

Una metodología de series de tiempo para el área de la salud; caso práctico

A methodology of time series for the health field —a practical case

León Darío Bello P.

Especialista en estadística, docente de la Facultad Nacional de Salud Pública, Universidad de Antioquia, Medellín Colombia.
Cibercorreo: ldbello@guajiros.udea.edu.co

Sandra Martínez C.

Estudiante de séptimo semestre de gerencia en sistemas de información en salud, Universidad de Antioquia, Medellín, Colombia

Recibido: 9 Diciembre 2005 Aprobado: 29 Mayo 2007

Bello LD, Martínez S. Una metodología de series de tiempo para el área de la salud – caso práctico. Rev. Fac. Nac. Salud Pública. 2007; 25 (2): 117- 122

Resumen

En los últimos años, las series de tiempo han tenido desarrollos y aplicaciones importantes en el área de la salud; por ello, el presente artículo muestra una metodología ensayada en estudios anteriores, en que se esbozan los pasos en el análisis de series de tiempo para un enfoque clásico. El objetivo permite identificar un buen modelo de pronóstico usando técnicas de suavización: exponencial, simple, Holt y Winter. Además, se

realiza la validación de los supuestos de los residuales, dando confiabilidad a los resultados del modelo encontrado. Para aplicar el procedimiento se utilizaron los datos del número de consultas externas realizadas en la IPS de la Universidad de Antioquia durante abril del 2001 y septiembre del 2005.

----- *Palabras clave:* estudios de series temporales, aleatoriedad, tendencias, servicio ambulatorio

Summary

During the last years time series have had developments and important applications in the field of health. The present article shows a methodology tried in previous studies and outlines the steps in the analysis of time series for a classical approach. The objective allows identifying a good model of prognosis by using smoothing methods: exponential, simple, Holt and Winter. In addition, the validation of the assumptions

of the residuals is made managing to give trustworthiness to the results of the found model. In order to apply the procedure, data of number of outpatient appointments performed at the University of Antioquia Health Service from April 2001 to September 2005 were used.

----- *Key words:* time series studies, randomness, trends, outpatient clinics

Introducción

Una serie de tiempo es el conjunto de datos numéricos que se obtienen en períodos regulares a través del tiempo; también se conoce como *series temporales* o *series cronológicas*. La unidad de tiempo puede ser: hora, día, mes, trimestre, año o cualquier período considerado de interés.

La suposición básica que sustenta el análisis de series temporales es que los factores que han ocasionado patrones o tendencias en el pasado y en el presente continuarán haciéndolo, más o menos, de la misma forma en el futuro. Por lo tanto, los principales objetivos del análisis de series temporales consisten en identificar y aislar tales factores de influencia con el propósito de realizar proyecciones, es decir, estimar los valores futuros de la variable en estudio.

Las series de tiempo poseen características diferentes de la mayoría de métodos estadísticos utilizados para tratar temas referentes al área de la salud. Por ejemplo, en la estadística paramétrica se supone que los datos provienen de observaciones independientes del fenómeno de interés, situación poco probable en datos provenientes de series de tiempo. El incremento o disminución de algunas enfermedades puede afectarse por el incremento o disminución de la población (desplazamiento, violencia), por lo tanto, a medida que se avanza en el tiempo, es probable que varíe el número de personas afectadas, dado el aumento de la población o el desarrollo científico y tecnológico, es decir, los datos están influenciados por el pasado. Además, puede suceder que algunos brotes o epidemias sean recurrentes en períodos similares cada año, debido a cambios climáticos, cambio de dietas alimenticias, situaciones folclóricas y culturales, entre otras.

Es importante mencionar que existe el enfoque de los modelos ARIMA (procesos autorregresivos integrados con promedios móviles), que requieren de una mayor conceptualización estadística. Sin embargo, lo expuesto en este artículo también es útil como metodología para identificar si esos procesos son estacionarios en media y en varianza, lo cual se requiere para identificar el orden de los parámetros utilizados en dichos métodos.

