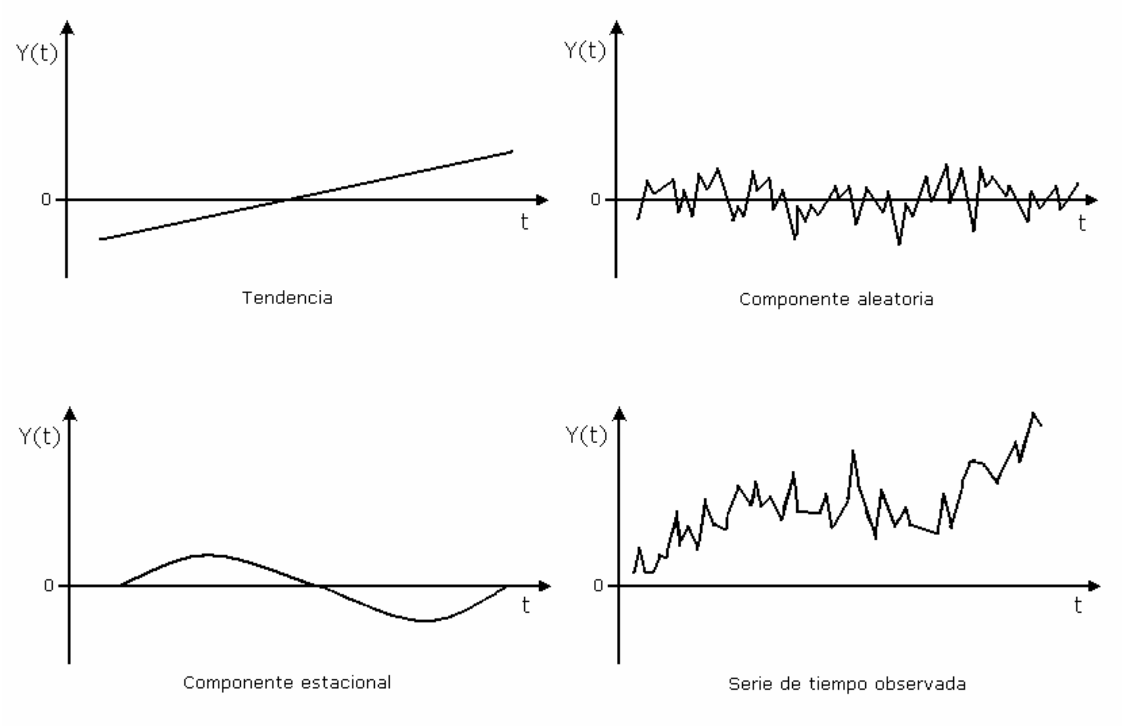
donde:

T: Tendencia de la serie.

E: Variación Estacional.

A: Variaciones aleatorias.

El gráfico siguiente muestra la serie y sus componentes, para el caso aditivo



El problema que se presenta es modelar adecuadamente las componentes de la serie.

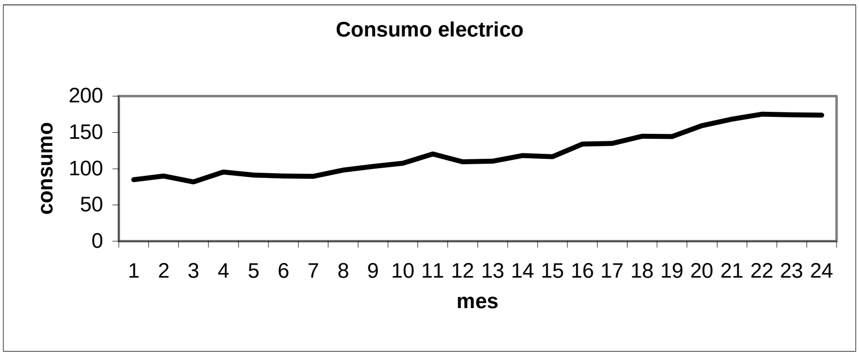
ESTIMACIÓN DE LA TENDENCIA

Hay varios métodos para estimar la tendencia T(t), uno de ellos es utilizar un modelo de regresión lineal. Se pueden utilizar otros tipos de regresiones, como regresión cuadrática, logística, exponencial, entre otros.

EJEMPLO 1: La tabla presenta parte de los datos de una serie de energía eléctrica. Son 24 datos mensuales referentes a los años 1977 a 1978.

Consumo de Energía Eléctrica

t	Y(t)	t	Y(t)
1	84,6	13	110,3
2	89,9	14	118,1
3	81,9	15	116,5
4	95,4	16	134,2
5	91,2	17	134,7
6	89,8	18	144,8
7	89,7	19	144,4
8	97,9	20	159,2
9	103,4	21	168,2
10	107,6	22	175,2
11	120,4	23	174,5
12	109,6	24	173,7



