

VARIACIONES SOBRE UN TEMA DE LA FUNCION LOGISTICA

Eduardo E. Arriaga
(Oficina del Censo
de los Estados Unidos de América)

RESUMEN

El uso de la función logística en el análisis de población y en técnicas demográficas se ha incrementado significativamente durante los últimos años. Sin embargo, no se ha dado mucha atención a la interpretación de las constantes de la función ni a las consecuencias de los supuestos implícitos resultantes de la aplicación de la función en el campo demográfico. En este artículo, el autor presenta el significado de las constantes de la función logística cuando ésta se utiliza en algunos casos demográficos. También se dan algunos ejemplos de los supuestos implícitos resultantes de la aplicación de la función logística a las tendencias de las poblaciones urbanas y rurales, esperanzas de vida al nacimiento, y el procedimiento de la transformación del logito.

<ANALISIS ESTADISTICO> <AJUSTE DE CURVAS> <TENDENCIA>
<POBLACION RURAL> <POBLACION URBANA>
<ESPERANZA DE VIDA>

SUMMARY

VARIATIONS ON THE LOGISTIC CURVE THEME

The use of the logistic function in demographic techniques and analyses has increased significantly during recent years. Nevertheless, not too much attention has been given to the function of the constants nor to the consequences of the implicit assumptions when the function is used for estimating some demographic parameters. In this article, the author underlines the meaning of the constants in the logistic function when applied in some demographic cases, and gives some examples of the assumptions involved in the application of the logistic to the trends of urban/rural populations, life expectancies, and the logit transformation.

<STATISTICAL ANALYSIS> <CURVE FITTING> <TREND>
<RURAL POPULATION> <URBAN POPULATION> <LIFE
EXPECTANCY>

Las funciones logísticas han sido usadas, cada vez con mayor frecuencia, dentro del análisis demográfico. En una primera etapa, la logística fue usada para predecir la población total de un país o área (Pearl and Reed 1920; Pearl, Reed and Kish, 1940); posteriormente se ha usado para predecir o reproducir la tendencia de ciertas características de la población, como porcentaje urbano, porcentaje de alfabetos, o para interpolación (Arriaga, 1975; Naciones Unidas, 1982). Más recientemente, la función logística ha sido utilizada para suavizar funciones de las tablas de mortalidad, para estimar distribuciones de población por edades correspondiente a subáreas, para proyectar poblaciones de pequeñas áreas, etc. (Brass, 1975; CELADE, 1982; Naciones Unidas, 1982). Aunque el uso de las funciones logísticas en el análisis demográfico se ha expandido y aumentado notablemente, los supuestos implícitos que encierran dichas aplicaciones no deben descuidarse. El propósito de esta reseña técnica es explicitar algunos de tales supuestos y, además, clarificar el significado de las constantes de la función.

LA FUNCION LOGISTICA

La función logística, en una de sus expresiones generales, es:

$$y = \frac{1}{1 + e^{a+wn}} \quad (1)$$

en la cual, a y w son constantes, y n representa un período de tiempo. Como e^a es una constante, la fórmula (1) puede también escribirse:

$$y = \frac{1}{1 + Ke^{wn}} \quad (2)$$

Debido a la frecuente aplicación de funciones logísticas en el análisis de índices demográficos, debemos tener siempre presente cuál es el significado de las constantes a y w , y cómo deben ser interpretadas.

EL CAMBIO DE UNA PROPORCION Y LA LOGISTICA

Si se tiene un índice I que presenta una variación posible entre un límite inferior y superior L y U respectivamente (o que se supone varía entre dichos límites), se puede realizar el siguiente desarrollo. La proporción del índice I en el año j en relación al intervalo de variación $V = U - L$ será:

$$\frac{I_j - L}{V} \quad (3)$$

pero V , el intervalo de variación, puede ser expresado como la suma entre a), el valor máximo U , menos el valor del índice I y b), el valor del índice I menos el límite mínimo L . En símbolos, $V = U - I + I - L$. Por lo tanto, la proporción del índice para el año j puede expresarse como

$$\frac{I_j - L}{V} = \frac{I_j - L}{U - I_j + I_j - L} = \frac{1}{1 + \frac{U - I_j}{I_j - L}} \quad (4)$$

Del mismo modo, si tenemos otra observación del índice en el momento $j+n$, la misma relación podría ser hecha para ese año.

