

## CIFRAS DE HOY

# Análisis de la evolución y tendencia de las tasas de hepatitis A en Chile desde 1975 a 2002

GABRIEL CAVADA<sup>(1)</sup>, ROSA MONTAÑO<sup>(1)</sup>,  
SERGIO ALVARADO<sup>(1)</sup>, MARÍA TERESA VALENZUELA<sup>(1)</sup>

Las tendencias temporales de las tasas de hepatitis A en Chile, resultan ser erráticas, tanto si se analizan en forma global como separadas por rango etáreo. Esta tendencia no regular en el tiempo, hace muy complejo intentar predecir una futura tasa mediante métodos tradicionales de análisis de series temporales tales como modelos ARIMA o ARMA entre otros. La Figura 1 muestra la tendencia temporal de la tasa global y desagregadas por rango etáreo de hepatitis A en Chile.

Un análisis muy simplista nos inclinaría a pensar que dichas evoluciones son constantes en el tiempo, resultado que puede obtenerse al descartar una tendencia creciente o decreciente en el tiempo, mediante un análisis de regresión lineal. Por ejemplo, si se opta por estudiar la tasa para el grupo etáreo de 0 a 4 años, el estudio de tendencia lineal arroja (Tabla 1) una tendencia estadísticamente significativa a la baja ( $p = 0,007$ ), conclusión que no es

satisfactoria, puesto que no considera los quiebres de la tendencia en determinados años.

La Figura 2 induce a pensar que la tasa de hepatitis A, en este primer rango etáreo, tenderá a crecer el año 2004 y 2005. Sin embargo, el modelo estimado en la Tabla 1 predice una tasa de 84,2 y 79,4, respectivamente, predicciones muy discutibles si se analizan las Figuras 2 y 3.

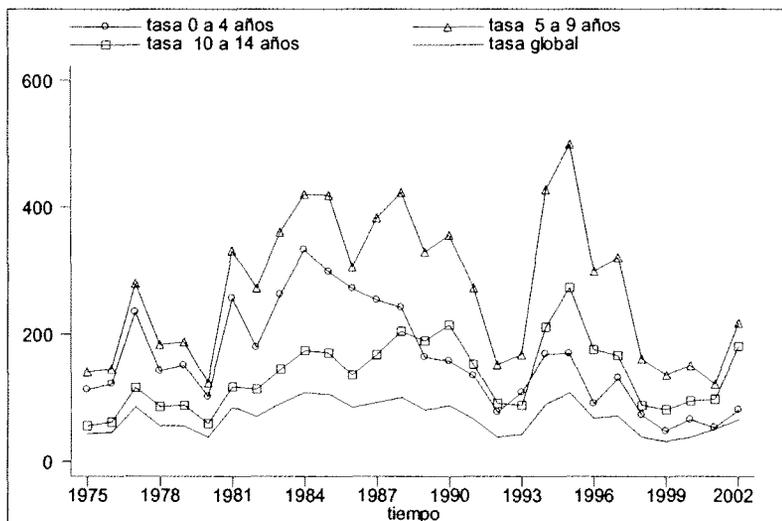
El análisis de la Figura 3, induce a modelar la evolución de esta tasa, mediante un modelo mínimo cuadrático de segundo grado, las estimaciones de dicho modelo se muestran en la Figura 4.

Este modelo predice tasas para el rango etáreo de 0 a 4 años para los años 2004 y 2005 de 186,0 y 263,5, dada la tendencia al alza de dichas tasas estas predicciones resultan ser más creíbles que las entregadas por el modelo lineal, pero igualmente discutibles, pues no consideran las variaciones y quiebres que se observan a través del tiempo, o dicho de otro modo, no se

