CALCULO DE LA POBLACION FUTURA DE ALBACETE. UN NUEVO METODO DE AJUSTE DE LA FUNCION LOGISTICA

Francisco Díaz Martínez

Francisco Díaz Martínez.
Dr. en Ciencias Económicas
Catedrático de Estadística de E.U.

1. INTRODUCCION

En el pasado se intentó frecuentemente la matematización de ciertos comportamientos demográficos que explicasen la evolución temporal de una población. La matematización a la que se alude no ha sido otra cosa que el adaptar a un lenguaje continuo aquellos problemas cuya base pertenecen al campo de las magnitudes discretas, lo que hace que la utilización de los elementos del cálculo diferencial sean un instrumento útil y operativo.

El análisis estadístico de la población tuvo, sin duda, su origen en los aritméticos políticos ingleses Graunt, Petty y Halley. Los dos primeros en el siglo XVII consideraron ya, que la población tiende a multiplicarse en progresión geométrica, adelantándose en más de un siglo a Malthus, por lo que no es de extrañar, que se generalizase la utilización de funciones exponenciales para el cálculo de poblaciones futuras.

No se pretende entrar en la polémica de si la logística tiene o no valor predictivo, pero lo que si está fuera de toda duda es que explica aceptablemente la evolución de una población en un determinado intervalo de tiempo pasado. Más aún, si dentro de determinados periodos o "ciclos culturales" se ajusta a una logística o un arco de logística cada ciclo, yuxtaponiendo convenientemente dichos arcos, se obtienen resultados muy aproximados a la realidad observada, e inclusive proporciona estimaciones admisibles siempre que no se modifiquen las condiciones del ciclo. Véase Alcaide [1], [2], [3].

(1) Alcaide Inchausti, A. "Nueva determinación de la curva logística de la población de España". *Economía Política*. Sept.-Dic. 1955.
(2) Alcaide Inchausti, A. "Estadística (Introducción). UNED, Madrid 1974.
(3) Alcaide Inchausti, A. "La población de España en el periodo 1970-2000". *ICE*. Dic. 1974.

Una presentación de la teoría sobre los ciclos demográficos fue realizada en 1949 por Cowgill.[4]

La curva logística no sólo se ha utilizado para explicar la evolución pasada y futura de poblaciones humanas y de otros seres, sino que también se usa para determinar la tendencia de ciertas series de datos sociales o económicos; puede verse, a este respecto Mills.[5]

El presente artículo tiene por objeto la exposición y aplicación de un método de ajuste del modelo matemático más divulgado en la literatura demográfica, el de la población logística. Creemos necesario hacer una breve exposición de este método para justificar las aplicaciones que posteriormente se van a realizar. La logística, utilizada de forma predictiva sin la necesaria acotación de la función por ella definida la hace inaplicable para cálculos a largo plazo; sin embargo, para breves intervalos de tiempo o para interpolaciones puede proporcionar resultados francamente aceptables.

El nuevo método que aquí se aplica incluye un procedimiento debido a García Sestafe[6] y evita la subjetividad de la elección de una o de las dos asíntotas de la curva.

2. AJUSTE DE LA CURVA LOGISTICA MEDIANTE EI CONOCIMIENTO DE LA PENDIENTE

En los métodos usuales de ajuste se empieza por fijar con mejor o peor criterio una de las dos asíntotas de la logística a ajustar. En el método que se incluye a continuación, se deducen a partir de los datos las ecuaciones de ambas asíntotas.

Este método está basado, como su nombre indica, en la obtención de la pendiente es una generalización de un procedimiento, alguna vez aplicado para la logística de asíntota horizontal $y = 0$[7]

La idea directriz de este método es expresar la pendiente de la curva –en cada punto– como función, únicamente, de la ordenada en el punto, o sea

$$y' = f(y)$$

Encontrada tal relación, las posibles asíntotas $y = y_o$, $y = y_1$ serán tales que y_o e y_1 serán raíces de la ecuación $f_{(y)} = 0$.

---

(4) Cowgill, D. "The Theory of Population Growth Cycles". *American Journal of Sociology*. Tomo LV. 1949.
(5) MILLS, F. "Stadistical Methods...", Henry Holt and Company. 1955.
(6) García Sestafe, J.V. "La curva logística", INE, 1989.
(7) Chacón, E. "Curso de Estadística", Volumen I. Publicaciones de la Universidad de Deustto. Bilbao 1955.