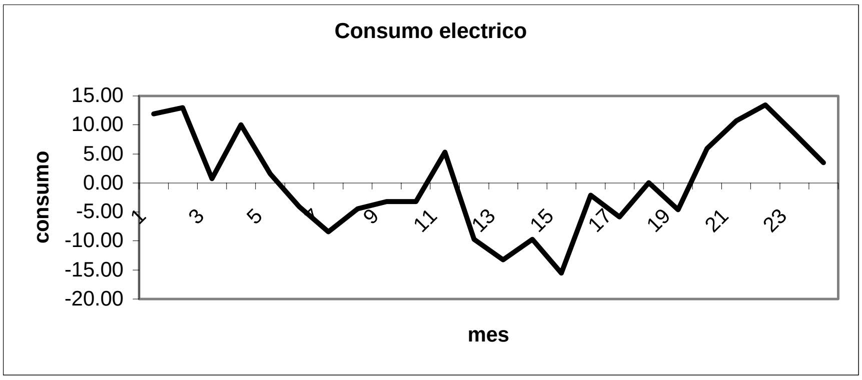
El modelo de tendencia propuesto es un modelo de regresión lineal:

$$Y(t) = a + b\,t + A(t)$$

Por mínimos cuadrados se obtiene

$$T(t) = 68.45 + 4.24 * t$$

La serie sin tendencia se ve de la siguiente manera:



Se observa un ciclo que dura casi todo el período observado, de 24 meses. □

ESTIMACION DE LA COMPONENTE ESTACIONAL

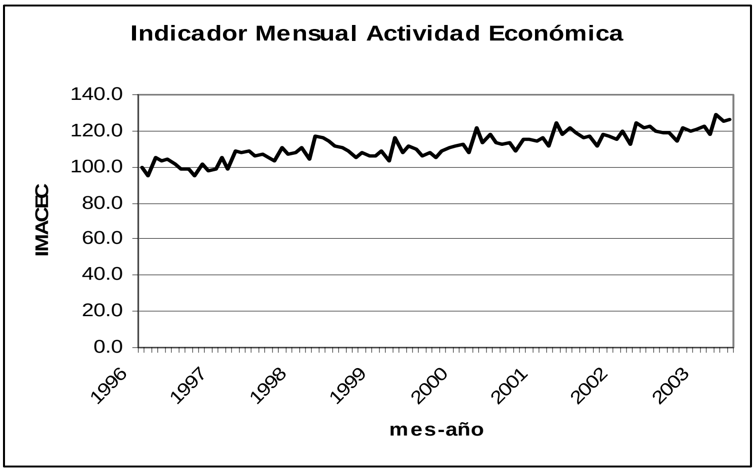
Para estimarla, se debe conocer el período, y se deben tener datos de varios períodos consecutivos. Por ejemplo, datos mensuales, estacionalidad de un año.

<u>EJEMPLO 2</u>

Indicador Mensual de Actividad Económica (IMACEC). Base del índice : 1996=100
Corresponde al nuevo Indicador Mensual de Actividad Económica (Imacec), estructurado a base de la matriz insumo-producto de 1996. La cobertura de este indicador comprende casi la totalidad de las actividades económicas incluidas en el PIB.

Las cifras de 2000 y 2001 son provisionales.
Las cifras de 2002 y 2003 son preliminares.
Fuente: Banco Central de Chile.

	1996	1997	1998	1999	2000	2001	2002	2003
Enero	99.6	105.0	110.8	109.2	112.6	116.4	119.7	122.6
Febrero	94.9	98.6	104.3	103.7	107.6	111.8	113.0	118.3
Marzo	105.4	109.1	117.5	116.4	121.2	124.3	124.4	128.8
Abril	103.4	108.1	116.1	108.0	113.8	118.0	122.0	125.3
Mayo	104.2	109.2	114.4	111.2	117.9	121.7	123.0	126.1
Junio	101.3	106.5	111.9	110.0	113.1	119.1	120.1	
Julio	98.7	107.1	110.9	106.4	112.3	116.0	118.9	
Agosto	98.7	105.6	109.0	108.1	113.4	116.9	119.1	
Septiembre	94.8	103.8	105.4	105.7	108.6	111.4	114.6	
Octubre	102.0	110.9	107.7	109.2	115.4	118.4	121.7	
Noviembre	98.0	106.8	106.1	110.7	114.9	117.3	119.9	
Diciembre	99.2	108.4	106.5	111.9	114.4	115.7	120.9	



Se estima la tendencia por regresión lineal $Y(t) = a + bt + A(t)$

dando el siguiente resultado:

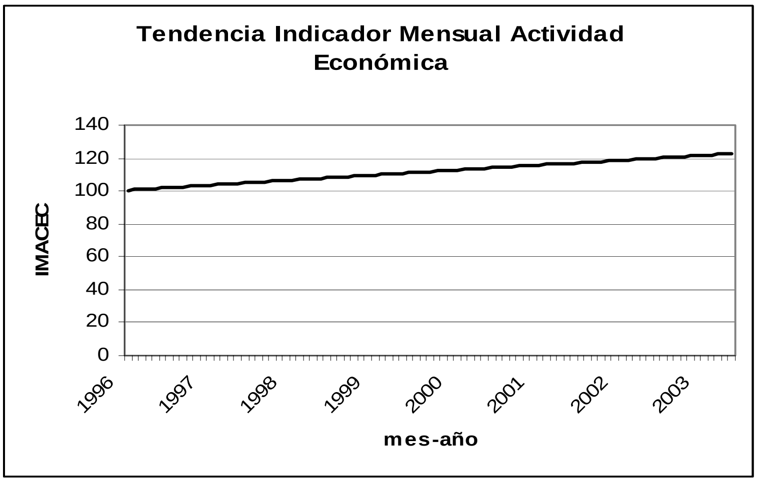
Intercepto a = 100.3. Corresponde al valor de partida.