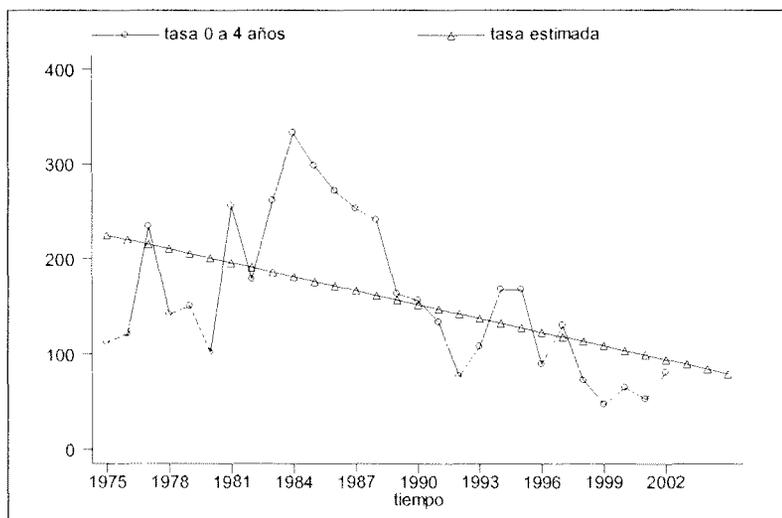
Tabla 1

Source	SS	df	MS	Number of obs = 28 F( 1, 26) = 8.64 Prob > F = 0.0068 R-squared = 0.2493 Adj R-squared = 0.2205 Root MSE = 70.709		
Model	43178.5763	1	43178.5763			
Residual	129992.134	26	4999.69745			
Total	173170.71	27	6413.73			
t0	Coef.	Std. Err.	t	P> t	[95% Conf.	Interval]
tiempo	-4.861439	1.654256	-2.94	0.007	-8.26181	-1.461069
cons	9826.569	3289.514	2.99	0.006	3064.876	16588.26

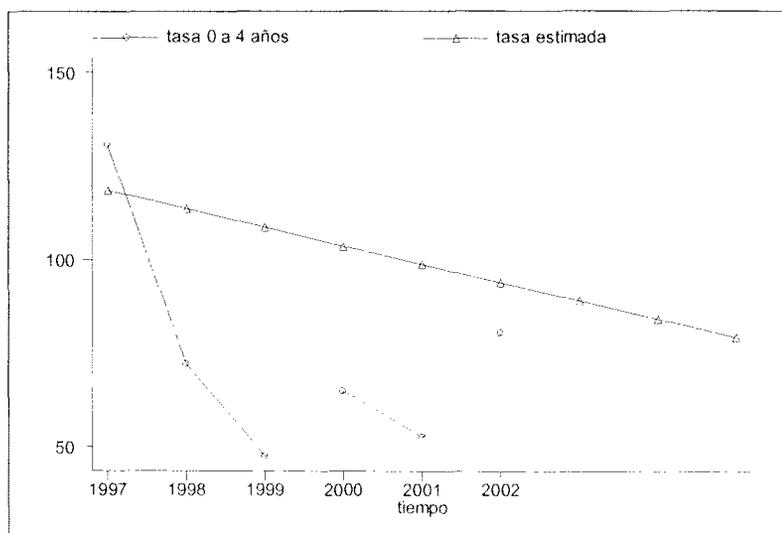
<sup>(1)</sup> Escuela de Salud Pública, Universidad de Chile. gcavada@med.uchile.cl