Sea la logística de ecuación

$$y = c + \frac{a}{1 + b\,e^{-ht}} \qquad c = y_0, \quad c + a = y_1$$

donde $N_{(t)} = y$; tomando logaritmos neperianos

$$\ln(y - c) = \ln a - \ln(1 + b\,e^{-ht}) \qquad \ldots\ldots\ldots$$

y derivando

$$\frac{y'}{y - c} = \frac{b\,h\,e^{-ht}}{1 + b\,e^{-ht}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

y como

$$1 + b\,e^{-ht} = \frac{a}{y - c} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

se tiene

$$\frac{y'}{y - c} = \frac{h\left(\frac{a}{y - c} - 1\right)}{\frac{a}{y - c}} = h\left(1 - \frac{y - c}{a}\right)$$

de donde

$$y' = \frac{-h}{a}(y - c)^2 + h(y - c) = \frac{-h}{a}y^2 + h\left(\frac{2c}{a} + 1\right)y - hc\left(\frac{c}{a} + 1\right)$$

Haciendo

$$-\frac{h}{a} = \alpha\ ;\ h\left(\frac{2c}{a} + 1\right) = \beta\ ;\ -hc\left(\frac{c}{a} + 1\right) = \gamma \qquad [1]$$

y como

$$\left|\frac{(\alpha, \beta, \gamma)}{(a, h, c)}\right| = \frac{h^2}{a^2} \neq 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots$$

basta ajustar por mínimos cuadrados el polinomio

$$y' = \alpha\,y^2 + \beta\,y + \gamma$$

a la nube de puntos (y_i, y'_i) donde se suponen conocidos los valores y'.

A partir de las ecuaciones [1] resulta

$$\alpha\,c^2 + \beta\,c + \gamma = 0$$

ecuación que, si la logística es ajustable, se debe cumplir que

$$\beta^2 - 4\,\alpha\,\gamma > 0\,, \quad \alpha < 0, \quad \gamma < 0$$

luego

$$\alpha c^2 + \beta c + \gamma = 0$$

tiene en dicho caso dos raíces reales distintas y positivas c_1 y c_2 ($c_1 < c_2$).

Por otra parte, como

$$c_2 - c_1 = \frac{\sqrt{\beta^2 - 4\alpha\gamma}}{\alpha} = \sqrt{\frac{\beta^2}{\alpha^2} - 4\frac{\gamma}{\alpha}} = a$$

las dos asíntotas buscadas son $y = c_1$ (asíntota inferior) e $y = c_2$ (asíntota superior).

Si no se cumplen $\beta^2 - 4\alpha\gamma > 0$ ó α y β no son ambos negativos, la curva logística no es ajustable –al menos por este procedimiento– a la nube de puntos dada.

Aunque h es deducible de cualquiera de las ecuaciones [1] es preferible, para obtener los parámetros h y b escribir la logística en la forma

$$\frac{a}{y - c} - 1 = b\,e^{-ht} \rightarrow \ln\left[\frac{a}{y - c} - 1\right] = \ln b - ht$$

haciendo

$$Y = \ln\left[\frac{a}{y - c} - 1\right]$$

basta ajustar a la nube de puntos (t_i, Y_i) la recta $Y = A + B_t$, para tener

$$A = \ln b, \qquad B = -h$$

o sea

$$b = e^A, \qquad h = B$$

con lo cual conocen los cuatro parámetros buscados.

....

Para el ajuste de $y' = \alpha y^2 + \beta y + \gamma$, quedan por obtener los valores de y'_i en los puntos (t_i, y_i) que se utilizan para dicho ajuste. Para dar la mayor generalidad al método, como las fechas censales, salvo en ocasiones, no son equidistantes en el tiempo, ha parecido conveniente, para el referido cálculo, aplicar el método de las diferencias divididas que puede verse en[8] o en[9]. Una variante de la

(8) Alcaide Inchausti, A. y García Sestafe, J.V. Ampliación de Matemáticas para Economistas, UNED, Madrid, 1977.

(9) Spiegel, M. Finite Differences and Difference Equations, Schaum's 1971.

educación puede consultarse en [10].

Por cada tres puntos consecutivos, de abcisas t_{i-1}, t_i, t_{i+1}, se interpola una parábola de segundo grado; su pendiente en t_i proporciona una aproximación aceptable para y' en $t = t_i$.

