Demographic Relationships at the Macro Level

Krishnan Namboodiri
Professor Emeritus
Ohio State University
190 N Oval Mall, 300 Bricker Hall
Columbus, Ohio, 43210
namboodivi.2@osu.edu

Abstract

The focus herein is on macro-level relationships in mathematical demography and regression analysis. Notions of vector space calculus are used in discussing the impact of nuptiality and fertility on the net reproduction rate and log-ratio techniques in the analysis of regression relationships involving compositions. The latter is illustrated using occupational data from the U. S. population census. The methods described herein permit using all the available information while the commonly used method, which replaces a vector by a summary measure, lets much of the flavor in the data slip away.

Key Words: Vector space calculus, net reproduction rate, nuptiality, fertility

Krishnan Namboodiri

Résumé

Cet article a pour sujet principal les relations de niveau macro dans la démographie mathématique et l'analyse de régression. Des principes de calcul d'espace vectoriel ont été appliqués pour examiner l'impact des nuptialités et fécondité sur le taux de reproduction nette; nous avons aussi employé des techniques de "log-ratio" dans l'analyse de régression des relations qui comportent des compositions. Ces dernières sont illustrées à l'aide de données sur les professions provenant du recensement démographique des Etats-Unis. Les méthodes décrites dans cet article permettent d'utiliser la somme des informations tandis que la méthode la plus souvent employée, qui remplace le vecteur par une statistique sommaire, laisse passer beaucoup de détails.

Mots-clés: Calcul d'espace vectoriel, taux de reproduction nette, nuptialité, fécondité

Introduction

Demographers are often interested in relationships among demographic variables at the macro level. An example is the relationship between the maternity function and the intrinsic rate of natural increase. Such problems involving the implications of a change in a mathematical function defined over a subset of the real number field can be approached from the perspective of what is known as functional analysis.

The net reproduction rate is examined as it relates to nuptiality and fertility. Then the analogous empirical setting in which a vector is a predictor or the 'response' in regression analysis is considered. An illustration in which the 'response' is the income composition of the labor force and the predictors are the corresponding educational and age compositions is provided. For this illustration U.S. data on occupational characteristics is used (U. S. Bureau of the Census 1963).

Preliminaries

A linear vector space is a set of mathematical entities which can be added together and multiplied by real (or complex) numbers, giving in each case an element of the set. These operations in linear vector space obey laws similar to the distributive, commutative, and associative laws that apply to the arithmetic of ordinary numbers. Examples of linear vector spaces include the set of *n*-tuples of real numbers and the set of all continuous real-valued functions on the interval [*a*, *b*].

Associated with a linear vector space is a scalar field, which may be the field of real numbers or of complex numbers. Any operator or rule that produces an element of a scalar field, given an element of a linear vector space, is called a functional. A simple example of a functional in demography is the integral operator in

$$\int_a^b m(x)\, dx \tag{1}$$

where $m(x)dx$ is the probability that a woman of age x will give birth to a daughter before reaching age $x + dx$, mortality being ignored. Note that this operator produces a real number, given the maternity function.

Generalizing the familiar concept of 'length,' or 'distance,' one defines a norm, a non-negative function, over a linear vector space. A norm must have the following properties: If V is a linear vector space and u and v are any two elements in it, then the norm of u is zero if and only if u is zero; the norm of the sum of u and v is $\leq$ the sum of the norms of u and v; and the norm of αu is α times the norm of u, where α is a real positive number. The norm of u is often represented by the symbol $||u||$. . An example of a norm is the Euclidean length function. If X is a linear vector space of n-tuples, $(x_1, x_2, \ldots, x_n)$, then $||x|| = \sqrt{(x_1^2 + \ldots + x_n^2)}$ is a possible norm over X.