**Figura 1.** Evolución de las tasas de hepatitis A. Chile 1975 a 2002.



**Figura 2.** Evolución de las tasas de hepatitis A, 0 a 4 años. Chile 1975 a 2002.



**Figura 3.** Evolución de las tasas de hepatitis A, 0 a 4 años. Chile 1997 a 2002.

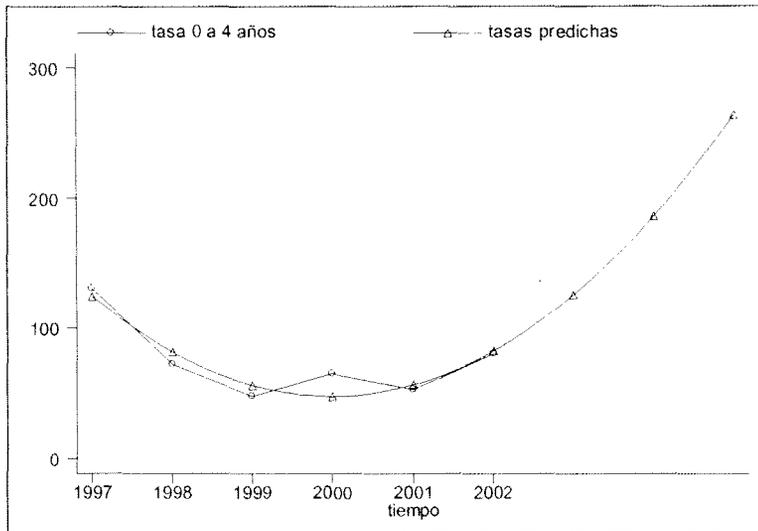


Figura 4. Tasas 0 a 4 años de hepatitis A con ajuste cuadrático.

ha considerado la historia de las evoluciones.

Una metodología alternativa y novedosa para la modelación de la evolución de estas tasas es la técnica multivariada no paramétrica “Multivariate Adaptive Regression Splines” (MARS).

### LOS MODELOS MARS

Jerome H. Friedman (1991) propone una novedosa metodología llamada MARS. Esta intenta construir un modelo de Regresión no-lineal, basado en un producto de funciones base spline. MARS es una generalización de la Recursive Partitioning Regression (PR), la que divide al espacio de las variables predictoras en diferentes subregiones, intentando una aproximación local en cada subregión (Lewis and Stevens, 1991; Friedman and Roosen, 1995).

Produce un modelo para la respuesta en estudio, que automáticamente selecciona las variables que aparecen en la ecuación final. También indica lo complejo de la relación entre la respuesta y cada variable predictora, a partir del número de funciones bases usadas (Silva et al., 2001).

Para una variable respuesta  $y$ , la variable predictora  $x$  y el error aleatorio  $\varepsilon$  se asume un modelo del tipo  $y = f(x) + \varepsilon$

$$B_k^{(q)}(x) = \begin{cases} x^j & j = 0, \dots, q \\ (x - t_k)_+^q & k = 1, \dots, K \end{cases}$$

Se seleccionan  $K$  nodos  $x = t_k, k=1, \dots, K$ , el cual define  $K+1$  regiones sobre el rango de  $x$ . Se asocia a cada nodo la función spline lineal, generando una familia de funciones bases.

Para la aproximación de orden  $q$ , se estima la función,

$$\hat{f}_q(x) = \sum_{k=0}^{K+q} a_k B_k^{(q)}(x)$$

generalmente el orden de la función spline que se tome debe ser menor o igual a tres, para que la función y sus  $q-1$  derivadas sean continuas en todas partes. Esta restricción y el uso de polinomios en cada subregión produce funciones suavizadas y ajustadas.

Ya que cada polinomio tiene grado  $q$ , tenemos  $q+1$  parámetros que deben ser ajustados usualmente por mínimos cuadrados.

Para evaluar este modelo, Friedman propone usar la estadística Generalized Cross Validation,

$$GCV = A \times \sum_i (y_i - \hat{f}_q(x_i))^2 / N$$

en donde  $A = (1-C(M)/N)^{-2}$  y  $C(M) = 1 + \text{traza}(B(B'B)^{-1}B')$ , el numerador es la falta de ajuste sobre los datos de entrenamiento y el denominador es un termino penalizado que refleja la complejidad del modelo.

## RESULTADOS

Al ajustar un modelo MARS con variable dependiente la tasa global de hepatitis A en Chile y como variable independiente el tiempo en años desde 1975 al 2002, se obtiene el siguiente modelo:

$$\text{Tasa total} = 87.04 - 5.42 * \text{máximo}(0, 1983 - \text{tiempo}) - 3.22 * \text{máximo}(0, \text{tiempo} - 1988.5) + \text{Error}$$