Ahora bien, el cambio del índice desde j a $j+n$ puede expresarse en función de una exponencial, esto es:

$$I_{j+n} - L = (I_j - L) e^{sn} \quad (5)$$

y, del mismo modo

$$U - I_{j+n} = (U - I_j) e^{un} \quad (6)$$

siempre habrán valores de s y u que satisfarán las dos expresiones anteriores.

Consecuentemente, la proporción del índice en el momento $j+n$, puede expresarse en función del índice en el momento j y las tasas de crecimiento de las fórmulas (5) y (6)

$$\frac{I_{j+n} - L}{V} = \frac{1}{1 + \frac{(U - I_j) e^{un}}{(I_j - L) e^{sn}}} = \frac{1}{1 + \frac{U - I_j}{I_j - L} e^{wn}} \quad (7)$$

donde $w = u - s$

Si ahora se hace el supuesto fundamental de que la diferencia de tasas w se mantiene constante durante todo el período de variación del índice, desde su valor mínimo al máximo, se tiene entonces (y solamente entonces) una logística. Para llegar a la fórmula general de la logística, solamente se necesita hacer

$$\frac{U - I_j}{I_j - L} = e^a \quad (8)$$

reemplazando (8) en (7), se llega a que la proporción del índice bajo análisis, que varía entre 0 y 1, sigue la tendencia de la siguiente logística:

$$\frac{I_{j+n} - L}{V} = \frac{1}{1 + e^{a+wn}} \quad (9)$$

fórmula que permite estimar la proporción del índice para cualquier fecha. Si se deseara obtener una estimación directa del índice, se despeja I , quedando la expresión:

$$I_{j+n} = L + \frac{V}{1 + \frac{U - I_j}{I_j - L} e^{wn}} \quad (10)$$

Esta última fórmula representa una logística modificada cuyos valores asintóticos son L y U , (recordemos que V es la diferencia entre U y L). Como se expresa en esta ecuación, es preferible trabajar directamente con el cociente que multiplica a la exponencial, fórmula (10), que con el valor de la constante a , fórmula (9).

Dado el requisito necesario para obtener una logística de que w sea constante, w puede ser fácilmente estimada, siempre que los valores de las asíntotas sean conocidos (0 y 1 en el caso de proporciones), o puedan ser supuestos. En cualquiera de los casos, la constante w se estima a partir de la fórmula

$$w = \frac{1}{n} \ln \left[\frac{U - I_{j+n}}{I_{j+n} - L} \bigg/ \frac{U - I_j}{I_j - L} \right] \quad (11)$$

que simplemente significa determinar la tasa anual de cambio desde el año j al $j+n$ del cociente entre lo que el índice aún puede cambiar ($U-I$) y lo que el índice ya ha cambiado ($I-L$).

EJEMPLOS DE CASOS ESPECIFICOS

a) Caso de la proporción urbana y rural

Si PU representa la proporción de la población en áreas urbanas, y si se usa en función logística para interpolar dicha proporción entre dos observaciones o para realizar una extrapolación en base a dichas observaciones, el cambio (logístico) de la proporción urbana puede escribirse del siguiente modo, por similitud a la fórmula (7).

$$PU_{j+n} = \frac{1}{1 + \frac{PR_j}{PU_j} e^{wn}} \quad (12)$$

donde PR es la proporción rural, o sea $1 - PU$. La constante w se estima en base a la fórmula (11) y dos observaciones en los años j y $j+n$

$$w = \frac{1}{n} \ln \left[\frac{PR_{j+n}}{PU_{j+n}} \bigg/ \frac{PR_j}{PU_j} \right] \quad (13)$$

Puede demostrarse que w es la diferencia de las tasas anuales de crecimiento de la población rural y urbana (Arriaga, 1975). Por lo tanto, al interpolar o extrapolar proporciones urbanas por medio de una logística, se supone implícitamente que la diferencia entre las tasas anuales de crecimiento de la población rural menos la tasa urbana es constante. Cuando la logística se usa con fines interpolativos, dicho supuesto es aceptable. Cuando la logística se usa con fines de extrapolación, debe ponerse en duda si el supuesto es aceptable. Indudablemente, la respuesta dependerá principalmente de la longitud del período de extrapolación.