Pendiente b = 0.253 . Corresponde al aumento medio mensual.

Coeficiente de determinación R^2 = 0.74, que indica un ajuste moderadamente bueno.

El error estándar de los errores se estimó en 3.98.

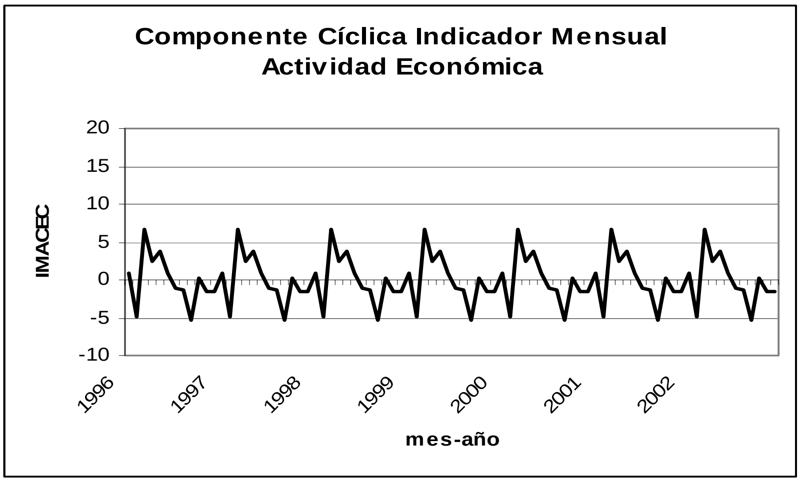
La recta de regresión correspondiente a la tendencia se muestra en el siguiente gráfico:



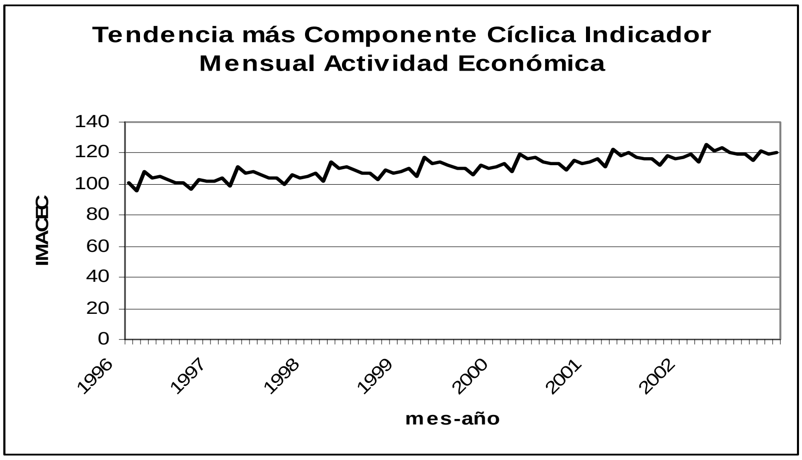
Asumiremos un modelo clásico aditivo. Entonces para obtener una estimación de la estacionalidad, restamos los valores ajustados de la tendencia a los datos, obteniendo una serie sin tendencia. Luego promediamos todos los valores de enero, los de febrero, los de marzo, etc., obteniendo doce valores mensuales promedio:

Mes	Prom.
Enero	0.8
Febrero	-4.9
Marzo	6.7
Abril	2.4
Mayo	3.8
Junio	0.8
Julio	-1.2
Agosto	-1.3
Septiembre	-5.4
Octubre	0.2
Noviembre	-1.7
Diciembre	-1.5

que se muestran en el siguiente gráfico:

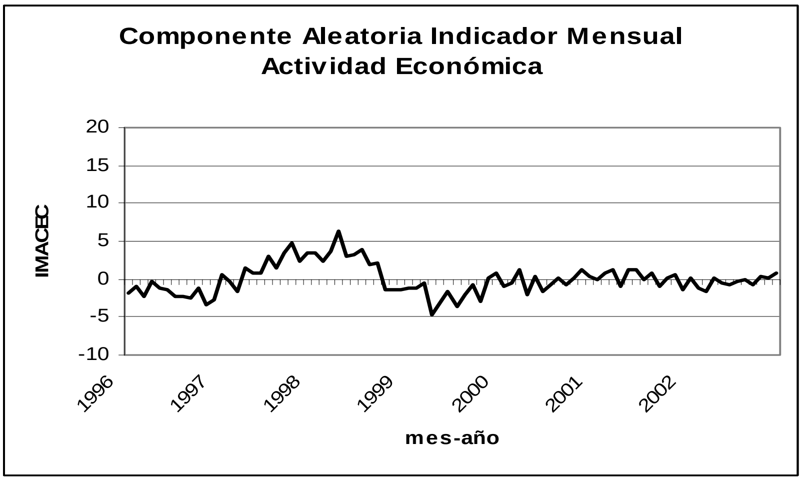


Se observan valores altos a partir de marzo, y bajos en torno a septiembre.
Si recomponemos la serie con tendencia y componente cíclica, sin la componente aleatoria, tenemos la situación que se ilustra en el gráfico siguiente:



Con esto se pueden hacer predicciones futuras, extrapolando la recta de regresión y sumándole la componente cíclica del mes correspondiente. Dentro de un rango limitado, estas predicciones pueden ser acertadas.