Además, en eventos no muy cambiantes, algo común en el área que nos ocupa, como la mortalidad y morbilidad en suicidios, diferentes tipos de cáncer o accidentalidad por diversos motivos, los datos —en casos e incluso en tasas— no son volátiles y, por lo tanto, los métodos clásicos se ajustan de una manera adecuada, como lo demostró un estudio realizado para la Secretaría Distrital de Santafé de Bogotá en el año 2002.

La serie con la cual se desarrolla la metodología propuesta, así como algunos elementos que se deben considerar previamente para garantizar un buen análisis, es el número de consultas externas realizadas en la Institución

Prestadora de Servicios de Salud de la Universidad de Antioquia (IPS Universitaria) entre abril del 2001 y septiembre del 2005. Los datos se obtuvieron mensualmente.

En la IPS Universitaria, la consulta externa —conformada por consulta médica general, especializada, subespecializada, atención domiciliaria y procedimientos terapéuticos— constituye uno de los servicios de mayor demanda dentro de la asistencia prestada por dicha entidad. La población a la cual se le brinda el servicio esta conformada por los afiliados al programa de salud (empleados de la universidad) y estudiantes; desde hace algunos años también se presta el servicio a usuarios de otras empresas promotoras de salud (EPS) y a personas particulares. Esta ampliación de cobertura ha generado una serie de cambios en la dinámica de la prestación de los servicios, en especial de consulta externa, por lo cual se quiere observar qué tanto ha cambiado el comportamiento de dicha variable a través del tiempo. El software utilizado para el caso práctico fue el SPSS versión 11.0.

Consideraciones previas sobre una serie de datos

Consistencia

Aun considerando que algunas series tienen problemas de calidad del dato como subregistro o muestreo insuficiente, el hecho de que la manera de recolectar el dato sea invariable en el tiempo puede garantizar una buena descripción de la serie. Todo esto tiene que ver con los mecanismos de notificación, los cuales pueden cambiar la forma de capturar la información.

Comparabilidad

Han de tenerse presente los posibles cambios que se originan a través del tiempo, por ejemplo, cuando se incluye la variable número de población es necesario obtener una medida de relación respecto de ese cambio de volumen.

Pasos en el análisis de una serie de tiempo: enfoque clásico

Determinar si la secuencia de datos forma una serie no aleatoria

Una manera empírica de identificar aleatoriedad es mediante el gráfico de secuencia: si muestra una tendencia hacia arriba o hacia abajo indica no aleatoriedad, como en este caso. Para verificar dicha percepción se realiza la prueba de Rachas, determinando la dependencia o no de la serie. En caso de que no sea aleatoria, la serie se conoce como ruido blanco o caminata aleatoria y, por ende, no tiene las componentes de tendencia, estacionalidad ni ciclos.

Para el caso práctico el valor del sig. en la prueba de Rachas es $0,000 > 0,05$, por lo tanto se concluye que la serie no es aleatoria.

Análisis exploratorio de datos

Es vital determinar la existencia de valores atípicos, extremos y valores perdidos antes de tratar de modelar la serie. Para ello se tienen las siguientes herramientas:

- gráfico de secuencias;
- gráfico de caja y sesgo;
- cálculo de estadísticas descriptivas.

La figura 1 muestra el gráfico de secuencia para la variable consultas externas. En ella se aprecia que la serie presenta una tendencia negativa durante el período analizado, más pronunciada en el año 2003, y a partir de ese año, el número de consultas se mantiene estable. No se identifican valores perdidos ni atípicos o extremos. Para el propósito que se busca, es indiferente trabajar con casos o tasas, debido a que el comportamiento en ambos casos es similar. La estacionalidad no es fácil de intuir según la figura, por lo tanto, se requiere de métodos para ello (índices estacionales).

Por otro lado, es importante calcular algunas medidas resumen con el fin de confrontar lo encontrado en el gráfico y, además, profundizar en el conocimiento de la serie en aspectos importantes como la variabilidad de los datos y su forma.