Una función logística modificada, puede utilizarse para interpolar o extrapolar directamente la población urbana (UP) en vez de la proporción urbana. En este caso la fórmula se expresaría como:

$$UP_{j+n} = L + \frac{V}{1 + \frac{RP_j}{UP_j} e^{wn}} \quad (14)$$

donde $V = U - L$, RP es la población rural, y w se estima como se indica en la fórmula (13), pero reemplazando las proporciones urbanas y rurales por los valores absolutos de estas poblaciones (aunque el resultado es el mismo si w se estima usando proporciones). Esta fórmula (14) puede usarse para proyectar la población de áreas relativamente pequeñas de un país. En este caso, los problemas se presentan en la determinación de los valores asintóticos, principalmente el de la asíntota superior U , ya que L puede suponerse cero. Aparentemente dichos problemas no surgirían cuando se usan proporciones, ya que las asíntotas son 0 y 1. Sin embargo, en el caso de proporciones, implícitamente se supone lo siguiente: *a*) si la proporción de la población de ciertas áreas crece, se estaría haciendo tender la población de cada área hacia la población total del país; *b*) si la proporción de la población de ciertas áreas disminuye, la población de dichas áreas se estaría haciendo tender hacia cero. Ambos casos no son lógicos. Los problemas tampoco se resolverían si se proyectase la población absoluta de cada área, ya que en este caso habría que determinar las dos asíntotas. Sin embargo, la predicción de la población total de subáreas por medio de la función logística, para períodos de tiempo menores de 20 años, depende mucho más del cambio observado de dicha población durante las dos observaciones que se usan como base para determinar la función logística, que de los supuestos realizados sobre los niveles de las asíntotas. Este sería el caso general, excepto cuando las asíntotas (superior o inferior para población creciente o decreciente, respectivamente) se fijan muy próximas al último valor observado de la población.

Finalmente, debe decirse que cuando la logística se usa para predecir poblaciones de áreas pequeñas, los resultados son muy similares a los de cualquier otro procedimiento mecánico (exponencial, lineal, proyección de la razón, etc.) siempre que el período predictivo no sea mayor de 20 años desde la última observación disponible, y que los resultados se ajusten a una población esperada para el total de las áreas.

b) *Esperanza de vida al nacimiento*

La aplicación de la logística para interpolar esperanzas de vida al nacimiento es bastante generalizada y se utiliza mucho más que para extrapolar dicho índice. Por ejemplo, si se supone que los valores asintóticos de L y U son 20 y 85 años respectivamente, la fórmula (5) puede escribirse como:

$$e_o^{j+n} = 20 + \frac{65}{1 + \frac{85 - e_o^j}{e_o^j - 20} e^{wn}} \quad (15)$$

y el valor de w :

$$w = \frac{1}{i} \ln \left[\frac{85 - e_o^j}{e_o^j - 20} \quad \frac{85 - e_o^{j-i}}{e_o^{j-i} - 20} \right] \quad (16)$$

En el caso de las esperanzas de vida al nacimiento (o tasas globales de fecundidad u otros índices), el supuesto que se haga sobre el valor de las asíntotas puede afectar la tendencia del crecimiento del índice. Por ejemplo, si los valores asintóticos son 20 y 85, y se tienen dos observaciones de 30 y 35 años de esperanza de vida al nacimiento para los años j y $j+10$, la tendencia de la tasa anual de cambio de las esperanzas de vida al nacimiento no será monotónicamente decreciente durante todo el período de variabilidad de la función logística (como sería el caso si la asíntota inferior fuese cero). En el ejemplo, las tasas de crecimiento de las esperanzas de vida al nacimiento aumentan hasta el año j , son máximas alrededor del año $j+10$, y posteriormente comienzan a tender hacia cero.

c) *La logística y la transformación del logito*

Como se sabe, el logito es la linealización de la función logística. La relación lineal entre los logitos de dos funciones del mismo tipo, ha sido usada para suavizar funciones de tablas de mortalidad construidas con datos deficientes o para generar tablas de mortalidad cuando sólo existe información fragmentaria (Brass, 1975). El procedimiento supone una relación lineal entre los logitos de una función de una tabla de mortalidad observada o empírica, y los logitos de la

misma función pero correspondiente a otra tabla de mortalidad que se toma como estándar. Si se trata de la función de sobrevivencia a una edad exacta (l_x) en una tabla de mortalidad cuya raíz es 1, y se designa con asterisco la función proveniente de la tabla de mortalidad estándar, se tiene:

$$\ln \frac{1 - l_x}{l_x} = A + B \ln \frac{1 - l_x^*}{l_x^*} \quad (17)$$

Esta expresión puede transformarse en:

$$\frac{1 - l_x}{l_x} = e^A \left[\frac{1 - l_x^*}{l_x^*} \right]^B \quad (18)$$

luego,

$$l_x = \frac{1}{1 + \left[\frac{1 - l_x^*}{l_x^*} \right]^B e^A} \quad (19)$$

y de aquí se obtiene:

$$l_x = \frac{1}{1 + \left[\frac{{}_x d_o^*}{l_x^*} \right]^B e^A} \quad (20)$$

donde

$${}_x d_o^* = 1 - l_x^* \quad (21)$$

donde, como se sabe, A afecta principalmente el nivel de la esperanza de vida al nacimiento, y B afecta principalmente a la estructura de la función ℓ_x . Puede verse, en la ecuación (20), que cuando B es igual a la unidad estamos en presencia de una logística.

Esta forma de expresar la relación de la transformación del logito permite fácilmente observar lo que ocurre cuando se cambian los valores de A y B . Supongamos primero que B es distinto de 1 y que A es igual a cero. La función ℓ_x que se estimará a partir de la función ℓ_x^* estándar difiere de esta última en la siguiente forma: a) para cualquier valor de B , ambas tablas de mortalidad tendrán una ℓ_x idéntica para la edad, en que la función ℓ_x^* de la tabla estándar tiene un valor de $1/2$; b) si llamamos i a dicha edad se tiene lo siguiente:

$$1) \text{ Para } B < 1; \ell_x \leq \ell_x^* \text{ cuando } x \leq i$$

$$2) \text{ Para } B > 1; \ell_x \geq \ell_x^* \text{ cuando } x \leq i$$

Si ahora se supone que $B = 1$, la relación entre las dos ℓ_x es una relación logística:

$$\ell_x = \frac{1}{1 + \frac{{}_x d_0^*}{\ell_x^*} e^A} \quad (22)$$

en la que A es la tasa de crecimiento de la función logística que relaciona a las dos funciones, y donde $n=1$. El símbolo ${}_x d_0^*$ significa las defunciones de una tabla de mortalidad acumuladas desde la edad 0 hasta la edad x . La equivalencia de la fórmula (22) con la (7) está en que ${}_x d_0^* = U-I$, ya que en este caso la asíntota superior es 1 y el índice es la función ℓ_x . Como es conocido, en la transformación del logito, el significado de mantener $B = 1$ es no cambiar la estructura de la función ℓ_x . Veamos explícitamente el significado de esto.

Algunas veces, cuando se tiene solamente un valor de ℓ_x (generalmente la mortalidad infantil, o la sobrevivencia hasta las edades 2, 3, ó 5), y se quiere estimar una tabla de mortalidad que tenga la

“misma estructura de mortalidad” que la estándar, la estimación de todos los valores de l_x se puede obtener directamente de la tabla de mortalidad estándar sin utilizar logitos, sino simplemente una función logística simplificada. Por ejemplo, como $n=1$, el valor de e^A (de acuerdo a la fórmula 11) puede expresarse como:

$$e^A = \frac{i d_o}{l_i} : \frac{i d_o^*}{l_i^*} = Q \quad (23)$$

Donde l_i es la única observación disponible del nuevo nivel de mortalidad para la edad i , luego reemplazando la fórmula 23 en la

$$l_x = \frac{l_x^*}{l_x^* + Q_x d_o^*} \quad (24)$$

Cuando se utiliza el procedimiento de la transformación del logito (con dos parámetros), debe considerarse si el procedimiento es adecuado y cuál es el significado real de “la misma estructura de mortalidad”.

El problema está en determinar el significado de “igual estructura de la mortalidad” cuando se tienen distintos niveles de mortalidad. En otras palabras, si se acepta que “la misma estructura de mortalidad” es la relación derivada del sistema de logitos con dos parámetros cuando $B = 1$, (equivalente a la relación logística de las fórmulas 22 ó 24), la pregunta es si se debe aceptar una “misma” estructura de mortalidad para diferentes niveles de mortalidad. Los modelos de mortalidad conocidos (Naciones Unidas, 1956; Coale y Demeny, 1968; Organisation de Cooperation et de Development Economiques, 1980; Naciones Unidas, 1982) no siguen dicha relación, ni tampoco ésta es seguida en observaciones empíricas. Aparentemente, el procedimiento de construcción de tablas de mortalidad por medio de logitos (con dos parámetros) manteniendo el coeficiente angular $B = 1$, no seguiría las tendencias de las observaciones empíricas ni de los modelos de tablas de mortalidad comúnmente usados. Sin embargo, cuando los valores de la constante A son pequeños en valor absoluto (significando casi el mismo nivel de mortalidad que la tabla estándar), las posibles inconsistencias que puedan cometerse no serían de una magnitud importante.