A continuación se muestra el gráfico de la componente aleatoria sola.



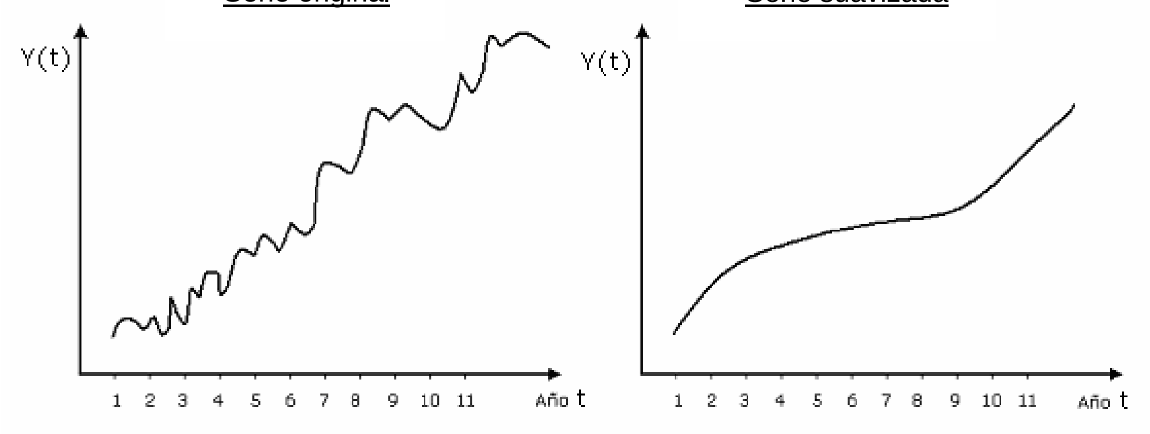
Si se usa el modelo multiplicativo, el procedimiento es parecido.
Nota: Junto con las series de datos como esta, el Banco Central también entrega series sin tendencia y desestacionalizadas.

□

SUAVIZAMIENTO DE SERIES DE TIEMPO

Una forma de visualizar la tendencia, es mediante suavizamiento de la serie. La idea central es definir a partir de la serie observada una nueva serie que filtra o suaviza los efectos ajenos a la tendencia (estacionalidad, efectos aleatorios), de manera que podamos visualizar la tendencia.

Serie original Serie suavizada



Promedio Móvil.

Este método de suavizamiento es uno de los más usados para describir la tendencia. Consiste en fijar un número k, preferentemente impar, como 3, 5, 7, etc., y calcular los promedios de todos los grupos de k términos consecutivos de la serie. Se obtiene una nueva serie suavizada por promedios móviles de orden k. De este modo se tienden a anular las variaciones aleatorias.

Por ejemplo, consideremos una serie de seis observaciones y fijemos el orden k=3. Entonces los términos de la serie suavizada son

t	$Y(t)$	$Z(t)$ *media móvil de orden* $k=3$
1	$Y(1)$	$-$
2	$Y(2)$	$Z(2)=\dfrac{Y(1)+Y(2)+Y(3)}{3}$
3	$Y(3)$	$Z(3)=\dfrac{Y(2)+Y(3)+Y(4)}{3}$
4	$Y(4)$	$Z(4)=\dfrac{Y(3)+Y(4)+Y(5)}{3}$
5	$Y(5)$	$Z(5)=\dfrac{Y(4)+Y(5)+Y(6)}{3}$
6	$Y(6)$	$-$

Nótese que Z(1) y Z(6) no se pueden calcular. En general, se pierden k/2 términos en cada extremo.

EJEMPLO 3.

Precio del dólar observado, días miércoles, enero a junio año 2003.

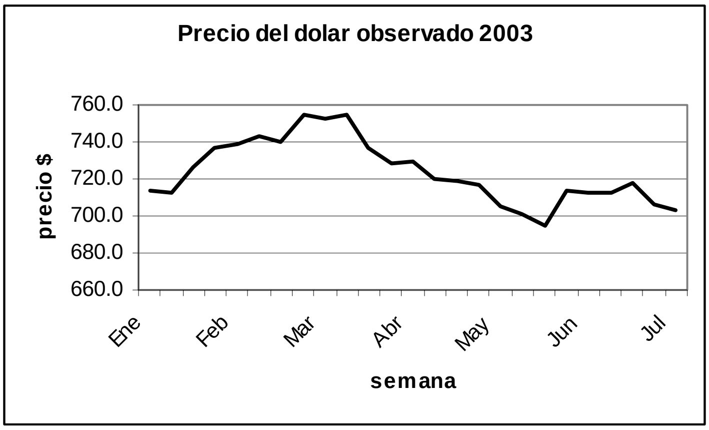
Fuente: Banco Central de Chile.