The derivative or variation of a functional can be defined in a number of ways, of which I mention just one. For others see, for example, Troutman (1996). A continuous operator or function $L\colon N \to M$ is said to be the Freschet derivative of $f\colon N \to M$ at the point $x \in N$ if

$$f(x + h) - f(x) = Lh + o(h) \text{ as } h \to 0 \tag{2}$$

Here L for $f'(x)$, if $f(x)$ is differentiable.   For a real-to-real function $f : \mathrm{R} \rightarrow \mathrm{R}$, the ordinary derivative at x , as defined in elementary calculus, is a number giving the slope of the graph of f at x . The Freschet derivative L at x , however, is not a number; it is a linear operator $\mathrm{R} \rightarrow \mathrm{R}$. If f is a differentiable function, then $f(x + h) - f(x) = h f'(x) + o(h)$. Comparing this with (2), one may view L as the operator which multiplies each $h \in R$ by the number $f'(x)$ (interpreted as in elementary calculus), or it may be the result of applying the linear operator $f'(x)$ (interpreted as in the definition given above)   to the element h of the space R.

To give an example (see Griffel, 1981, pp. 311-312), consider a functional $J: C$ $[0, 1] \rightarrow \mathrm{R}$, defined as

$$J(u) = \int_0^1 g(x)[au(x) + b\{u(x)\}^2]\mathrm{d}x, \qquad (3)$$

where a and b are constants and $g(x)$ a given function.

$$
\begin{aligned}
J(u + h) - J(u) &= \int_0^1 g(x)[au(x) + ah(x) + b\{u(x) + h(x)\}^2 \qquad (4)\\
&\quad -au(x) - b\{u(x)\}^2]\mathrm{d}x\\
&= \int_0^1 g(x)[ah(x) + 2bh(x)u(x) + b\{h(x)\}^2]\mathrm{d}x\\
&= \int_0^1 g(x)[a + 2bu(x)]h(x)\mathrm{d}x + \int_0^1 g(x)b\{h(x)\}^2]\mathrm{d}x
\end{aligned}
$$

The last term here is $o(h)$; hence $J'(u)$ is the linear functional defined by

$$J'(u)h = \int_0^1 g(x)[a + 2bu(x)]h(x)\mathrm{d}x. \qquad (5)$$

 $J'(u)h$ denotes the number into which the functional $J'(u)$ maps the function h .

Net Reproduction Rate as it Relates
to Nuptiality and Fertility

Let $m(a)$ be the fertility function by age, $p(a)$ the corresponding survival probability function, and $n(t)$ the force of nuptiality so that $\exp(-\int_0^a n(t)\,dt)$ is the probability of remaining single until age a , all of these referring to females. Let $m_w(a)$ and $m_o(a)$ be the fertility functions within and outside marriage, respectively. Then assuming no mortality differential according to marital status, the net reproduction rate is given by

$$J(n) = R_0 = \int_0^\omega p(a)[1 - \exp(-\int_0^a n(t)\, dt)]m_w(a)\, da \qquad (6)$$
$$+ \int_0^\omega p(a)\exp(-\int_0^a n(t)\, dt)m_o(a)\, da$$

Considering $p(a)$ as a given function,

$$J(n + h) - J(n) = \int_0^\omega p(a)[1 - \exp(-\int_0^a \{n(t) + h(t)\}dt)]m_w(a)\, da$$
$$+ \int_0^\omega p(a)\exp[-\int_0^a \{n(t) + h(t)\, dt]m_o(a)\, da$$
$$- \int_0^\omega p(a)[1 - \exp(-\int_0^a n(t)\, dt)]m_w(a)\, da$$
$$- \int_0^\omega p(a)\exp(-\int_0^a n(t)\, dt)m_o(a)\, da$$
$$= \int_0^\omega p(a)\exp(-\int_0^a \{n(t)dt) \int_0^a h(t)dt\, [m_w(a) - m_o(a)]\, da + o(h)$$

Hence $J'(n)$ is the linear functional defined by

$$J'(n)h = \int_0^\omega p(a)\exp(-\int_0^a \{n(t)dt) \int_0^a h(t)dt\, [m_w(a) - m_o(a)]\, da$$

that states that if, at each age, fertility within marriage is higher than that outside marriage, then the sign of $J'(n)h$ depends upon the sign of $\int_0^a h(t)\, dt$. Obviously, $\int_0^a h(t)\, dt \geq 0$, if $h(t) \geq 0$ throughout $0 \leq t \leq a$, i.e., for all ages. Thus, as a particular case of the setup described above, if the probability of remaining single till age a is reduced, for all a, then the net reproduction rate will increase. The same applies if the probability of remaining single is reduced in one age interval, with no change in this respect at other ages. (See Ruzicka (1974) for another way of arriving at this particular result.) Note that the result obtained above is quite general, although in interpreting it emphasis was given to simple scenarios such as $h(t) \geq 0$ for all t. Also note that in deriving the results, no restriction was put on the form of the functions involved. When focusing on $J'(n)h$, however, one will be concerned with functions such as $n(t)$ restricted to the forms observed in human populations.