Tabla 1. Medidas de resumen para la variable *consultas externas*; abril 2001-septiembre 2005

N	Válidos Pérdidos	54 0
Media		6177.56
Mediana		6110.50
Moda		2853 ^a
Desv. tip.		1746.24
Mínimo		2853
Máximo		9609

- a. Existen varias modas. se mostrará el menor de los valores

En promedio se presentaron aproximadamente 6.177 consultas externas por mes, con variabilidad moderadamente alta (1.746), es decir, el número de consultas externas varía aproximadamente en 28% (coeficiente de variación) mes a mes. En el 50% de los meses se presentaron 6.110 consultas externas o menos. El menor número de consultas externas realizadas en la ips en un mes —y a su vez uno de los valores más frecuentes— es de 2.853.

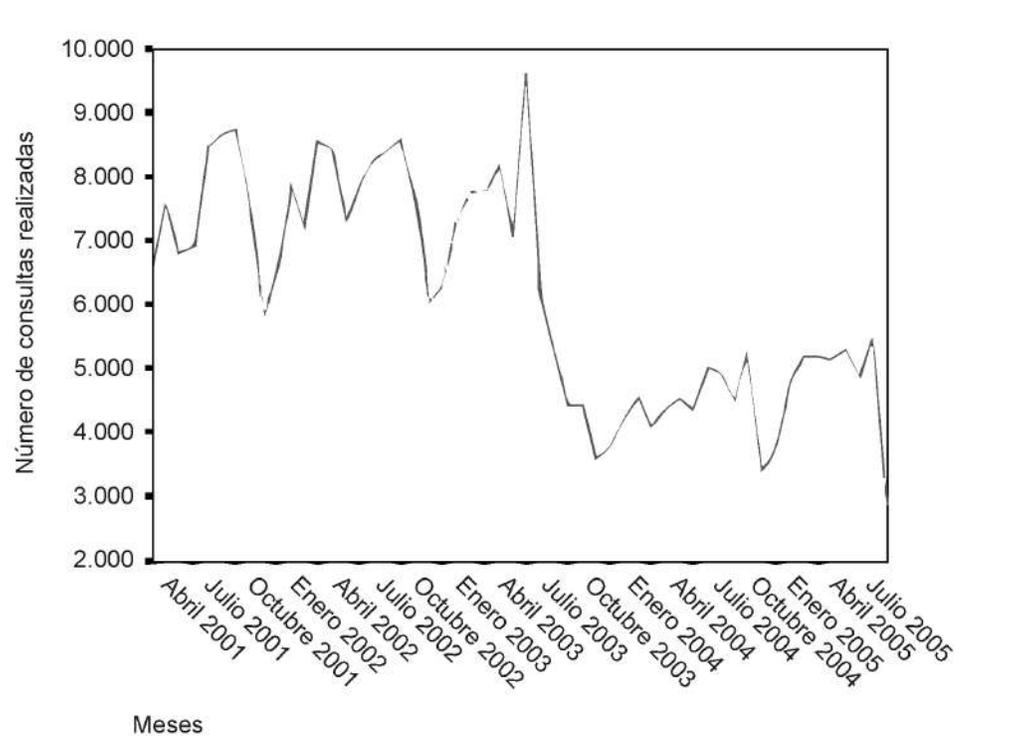


Figura 1. Consultas externas realizadas por la ips Universitaria entre abril del 2001 y septiembre del 2005

Identificación de las componentes de una serie

Con el fin de identificar si se realizan pruebas paramétricas o no paramétricas para determinar la presencia de la tendencia, es necesario validar los siguientes supuestos: aleatoriedad y normalidad, teniendo en cuenta el período estacional de la serie, para este caso se trabaja anualmente, es decir, se validarán los supuestos para cada año. En algunos casos no es necesario comparar los promedios para cada año, ya que se puede partir la serie en varios tramos y comparar dichos promedios, esa partición se puede hacer teniendo en cuenta el gráfico de secuencias. Para la serie que se está analizando los resultados arrojados por SPSS son:

Tabla 2. Significancia de aleatoriedad y anormalidad en una serie de tiempo

Periodos	Aleatoriedad Significancia	Normalidad Significancia
2001	1,000	0,569
2002	0,364	0,240
2003	0,034	0,858
2004	1,000	0,927
2005	1,000	0,010

Con respecto a la normalidad, se observa que solo en el 2005 el número de consultas externas no se comporta de manera normal, por lo tanto, se puede aplicar la prueba de ANOVA (análisis de varianza). En cuanto a la aleatoriedad, el número de consultas en el 2003 muestra una tendencia persistente a la disminución y, por ende, no cumplió el supuesto de aleatoriedad, sin embargo esto no impide la validez de la prueba de ANOVA.