En las columnas 3 a 6, se entregan los promedios móviles de orden 3, 5, 7 y 9,

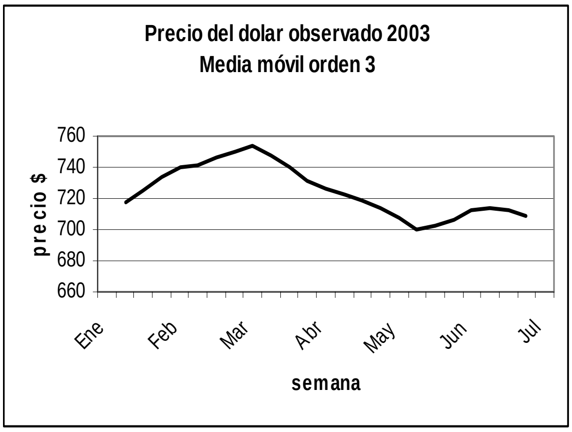
mes	precio	k=3	k=5	k=7	k=9
Ene	713.4				
	712.6	717.5			
	726.6	725.4	725.7		
	737.1	734.2	731.6	730.3	
Feb	738.9	739.7	737.2	736.2	735.5
	743.1	740.7	742.9	742.0	740.1
	740.2	746.2	746.0	746.0	742.8
	755.3	749.4	749.2	745.9	743.0
Mar	752.8	754.3	747.9	744.4	742.2
	754.8	748.0	745.5	742.5	740.1
	736.5	739.9	740.4	739.6	737.4
	728.3	731.5	733.9	734.4	734.7
Abr	729.8	726.0	726.6	729.2	729.2
	720.0	722.8	722.6	722.1	723.4
	718.7	718.4	718.0	717.1	716.7
	716.5	713.5	712.3	712.2	714.2
May	705.3	707.7	707.2	709.9	712.4
	701.2	700.2	706.2	708.8	710.5
	694.2	703.1	705.3	708.0	710.2
	713.7	706.7	706.9	708.2	708.8
Jun	712.2	713.0	710.1	708.2	707.3
	713.0	714.2	712.5	708.5	
	717.4	712.1	710.3		
	706.0	708.8			
Jul	703.1				

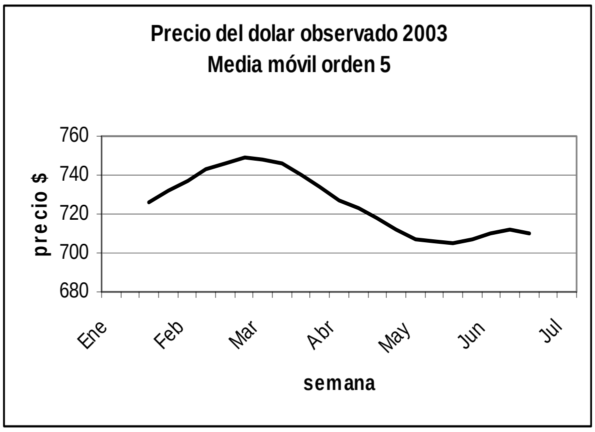
respectivamente.

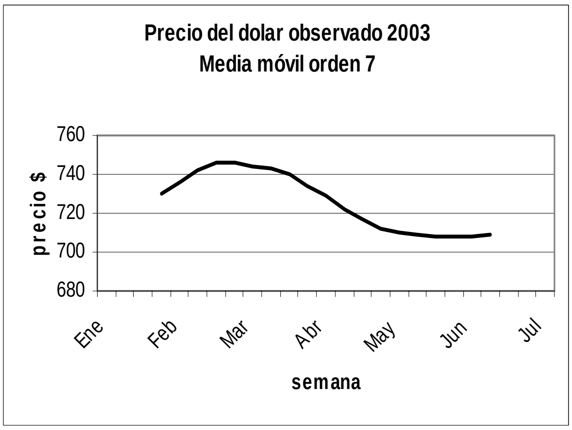
La serie original aparece graficada a continuación.

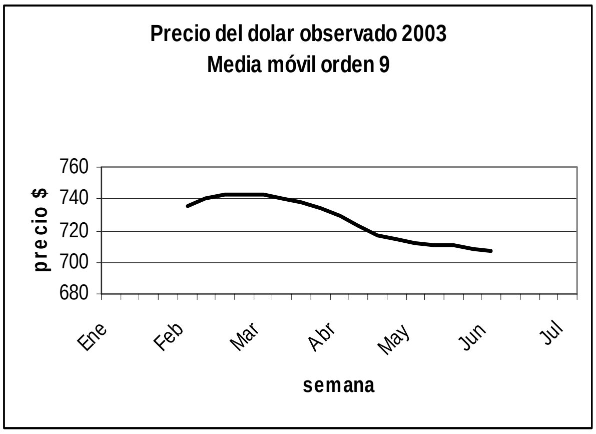


Los gráficos siguientes corresponden a las medias móviles. Se observa cómo a medida que aumenta el orden, el efecto del suavizado es mayor. Pero también se pierden más datos en los extremos.









El suavizamiento de media móvil es muy fácil de aplicar, permite visualizar la tendencia de la serie. Pero tiene dos inconvenientes: No es posible obtener estimaciones de la tendencia en extremos y no entrega un medio para hacer predicciones.

Si la serie presenta un efecto estacional de período k, es conveniente aplicar un suavizamiento de media móvil de orden k. En tal caso se elimina el efecto estacional, junto con la variación aleatoria, observándose solamente la tendencia.