Designando por

$$\phi_j \, (t_{i-1}, t_i \, , ..., t_{i+j-1}) \qquad ....$$

a la diferencia dividida de orden j correspondientes a los valores t_{i-1}, $t_i, \ldots, t_{i+j-1}$, se tiene

$$y = \phi_0 \, (t_{i-1}) + (t - t_{i-1}) \quad \phi_1 \, (t_{i-1}, t_i) + (t - t_{i-1}) \, (t - t_i) \; .... \; \phi_2 \, (t_{i-1}, t_i \, , t_{i+1})$$

de donde, derivando y haciendo $t = t_i$, se deduce

$$y'_i = (t_i - t_{i-1}) \quad \phi_2 \, (t_{i-1}, t_i \, , t_{i+1}) + \quad \phi_1 \, (t_{i-1}, t_i)$$

3. APLICACION DE LA CURVA LOGISTICA A LA POBLACION DE LA PROVINCIA DE ALBACETE

Para la aplicación de la curva logística a la población de la provincia de Albacete se han elegido los Censos de Población de los años que se indican en el siguiente cuadro. La población que figura es la de hecho; todos los censos están referidos al 31 de diciembre de cada año. El valor de t que figura es el que resulta de tomar

$t_{1860} = 0$

AÑO	HABITANTES (EN MILES)	t
1857	201,1	-3
1860	206,1	0
1877	219,1	17
1887	229,5	27
1900	237,9	40
1910	264,7	50
1920	291,8	60
1930	332,6	70
1940	374,5	80
1950	397,1	90

(10) Keyfitz, N. Introduction to the Mathematics of Population, Addison Wesley, 1968.

El primer paso a realizar es el cálculo de y' mediante la obtención de la diferencias divididas

t_i	Δt_i	$f(t_i)$	$f_1(t_{i-1}, t_i)$	$f_2(t_{i-1}, t_i, t_{i+1})$
-3		201,1		
	3		1,6667	
0		206,1		-0,0451
	17		0,7647	
17		219,1		0,0102
	10		1,0400	
27		229,5		-0,0171
	13		0,6461	
40		237,9		0,0884
	10		2,6800	
50		264,7		0,0015
	10		2,7100	
60		291,8		0,0686
	10		4,0800	
70		332,6		0,0055
	10		4,1900	
80		374,5		-0,0965
	10		2,2600	
90		397,1		

y aplicando

$$f'(t_i) = f_1(t_{i-1}, t_i) + (t_i - t_{i-1})\, f_2(t_{i-1}, t_i, t_{i+1})$$

se obtienen los valores de $f'(t_i)$

$$f'(t_1) = 1{,}6667 + 3\,(-0{,}0451) = 1{,}5314$$

$$f'(t_2) = 0{,}7647 + 17\,(0{,}0102) = 0{,}9380$$

$$f'(t_3) = 1{,}0400 + 10\,(-0{,}0171) = 0{,}8687$$

$$f'(t_4) = 0{,}6461 + 13\,(0{,}0884) = 1{,}7957$$

$$f'(t_5) = 2{,}6800 + 10\,(0{,}0015) = 2{,}6950$$

$$f'(t_6) = 2{,}7100 + 10\,(0{,}0685) = 3{,}3950$$

$$f'(t_7) = 4{,}0800 + 10\,(0{,}0055) = 4{,}1350$$

$$f'(t_8) = 4{,}1900 + 10\,(-0{,}0965) = 3{,}2250$$

Conocidos los valores de $f'(t_i) = y'_i$, basta ajustar a la nube de puntos (y_i, y'_i) la parábola

$$y'_i = \alpha y_i^2 + \beta y_i + \gamma$$

obteniéndose

$$\alpha = -1{,}560186 \cdot 10^{-4}$$

$$\beta = 0{,}107251$$

$$\gamma = -14{,}77852$$

Cumpliéndose que

$\beta^2 - 4\alpha\gamma > 0$ y que α y γ son ambos negativos, la ecuación

$$\alpha c^2 + \beta c + \gamma = 0$$

admite dos raíces reales positivas y distintas

$$c_1 = 190{,}6897 \quad \text{y} \quad c_2 = 496{,}7379$$

que proporciona las ecuaciones de las dos asíntotas

$y = c_1$, $y = c_2$; como $c_2 - c_1 = a$, se obtiene a = 306'0482

Para determinar b y h se introduce la variable

$$Y_i = \ln\left[\frac{a}{y_i - c} - 1\right]$$

donce $c = c_1$ (asíntota inferior)
dicho cambio de variables proporciona

t_i	Y_i
0	2,937041
17	2,279565
27	1,929454
40	1,701590
50	1,142696
60	0,706495
70	0,145512
80	-0,407935

y se ajusta una recta a la nube de puntos (t_i, Y_i), obteniéndose

$$Y = -0'04069901 + 3'05436 = Bt + A$$
$$(-17'6209)\ (26'4362)$$

donde los números entre paréntesis son los respectivos valores de la t de Student, siendo F = 310'49537 y el valor del coeficiente de correlación