Tendencia

Movimiento general o persistente a través del tiempo, movimiento hacia arriba o hacia abajo a largo plazo. Se presenta por cambios en tecnología, avances científicos, incrementos en la población, etc. Para identificar su existencia en una serie, se realiza la prueba de ANOVA con respecto a los años para determinar si los promedios de consultas externas por año presentan o no diferencias.

Para el caso práctico, la significancia en la prueba de ANOVA con respecto a los años es $0,000 < 0,05$, lo que significa que al menos uno de los promedios de consultas externas por año es diferente, lo cual confirma que la serie presenta tendencia tal como se había observado en la figura 1.

Estacionalidad

Fluctuaciones periódicas bastante regulares que se presentan dentro de cada período de 12 meses, año tras

año. Se presenta por condiciones climatológicas, costumbres sociales y religiosas, entre otras. Para identificar su existencia en una serie, se puede realizar la prueba de ANOVA con respecto a los meses para determinar si los promedios de consultas externas por mes presentan o no diferencias; o probar los índices estacionales para observar si existen valores que se encuentren por fuera del rango de 90-110.

Tabla 3. Índices estacionales en una serie de tiempo

Período	Seasonal index (*100)
1	108.434
2	109.102
3	101.940
4	102.519
5	107.203
6	109.353
7	105.347
8	100.074
9	74.481
10	80.496
11	98.024
12	103.028

De acuerdo con los índices estacionales, se determina que la serie sí presenta estacionalidad, pues existen índices que se encuentran por fuera del rango 90 y 110, el cual es un criterio empírico pero reconocido en la literatura de series de tiempo para determinar la estacionalidad de una serie. Así, por ejemplo, en los meses de septiembre se presenta una marcada disminución de 25,519% con respecto al promedio de consultas mensuales, mientras que en los meses de febrero y junio se incrementa el número de consultas externas.

Cíclica

Oscilaciones repetidas hacia arriba y hacia abajo, no con la misma duración ni por los mismos períodos. Por lo general se presenta por decisiones políticas que impactan los eventos a largo plazo, por lo tanto, en el área de la salud, es poco probable encontrar series que permitan un análisis de ciclos. En este caso se tienen datos para 4 años, lo que indica que esta serie no presenta el componente cíclico.

Irregular

Comportamiento no sistemático. Fluctuaciones aleatorias sin ningún patrón definido. Todas las series la poseen.

Se concluye que la serie presenta las componentes de tendencia, estacionalidad y variación irregular. Con estos elementos se puede identificar el modelo.

Identificación del modelo

Existen diferentes procedimientos para encontrar un modelo que permita pronosticar los valores de una serie: métodos de suavización, modelos mínimos cuadráticos y modelos ARIMA, entre otros. Para el caso que nos ocupa, el número de consultas externas no es muy cambiante, por lo tanto, se utilizan los modelos de suavización.

Suavizado exponencial simple

Es una técnica útil para suavizar una serie de tiempo; proporciona una impresión de los movimientos o patrones generales de los datos. Además, puede utilizarse para obtener predicciones a corto plazo. Esta técnica supera al método de promedios móviles en el sentido de que no elimina valores. Además, les da mayor ponderación a los últimos valores de la serie y menos peso a los primeros.

Esta herramienta de análisis predice un valor basándose en el pronóstico del período anterior. El suavizado exponencial simple es apropiado cuando las series a predecir son no estacionales y no tienen una tendencia constante ni ascendente ni descendente.

Suavizado exponencial de Holt

Es una extensión del planteamiento de suavizado exponencial, la diferencia radica en que el procedimiento de suavizado exponencial proporciona una visión de los movimientos a largo plazo sin tener en cuenta la estacionalidad ni la tendencia, mientras que el Holt permite pronosticar la tendencia. Para utilizar el método de Holt para predicción, suponemos que todos los movimientos de tendencia futuros continuarán a partir del último nivel suavizado.