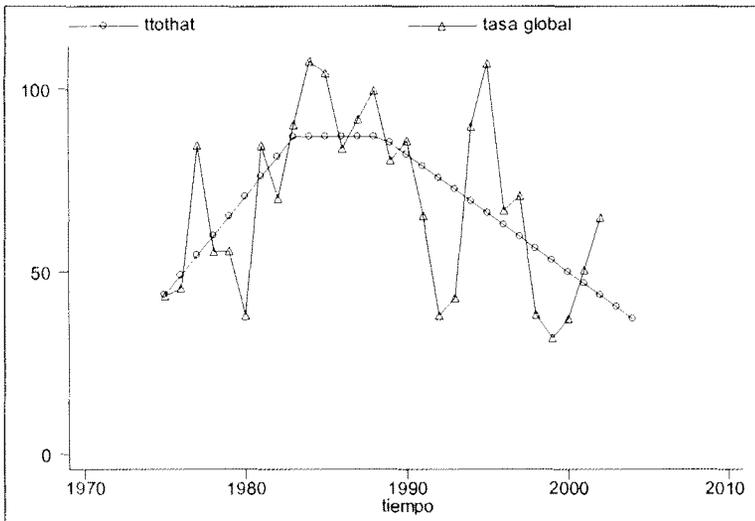


Figura 5. Tasa total observada y ajustada por MARS.

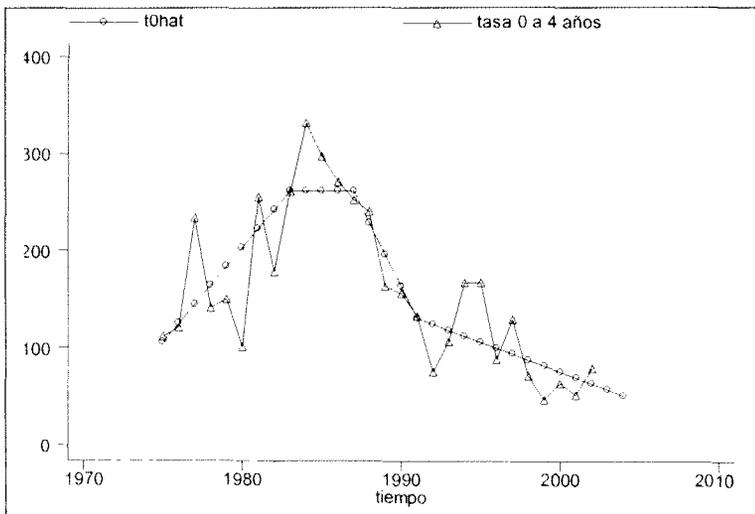


Figura 6. Tasa observada y ajustada por MARS, 0 a 4 años.

Esta expresión indica que tanto en 1983 y a mediados de 1988 se produjeron cambios en el comportamiento de la tendencia en la tasa total de hepatitis A. La Figura 5 muestra dichos quiebres y una tendencia a la baja en dichas

tasas, de hecho predice una tasa global para el año 2004 de 37,1.

Idéntico análisis para la evolución de la tasa en el grupo etáreo de 0 a 4 años entrega el siguiente modelo:

$$t_0 = 261.916 - 19.425 * \text{máximo}(0, 1983 - \text{tiempo}) + 26.580 * \text{máximo}(0, \text{tiempo} - 1991) - 32.666 * \text{máximo}(0, \text{tiempo} - 1987) + \text{Error}$$

Este modelo señala quiebres en la evolución de esta tasa en los años 1983, 1987 y 1991.

La Figura 6 muestra la tasa observada y predicha por MARS para el grupo etáreo de 0 a 4 años. Y predice tasas globales de hepatitis A para el año 2004 52,1.

## CONCLUSIONES

A pesar de la contradicción presentada por las predicciones para el año 2004, según los modelos aplicados, sería más razonable pensar

en que los modelos MARS tienen más credibilidad, dado este reconocimiento histórico que hacen en la evolución de las tasas.

Un mero juicio visual indicaría que la tasa de hepatitis A para el grupo etáreo de 0 a 4 años, estaría mejor predicha por la metodología MARS.

Dado el comportamiento errático de la evolución de las tasas estudiadas, hacer predicciones a más largo plazo sería una aventura demasiado peligrosa.