$\rho = -0'9904759$

Luego $h = B = -04069901$ y $b = e^A = 21'207608$

La ecuación logística será

$$y = 190{,}6897 + \frac{306{,}0482}{1 + 21{,}207608 \cdot e^{-0{,}04069901\,t}}$$

A continuación se estudia la bondad del ajuste de la logística; para ello se va a tomar como medida el denominado "índice de bondad"(11) que se define

$$I = 1 - \frac{\sum (y_i - y_i^*)^2}{\sum y_i^2 - n\,\bar{y}^2}$$

donde por y' se denota la estimación de y, I, como es sabido, indica un ajuste tanto mejor cuanto más próximo es a la unidad.

y^*_i	y_i	% de error
204,5	206,1	0,7
217,0	219,1	0,9
228,6	229,5	0,3
249,9	237,9	-4,8
271,8	264,7	-2,6
298,3	291,8	-2,1
328,1	332,6	1,3
359,1	374,5	4,2

(11) García España E. y Sánchez Crespo, J.L. "Estadística Descriptiva". INE. Madrid 1961.

En el presente caso se obtiene

$$I = 1 - \frac{648,88343}{605835,02 - 8.\ 72643,726} = 1 - \frac{648,88343}{24685,212} = 0,9737136$$

A partir de 1950 la población de la provincia de Albacete empieza a disminuir por lo que no es posible explicar su comportamiento mediante una función logística. Es a partir de 1975 cuando inicia la recuperación después de haber estado sometida a la fuerte sangría que ha supuesto durante más de 25 años la emigración.

En el periodo 1975-1988 el comportamiento de la población de Albacete se puede explicar mediante una función logística.

El valor de t para la obtención de las diferencias divididas que se ha utilizado es $t_{1981} = 0$.

AÑO	t_i	Δt_i	$f(t_i)$	$f_1(t_{i-1}, t_i)$	$f_2(t_{i-1}, t_i, t_{i+1})$
1975	-5,1667		331,4		
		5,1667		0,599996	
1981	0		334,5		0,094426
		5		1,56000	
1986	5		342,3		0,490000
		1		4,500000	
1987	6		346,8		-1,600000
		1		1,300000	
1988	7		348,1		

Siendo

$$f'(t_1) = 0,0599996 + 5,1667\ (0,0944263) = 1,087868$$

$$f'(t_2) = 1,56 + 5\ (0,49) = 4,01$$

$$f'(t_3) = 4,5 + 1\ (-1,6) = 2,9$$

La parábola que se ajusta a la nube d puntos (y_i, y'_i)

y_i	y'_i
334,5	1,087868
342,3	4,010000
346,8	2,900000

tiene por coeficientes

$$\alpha = -4{,}577637 \cdot 10^{-5}$$

$$\beta = 0{,}0234375$$

$$\gamma = -2$$

Resolviendo la ecuación $\alpha c^2 + \beta c + \gamma = 0$, se obtienen las asíntotas $c_1 = 108{,}1984$, $c_2 = 403'8016$, $a = c_2 - c_1 = 295'6032$

Realizando el cambio de variable

$$Y_i = \ln\left(\frac{a}{y_i - c} - 1\right)$$

donde $c = c_1$ (asíntota inferior), se obtiene la nube de puntos

t_i	Y_i
0	-1,133401
5	-1,336692
6	-1,431716

Al ajustar una recta a dichos puntos resulta

$$Y = -0{,}0379253\, t + (-1{,}17821) = b\, t + A$$

$$\rho = -0{,}9730407$$

donde $h = B = -0'0379253$; $b = e^A = 0'3078292$

Es decir, la ecuación de la función logística que explica la evolución reciente de la población de Albacete, además de tener la utilidad instrumental de función de predicción, es

Para el cálculo de la bondad de ajuste

y_i	y^*_i	% de error
334,5	334,3	-0,06
342,3	343,7	0,40
346,8	345,6	0,35

$$I = 0{,}9555899$$

Las predicciones a medio plazo concuerdan con una extraordinaria aproximación con las obtenidas mediante el método de las componentes para la hipótesis que contempla una evolución en la mortalidad y fecundidad equivalente a la experimentada en 1985, lo que puede interpretarse en la terminología de Cowgill por la no modificación de las condiciones del "ciclo natural".[12]