SUAVIZAMIENTO EXPONENCIAL

Este modelo se basa en que una observación suavizada, en tiempo *t*, es un promedio ponderado entre el valor actual de la serie original y el valor de la serie suavizada, en el tiempo inmediatamente anterior. Si *Y(t)* representa la serie de tiempo original, y *Z(t)* la serie de tiempo suavizada, entonces lo anterior se puede escribir

$$Z(t) = a \cdot Y(t) + (1 - a) \cdot Z(t-1)$$

en que *a* es un número entre 0 y 1.

Si *a* es cercano a 1, la serie suavizada pondera más fuertemente el valor original, luego ambas se parecen, y en consecuencia, el suavizamiento es poco.

Si *a* se acerca a 1/2, se ponderan moderadamente la serie original y la suavizada, por lo que el suavizamiento es moderado.

Si *a* es cercano a cero, *(1-a)* es cercano a 1, y la serie suavizada pondera más fuertemente el valor suavizado inmediatamente anterior, por lo que el suavizado es importante.

Consecuencia de la fórmula anterior es que la serie suavizada se puede expresar como

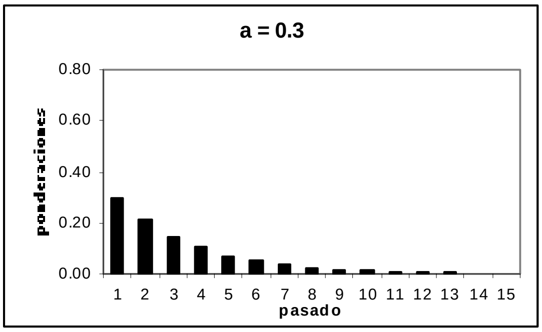
$$\mathrm{Z}(t) = \mathrm{a} \cdot \mathrm{Y}(t) + \mathrm{a} \cdot (1\text{-}\mathrm{a}) \cdot \mathrm{Y}(t-1) + \mathrm{a} \cdot (1\text{-}\mathrm{a})^2 \cdot Y(t-2) + \mathrm{a} \cdot (1\text{-}\mathrm{a})^3 \cdot Y(t-3) + \cdots$$

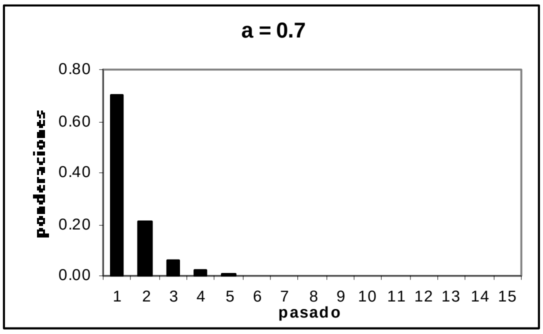
Es decir, cada término suavizado es un promedio ponderado de todos los términos históricos de la serie original, con ponderaciones

$$a, \qquad a \cdot (1-a), \qquad a \cdot (1-a)^2, \qquad a \cdot (1-a)^3, \qquad a \cdot (1-a)^4, \quad \text{etc.}$$

Como *a* está entre 0 y 1, estos números se van achicando a medida que avanzan. Eso significa que a medida que nos alejamos hacia el pasado, los término van influyendo cada vez menos en el término presente. La rapidez con que disminuye la influencia es mayor mientras más grande (cercano a 1) es *a.*

Los gráficos siguientes muestran las ponderaciones de los términos hacia el pasado, cuando *a = 0.3* y cuando *a = 0.7*





Criterio para elegir *a*: Si la serie varía lentamente, se eligen valores de *a* cercanos a 0. (valor típico *a* = 0.3). En cambio, si varía bruscamente, se eligen valores de *a* cercanos a 1 (valor típico *a* = 0.7).

El método de suavizamiento exponencial sirve para hacer predicciones, pero sólo de un valor, siguiente al último valor observado. Si se tienen observaciones Y(n), Y(n-1), Y(t-2), ... Y(n-k).

Si tratáramos de obtener el término Z(n+1) con la fórmula para el suavizamiento exponencial, nos daría

$$Z(n+1) = a \cdot Y(n+1) + (1-a) \cdot Z(n)$$

pero como no tenemos una observación Y(n+1), la aproximamos por Y(n). Por lo tanto podemos
hacer una predicción para Y(n+1) con la fórmula del suavizamiento exponencial modificada de la siguiente forma:

$$Z(n+1) = a \cdot Y(n) + (1-a) \cdot Z(n)$$

Si se intentara hacer más de una predicción, daría el mismo valor, por eso que sólo se usa para predecir un valor a la vez. Sin embargo, en la práctica, cada vez que aparece una nueva observación real, se actualiza la fórmula anterior, pera predecir la siguiente. Así, cada vez que el tiempo avanza en una unidad, se predice un nuevo valor a futuro.

El valor de *a* se que sirve mejor se suele buscar por un sistema de prueba y error, hasta encontrar el que permite predecir mejor.

EJEMPLO 3

En la página siguiente se muestra el Producto Interno Bruto trimestral de Chile, desde el primer trimestre de 1996 hasta el primer trimestre de 2003, en millones de pesos. Fuente: Banco Central de Chile.