Suavizado exponencial de Winter

Modela el nivel general de la serie, la tendencia y la estacionalidad, por ello se requiere estimar el componente estacional por medio de lo que se conoce como índices estacionales, los cuales sirven para ajustar el modelo. Para encontrar el modelo adecuado de la serie en estudio, se recurrió a esta técnica dado que la serie presenta estacionalidad y tendencia. En este caso se requieren tres constantes de suavización, a saber:

The SSE is: alpha gamma delta SSE
.6000000 .0000000 .2000000 38822536.774

Donde alpha modela el nivel general de la serie, gamma la tendencia y delta la estacionalidad. SSE es la suma de cuadrados del error, luego el modelo que tenga un menor valor de SSE es el que se prefiere. Se corrieron varios modelos (constantes diferentes) y el que mejor se ajustó a los datos fue: alpha = 0,6, gamma = 0,0 y delta = 0,2.

Análisis de residuales

Es necesario validar los supuestos de aleatoriedad y normalidad de los residuales para determinar si el modelo

encontrado es o no el adecuado para realizar pronósticos. La aleatoriedad de los residuales se puede determinar por medio de la prueba de Rachas y la normalidad, mediante la prueba de Smirnov Kolmogorov.

En el caso de los residuales de la serie de consultas externas, los resultados son los siguientes:

Aleatoriedad	Normalidad
Prueba de Rachas significativamente mayor de 0,05, por lo tanto, no se cumple.	Prueba de Smirnov Kolmogorov significativamente menor de 0,05, por lo tanto no se cumple.

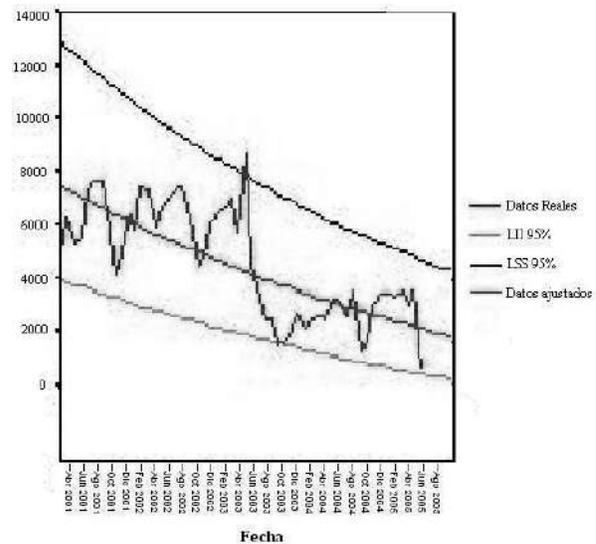


Figura 2. Proyección método exponencial, consultas externas, 2001-2005

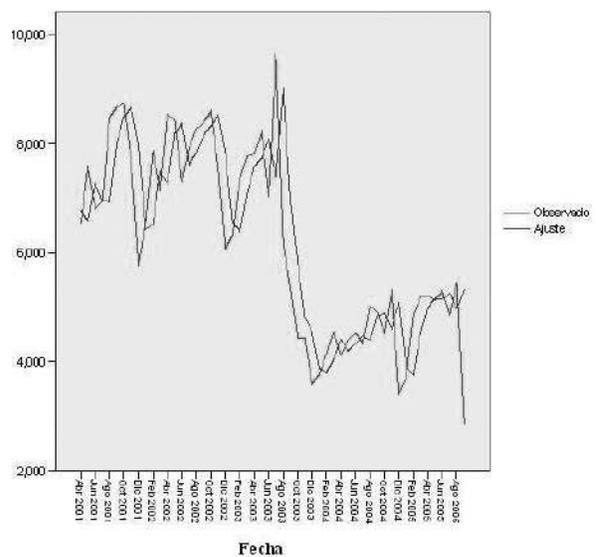


Figura 3. Proyección método suavización exponencial, consultas externas, 2001-2005

Tabla 4. Pronósticos para los modelos *suavizado exponencial* y *método exponencial*

Fecha	Suavizado Exponencial	Método exponencial
Octubre 2005	3.589	4.035
Noviembre 2005	3.400	3.979
Diciembre 2005	2.468	3.924
Enero 2006	2.599	3.869
Febrero 2006	3.071	3.816
Marzo 2006	3.140	3.763
Abril 2006	3.144	3.710
Mayo 2006	3.164	3.659
Junio 2006	2.891	3.608

En la prueba de Rachas se cumple aleatoriedad para los residuales, pero la normalidad no; sin embargo, si se realiza la prueba por año, estos se comportan de manera normal. Dado lo anterior, se debe buscar otro modelo mínimo cuadrático, como los ofrecidos por el programa SPSS, en que se destacan logarítmico, logístico y exponencial, entre otros.