It is worth emphasizing that just as in elementary calculus we think of $f'(x_0)$ as the derivative of a function $f(x)$ evaluated at the point $x = x_0$ in vector space calculus of the kind illustrated above, we think of the behavior of an element of the space at a particular point in the space, but the 'points' are not real numbers; they are real-valued functions, n-tuples, or the like.

In the context of regression analysis an issue is how a composition variable (e.g., an age pattern of something) affects a response variable. But such issues are discussed in connection with the usual regression analysis after representing

the composition variables by some summary measures. There are, however, ways to treat the composition variables more fully as discussed in the next section.

Composition Variables in Regression Analysis

In the context of regression analysis, when a composition variable is considered as a predictor or response, the common strategy is to represent the composition variable by a summary measure such as the mean or median. Thus Parcel and Muller (1989) regressed average earning on mean year of schooling, mean year of work experience, and so on, using occupational groups as the units of analysis. Such approaches sacrifice a great deal of available information and more importantly fail to shed light on the details of how a predictor composition affects the response.

A k-part composition (w_1, w_2, ..., w_k) may be a relative frequency distribution (e.g., income composition, which could be treated as a multinomial) or an ingredient-makeup (e.g., a budget composition, which cannot be treated as a multinomial). One can handle both types using the so-called log-ratios (see Aitchison 1986): ln (w_i /w_k), $i = 1$, ..., (k -1). The literature on mixture-experiments (see, e.g., Cornell 1981) may be worth mentioning in this connection since it shows how to treat an ingredient-makeup as a predictor.

After estimating equations predicting y_i defined as ln (w_i /w_k), $i = 1$, ..., (k -1), one obtains the 'predicted' distribution of (w_1, w_2, ..., w_k), corresponding to any given configuration of the predictors, noting that

$$w_i = w_k \exp(y_i), i = 1,2,...,(k-1) \text{ and } w_1 + w_2 + ... + w_k = 1, \text{ giving}$$

$$w_i = [1 + \sum_1^{k-1} \exp(y_i)]^{-1}$$

Using the predicted response compositions for specified configurations of the predictors, it is possible to examine the impact of the predictors on the response. To illustrate, consider a simple set up in which there are two predictors, U and V, both scalars (not distributions). Suppose we obtain the predicted distributions corresponding to say (minimum of U, median of V) and (maximum of U, median of V). Then comparing the two predicted distributions, one can assess how the response distribution shifts when U changes from its minimum to its maximum, while V remains 'fixed' at its median value. This strategy can be expanded to involve compositions as predictors. See Namboodiri, Corwin, and Dorsten (1993) for an application example.

An Illustration

For an illustration consider the relationship of the educational composition and age distribution to the income composition of the occupation-specific U.S. male labor force (U. S. Bureau of the Census 1963). The available data are all categorical (see Table 1). The results presented in Table 2 are from an analysis using the following log-ratios for 237 occupational classes (excluding some classes with small n's):

$y_i = $ ln (number of individuals in the ith income category ÷ number of individuals in the highest, i.e., 11th, income category), $i = 1, 2, ..., 10$.

$E_i = $ ln (number of individuals in the ith education category ÷ number of individuals in the highest, i.e., 8th, education category), $i = 1, 2, ..., 7$.

$A_i = $ ln (number of individuals in the ith age category ÷ number of individuals in the highest, i.e., 15th, age category), $i = 1, 2, ..., 14$.

Notice that the relationship of the E_i 's and the A_i 's to any single y_i is analogous to a mapping from a vector space to a real number field. For interpreting the results one estimates the predicted composition vector from the predicted y_i's, corresponding to different configurations of the predictor vectors. See Namboodiri, et al. (1993) for a demonstration of how this is done.