AÑO	PREDICCIÓN LOGÍSTICA	PREDICCIÓN MÉTODO COMPONENTES (Hipótesis continuista)	DIFERENCIA %
1990	350,4	353,4	0,86
1995	358,3	359,9	0,45
2000	365,1	367,1	0,55
2005	371,0	372,8	0,48
2010	376,2	376,3	0,02

4. APLICACION DE LA CURVA LOGISTICA A LA POBLACION DEL MUNICIPIO CAPITAL

Una nueva aplicación es el ajuste de la población del municipio de Albacete a la función logística, para ello, y debido a que en 1950 la población de este municipio cambia de "ciclo cultural", dicho ajuste se va a realizar en dos tramos, 1900-1950 y 1950-1981, al mismo tiempo que va a permitir "la predicción" de la población total para diferentes años.

El valor de t para la obtención de las diferencias divididas que se ha utilizado en $t_{1900} = 0$.

(12) Díaz Martínez, F. "Demografía de la provincia de Albacete: Evolución histórica. Análisis y proyecciones. Aspectos socieconómicos". Tesis Doctoral.

AÑO	t_i	Δt_i	$f(t_i)$	$f_1(t_{i-1}, t_i)$	$f_2(t_{i-1}, t_i, t_{i+1})$
1887	-13		20,9		-0,006254
		13		0,04615	
1900	0		21,5		0,012341
		10		0,33000	
1910	10		24,8		0,019500
		10		0,72000	
1920	20		32,0		0,013500
		10		0,99000	
1930	30		41,9		0,062000
		10		2,23000	
1940	40		64,2		-0,073500
		10		0,76000	
1950	50		71,8		-0,025000
		10		0,26000	
1960	60		74,4		0,081000

Como se hizo en el ejemplo anterior, a partir de las diferencias divididas se obtienen los valores de $f'(t_i)$, lo que permite, una vez conocidos los valores (y_i, y'_i) ajustan la parábola

$$y'_i = \alpha y_i^2 + \beta y_i + \gamma \cdots\cdots\cdots\cdots\cdots\cdots$$

Obteniéndose

$$\alpha = -2{,}043246 \,.\, 10^{-3}$$

$$\beta = 0{,}2017265$$

$$\gamma = -3{,}274202$$

cumpliéndose que

$\beta^2 - 4\alpha\gamma > 0$ y que α y γ son ambos negativos, la ecuación $\alpha c^2 + \beta c + \gamma = 0$ admite dos raices reales positivas y distintas

$$c_1 = 20'47867,\ c_2 = 78'24977 \text{ y } a = c_2 - c_1 = 57'7711$$

La determinación de b y h se obtiene mediante el cambio de variable

$$Y_i = \ln\left(\frac{a}{y_i - c} - 1\right)$$

donde $c = c_1$ (asíntota inferior)

Al ajustar una recta a la nube de puntos (t_i, Y_i)

t_i	Y_i
0	4,017546
10	2,51518
20	1,389856
30	0,5288006
40	-1,13523
50	-2,074063

se obtiene

$$Y = -0{,}1207724\ t + 3{,}892991 = B\ t + A$$

$$\rho = -0{,}9968171$$

Luego $h = B = -0'1207724$ y $b = e^A = 49'057398$

Por lo que la ecuación de la logística será

$$y = 20{,}47867 + \frac{57{,}7711}{1 + 49{,}057398 \cdot e^{-0{,}1207724\,t}}$$

y_i	y^*_i	% de error
21,5	21,6	0,4
24,8	24,2	-2,4
32,0	31,2	-2,5
41,9	45,5	8,6
64,2	62,0	-3,4
71,8	72,2	0,6

La bondad de ajuste es I = 0'9982659

Análogamente, para el periodo 1950-1981, considerando $t_{1950} = 0$, resulta

La ecuación de la logística siguiente

$$y = 66{,}08992 + \frac{77{,}12478}{1 + 16{,}546308 \cdot e^{-0{,}1103858\,t}}$$

y_i	y^*_i	% de error
71,8	70,5	-1,8
74,4	77,9	4,7
93,2	93,4	-0,2
117,1	114,5	-2,2

Cuya bondad de ajuste es I = 0'9855259

De la función logística se deduce que la población límite prevista dentro de este ciclo cultural, para el municipio capital es de 143. 215 habitantes y las diferentes predicciones futuras son,

AÑO	POBLACIÓN PREVISTA
1990	130359
1995	135248
2000	138417
2005	140377
2010	141555