Por último, debe mencionarse algo sobre la fórmula (22), que relaciona a dos l_x por medio de una logística. A veces se han hecho

predicciones de la función l_x por medio de una función logística, la cual se ajusta en base a dos observaciones para cada edad. La crítica a dicho procedimiento ha estado en que podría producir un juego proyectado de l_x con una estructura de mortalidad no apropiada, debido a que las tasas de crecimiento de las funciones logísticas podrían diferir mucho de acuerdo a las edades que se usen. Sin embargo, si en vez de ajustar una función logística a cada par de l_x , se determina una tasa de crecimiento de la función logística que sea constante para todas las edades, donde $nw = A$ (Fórmula 22 ó 24), la técnica sería exactamente igual a la que utiliza logitos cuando $B=1$ (fórmula 17).

CONCLUSIONES

No hay dudas que la aplicación de las funciones logísticas en el análisis demográfico ha sido aceptable y útil. La función logística no sólo es simple, sino que, además, su tendencia puede ser fácilmente comprendida.

El principal supuesto implícito en la función logística es que cuando se determina w (la tasa de crecimiento de la logística) en función de las asíntotas y del índice bajo estudio, w se mantiene constante. El requisito de mantener w constante es lo que más debe considerarse al usar una función logística, principalmente con propósitos extrapolativos o cuando se usa para generar una tabla de mortalidad en base a una tabla estándar y una observación de alguna función específica de la nueva tabla de mortalidad.

Por ejemplo, cuando la función logística se usa para extrapolar la proporción de la población perteneciente a un área, la pregunta que debe hacerse es si la diferencia de tasas anuales de crecimiento entre la población del área y la correspondiente al resto del país se mantendrá constante en el futuro. Como estamos casi seguros que a largo plazo esto no ocurrirá, la segunda pregunta es: ¿cuál deberá ser el período de tiempo de extrapolación para que el supuesto de mantener constante a w no difiera mucho de la realidad? Las mismas preguntas pueden hacerse cuando se trata de la extrapolación de cualquier otro índice demográfico.

Preguntas similares surgen al usar la función logística para generar una tabla de mortalidad cuando se tiene una observación y se supone que la estructura de mortalidad será "la misma" que la correspondiente a una tabla de mortalidad estándar ($B=1$). Si la observación disponible no difiere mucho de la estándar correspondiente, la

nueva tabla de mortalidad podría aceptarse (el valor absoluto de A sería cercano a cero). Sin embargo, si la nueva tabla de mortalidad generada tiene un nivel de mortalidad muy distinto al de la tabla estándar (cuando A difiere considerablemente de cero), el procedimiento podría ser juzgado inapropiado, y debería utilizarse otra técnica.

REFERENCIAS BIBLIOGRAFICAS

- Arriaga, Eduardo; 1975 "Selected Measures of Urbanization", en *The Measurement of Urbanization and Projection of Urban Population*, editado por Goldstein y Sly, Union Internacional Para el Estudio Científico de la Población, Ediciones Ordina, Bélgica.
- Brass, William; 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*, POPLAB, The University of North Carolina at Chapel Hill.
- CELADE; 1982. *Metodología de las Proyecciones de Población Urbana-Rural y Población Económicamente Activa Elaboradas en CELADE*, José M. Pujol y Juan Chackiel, Presentado en el Seminario de Proyecciones de Población, San José, Costa Rica, Octubre 4-13, 1982.
- Pearl, Raymond y Reed, Lowell; 1920. "On the Rate of Growth of the Population of the United States Since 1790 and Its Mathematical Representation" *Proceedings of the National Academy of Sciences*, Vol. 6, pp. 275-288.
- Pearl, R; Reed, L; y Kish, J; 1940. "The Logistic Curve and the Census Count of 1940" *Science*, Vol. 92, pp. 486-488.
- United Nations; 1982. *Estimates and Projections of Urban, Rural and City Populations 1950-2025: The 1980 Assessment*, ST/ESA/SER.R/45, Nueva York.