Luego de los ensayos de rigor, se encontró que el ajuste exponencial, diferente al suavizado exponencial, permite una buena aproximación a los datos de consulta externa. Los residuales de este último modelo cumplen aleatoriedad y normalidad por año. En ambos casos se cumplieron los mismos supuestos. Otros modelos, por lo tanto, podrán cumplir las mismas condiciones, e incluso mejores. Por lo tanto, es prudente tener en cuenta a los expertos en la temática específica para darles contexto a las proyecciones meramente estadísticas.

Pronósticos

Con la ayuda del programa SPSS, observamos la forma de presentar la información según el procedimiento. Para el ajuste exponencial, se observa cómo la proyección va disminuyendo de una manera leve; además, se presentan dos bandas construidas por los límites de los intervalos de confianza, dando un rango de posibilidades donde deberá estar la demanda de consultas externas en la IPS Universitaria.

Para el suavizado exponencial, se aprecia que el pronóstico está fuertemente influenciado por el último mes (septiembre 2005), sin embargo, el modelo parece ajustar bien los datos. La situación es que a medida que lleguen nuevos datos, estos se introduzcan en la base de datos

para robustecer el modelo. En cualquier caso, no es prudente realizar proyecciones más allá de los 6 meses.

Como ya se mencionó, el darles contexto a los valores encontrados es fundamental.

Conclusiones

- Existen diferentes metodologías para abordar una serie de tiempo, sin embargo, se sugiere que la selección del método sea lo más práctica y entendible posible.
- La ayuda de los programas de software permite explorar e implementar el procedimiento de ensayo/error, es decir, permite ensayar varias hipótesis para luego escoger la que cumpla un mayor número de supuestos y, además, entregue valores razonables para el evento en estudio.
- La discreción de los expertos en el tema no se puede dejar de lado; es más, deben refinar los pronósticos entregados por los procedimientos estadísticos y darle la cuota de racionalidad al número.

Referencias

1. Bello L. Estadística como apoyo a la investigación. 1ª ed. Medellín: L. Vieco e Hijas; 2005.
2. Aguirre J. Introducción al Tratamiento de Series de Tiempo, Aplicación a las Ciencias de la Salud. Madrid: Diaz de Santos; 1994. p. 90-115
3. Newbold P. Estadística para los Negocios y la Economía. 4ª ed. España: Prentice Hall; 1998. 752 p.
4. Díaz A. Diseño estadístico de experimentos. Medellín: Universidad de Antioquia; 1999. 347 p.
5. Manson R. Lind D. Estadística para Administración y Economía. 11ª ed. Bogotá: Alfaomega; 1992. 830p.
6. Molinero LM. Análisis de series temporales. Asociación de la Sociedad Española de Hipertensión. (Internet) 2004 (25 de octubre de 2005) Disponible en: <http://www.seh-lelha.org/tseries.htm>.
7. Biblioteca Virtual de la Salud Colombia. Secretarías De Salud Departamentales/Municipales. [Sitio de Internet] Colombia: Biblioteca Virtual de la Salud; 2005 [17 de noviembre de 2005] Disponible en: <http://www.col.ops-oms.org/iah/lisnalsds.htm>.
8. Areyano M. Introducción al Análisis Clásico de Series de Tiempo. 5campus.com [Internet]2001[17 de noviembre de 2005] Disponible en: <http://ciberconta.unizar.es/LECCION/seriest/000F2.HTM>
9. Bolivar M. Estadística. [Monografía en Internet]1997 [25 de noviembre de 2005] Disponible en: <http://www.monografias.com/trabajos14/estadistica/estadistica.shtml#ma>.