Año	Trim.	PIB	Suavizamiento exponencial a = 0.3	a = 0.5	a = 0.7
1996	I	7804934	7804934	7804934	7804934
	II	8038772	7875085	7921853	7968620
	III	7604665	7793959	7763259	7713852
	IV	7788918	7792447	7776088	7766398
1997	I	8141434	7897143	7958761	8028923
	II	8431424	8057427	8195093	8310674
	III	8238641	8111791	8216867	8260251
	IV	8489194	8225012	8353030	8420511
1998	I	8658075	8354931	8505553	8586806
	II	8910964	8521741	8708258	8813716
	III	8468861	8505877	8588559	8572317
	IV	8338698	8455723	8463629	8408784
1999	I	8572437	8490737	8518033	8523341
	II	8571428	8514945	8544730	8557002
	III	8334807	8460903	8439769	8401465
	IV	8636370	8513543	8538069	8565899
2000	I	8887713	8625794	8712891	8791169
	II	8972978	8729949	8842935	8918435
	III	8704303	8722255	8773619	8768543
	IV	8971751	8797104	8872685	8910788
2001	I	9175377	8910586	9024031	9096000
	II	9342846	9040264	9183438	9268792
	III	8960882	9016449	9072160	9053255
	IV	9146981	9055609	9109571	9118863
2002	I	9292810	9126769	9201190	9240626
	II	9503990	9239935	9352590	9424981
	III	9179436	9221786	9266013	9253099
	IV	9435570	9285921	9350792	9380829
2003	I	9621810	9386688	9486301	9549516

Junto a los datos se muestran tres suavizamientos exponenciales con *a=0.3*, *a=0.5* y *a=0.7.*

Como no hay datos indefinidamente hacia el pasado, los primeros términos de la serie sauvizada salen algo distorsionados, pues no consideran suficientes términos hacia atrás. La forma de calcular es la siguiente, partiendo del primer trimestre 1996, que llamaremos t=1 :

$$Z(1) = Y(1)$$

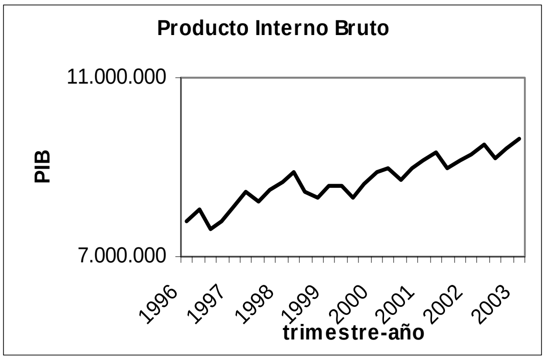
$$Z(2) = a\, Y(2) + (1-a)\, Z(1)$$

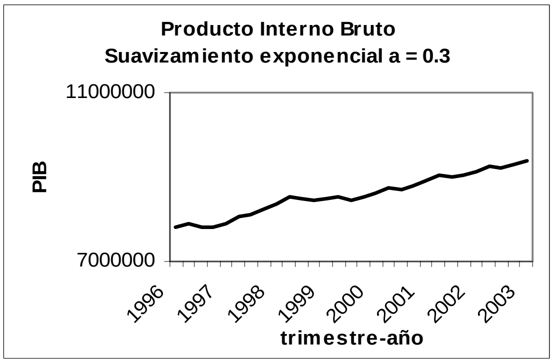
$$Z(3) = a\, Y(3) + (1-a)\, Z(2)$$

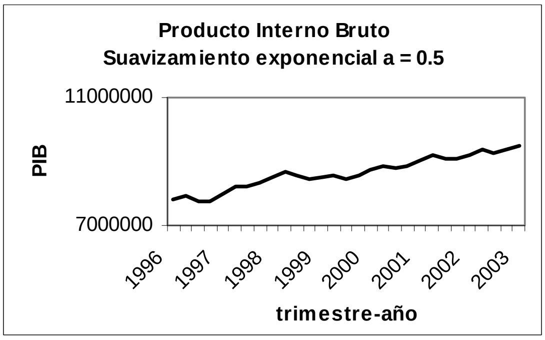
etc.

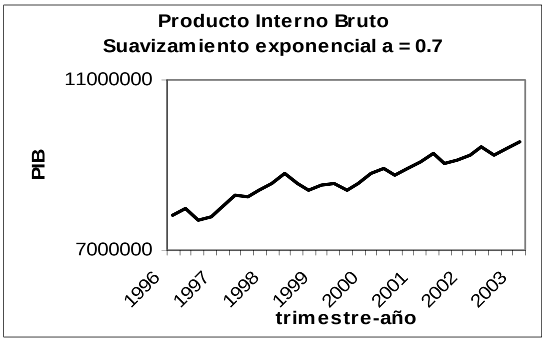
Como se ve, *Z(1)* no contiene toda la historia hacia atrás, *Z(2)* sólo un término hacia el pasado, *Z(3)* sólo 2, etc.

Los gráficos de la serie y los tres suavizamientos se muestran a continuación.









En los gráficos se puede apreciar que cuando la constante *a* es pequeña, cercana a cero, el suavizamiento es significativo. A medida que aumenta *a* acercándose a 1, el suavizamiento es menos y la serie suavizada se parece más a la serie original.