Concluding Remarks

In this paper the emphasis has been that many demographic phenomena can be viewed as problems of mapping from a vector space to a real number field. In the regression context, this type of a setup arises when one is interested in composition variables as predictors or response. When examining such relationships, it is a common strategy to represent a vector by means of a summary measure. This strategy is not always advisable. In many cases, the summary measure used may not capture the full flavor of the vector (e.g., a probability distribution). By using the approach outlined herein it is possible to examine the relationships of interest in a more general fashion.

Table 1
Specification of Variables

Variable	Categories and Codes
Education **(in years of schooling completed)**	(1) = <5 (2) = Grades 5 - 7 (3) = Grade 8 (4) = High School 1 - 3 yrs. (5) = High School 4 yrs. (6) = College 1-3 yrs. (7) = College 4 yrs. (8) = College 5+ yrs.
Age **(in years)**	(1) = 14 - 15 (2) = 16 - 17 (3) = 18 - 19 (4) = 20 - 24 (5) = 25 - 29 (6) = 30 - 34 (7) = 35 - 39 (8) = 40 - 44 (9) = 45 - 49 (10) = 50 - 54 (11) = 55 - 59 (12) = 60 - 64 (13) = 65 - 69 (14) = 70 - 74 (15) = 75 +
Income **(in US Dollars)**	(1) = Under 1,000 (2) = 1,000-1,999 (3) = 2,000-2,999 (4) = 3,000-3,999 (5) = 4,000-4,999 (6) = 5,000-5,999 (7) = 6,000-6,999 (8) = 7,000-7,999 (9) = 8,000-8,999 10) = 9,000 - 14,999 (11) = 15,000 +

Note: For each variable, the log ratio in the regression analysis = In (the frequency in the ith category/the frequency in the last category).

Table 2. Estimated Equations for Income Represented by Log-ratios {y_j} using Education and Age as Predictors Represented by Log-ratios, {E_i} for Education and {A_i} for Age

Predictor	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
Intercept:	4.47	4.39	3.96	3.39	2.4	1.8	1.41	0.84	0.6	0.56
E_1						-0.31	-0.53	-0.45	-0.13	-0.07
E_2		0.63	0.89	0.38		0.64	0.84	0.66		
E_3	0.34		-0.55							
E_4	0.91									
E_5	-1	-0.4			0.44				0.45	0.16
E_7	-0.41	-0.34	-0.46	-0.52	-0.56	-0.35	-0.26		-0.27	
A_1	0.13									
A_2					-0.18	-0.26	-0.22	-0.19	-0.1	-0.1
A_4	1.21	1.31	1.11	0.97	1.04	1.05	0.78	0.63	0.32	0.35
A_7	-1.91	-1.88	-1.75	-1.57	-1.55	-1.24	-1.7	-1.43	-1.26	-0.42
A_8							1.68	1.49	1.54	
A_9										
A_{10}	0.93	0.93	1.12	1.24	1.4	1.19				0.47
A_{12}	-0.47	-0.4	-0.46	-0.54	-0.55	-0.57	-0.35	-0.36	-0.28	-0.29
R^2	0.77	0.73	0.73	0.69	0.64	0.62	0.6	0.54	0.47	0.22

Note: Only significant coefficients (regression slopes) shown in the table. All coefficients shown are significant at the 5% level of probability.

References:

Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. London: Chapman & Hall.

Cornel, J. A. 1981. *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*. NY: Wiley.

Griffel, D. H. 1981. *Applied Functional Analysis*. Chichester, West Essex, England: Ellis Horwood.

Namboodiri, Krishnan, Ronald G. Corwin, and Linda Eberst Dorsten. 1993. "Analyzing Distributions in School Effects Research: An Empirical Illustration." *Sociology of Education* 66: 278-294.

Parcel, Toby L. and Charles W. Muller. 1989. "Temporal Changes in Occupational Earnings Attainment." *American Sociological Review* 54: 622-634.

Ruzicka, Ladislav T. "Nuptiality and Fertility of Birth Cohorts." *Demography* 11: 397 – 406.

U. S. Bureau of the Census. 1963. *U. S. Census of the Population: Occupational Characteristics*. Washington, D. C.: U. S. Government Printing Office.

Troutman, John L.1996. *Variational Calculus and Optimal Control*. Second edition. NY: Springer.