Se dispone de 29 datos. Es posible hacer una predicción del término de orden 30, que corresponde al segundo trimestre de 2003, mediante la fórmula

$$Z(30) = a\, Y(29) + (1-a)\, Z(29)$$

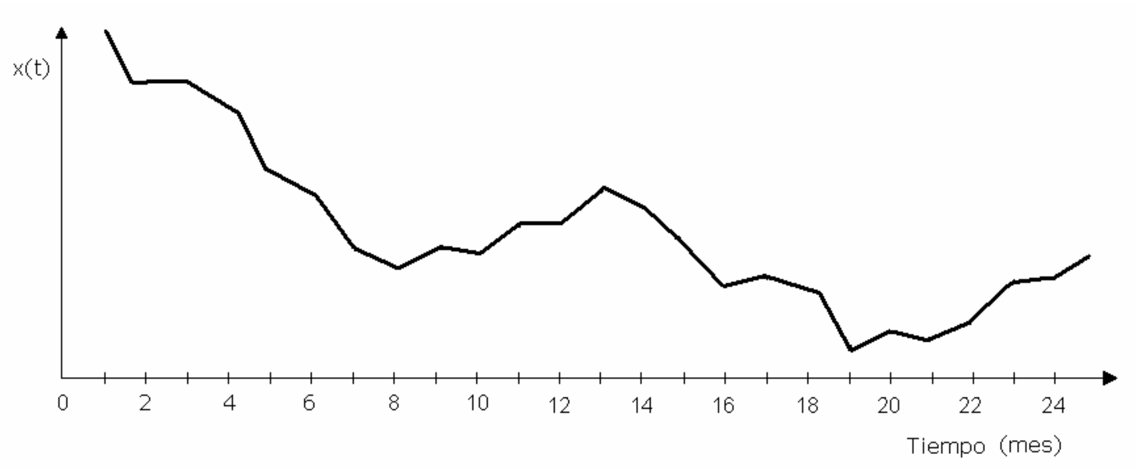
En el caso de a=0.3, se tiene

$Z(30) = 0.3 * 9621810 + 0.7 * 9386688 = 9457224$ millones de pesos.

En el caso a=0.5, la predicción da 9554055 millones de pesos. Y en el caso a=0.7, se obtiene el valor 9600122 millones. Observando el gráfico, ¿ cuál de las tres predicciones parece ser mejor?

PREGUNTAS

1. ¿Que puede decir de la serie de tiempo de datos mensuales, cuyo gráfico se muestra a continuación?



2. Cuál es el objeto de suavizar una serie de tiempo?

3. Explique por qué los promedios móviles y el suavizamiento exponencial "suavizan" una serie de tiempo, es decir, reducen la variación aleatoria.

4. ¿En qué consiste el suavizamientos de media móvil ?

5. Señale una diferencia entre el modelo clásico de series de tiempo y el suavizamiento por medias móviles, referente al uso que pueden tener.

6. Se tiene una serie de tiempo con datos económicos mensuales. A partir de ella se obtuvo una serie desestacionalizada y sin tendencia. Explique qué significa esto.

7. Señale una diferencia esencial entre el suavizamiento de una serie de tiempo por promedios móviles y el suavizamiento exponencial.

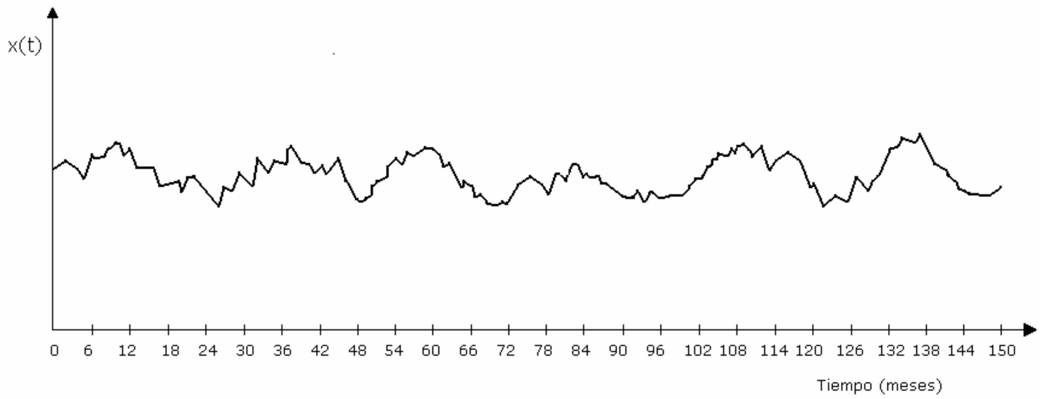
8. Se miden dos variables, x e y. Se calcula su coeficiente de correlación. ¿Qué mide este coeficiente?

9. ¿Cómo se interpreta el coeficiente de determinación, en una regresión lineal?

10. ¿En qué consiste el suavizamiento de media móvil, de una serie de tiempo ?

11. Se tiene una serie de tiempo con datos económicos mensuales. A partir de ella se obtuvo una serie desestacionalizada y sin tendencia. Explique qué significa esto.

12. ¿Que puede decir de la serie de tiempo de datos mensuales, cuyo gráfico se muestra a continuación, en relación a tendencia y estacionalidad?



13. ¿En qué consiste el suavizamiento exponencial?

14. Explique en qué consiste la aplicación de un modelo clásico aditivo a una serie de tiempo.

15. ¿En qué consiste el suavizamientos de media móvil ?