A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 255, 2011

Aleksander Susel[*]

MULTIPLE CLASSIFICATION ANALYSIS. THEORY AND APPLICATION TO DEMOGRAPHY

Abstract. Multiple classification analysis (MCA) is an additive model having a wider spectrum of application than, e.g. linear regression models. The variables in MCA model could represent any scale, like interval or nominal scale. Moreover, an influence of independent variable on dependent one could be estimated in the light of control variables. Finally, a linearity assumption between variables is relaxed. The presented paper is devoted to the application of MCA model to measure an influence of income on fertility of women in the United States. The theoretical characteristics of MCA model is described as well.

Key words: income, fertility, the USA, MCA.

I. INTRODUCTION

One of the areas of investigation conducted by population scientists is an influence of the level of income on fertility. Cain and Weininger (1973) analyzed the relationship between income and fertility in the light of theoretical approach proposed by Becker (1960, 1965). They proved that income is positively and significantly correlated with the level of fertility among American women. The influence was weak and significant in case of women aged 45–49, but strong and significant in case of group of teenagers. Thornton's (1979) paper is based on the theoretical framework of Easterlin (1969, 1978). The author used three income approximation like husband's income, household's net assets and expected income. The results show that the husband's income determined positively current fertility and total fertility but negatively planning fertility. Households with a medium level of assets were characterized by the highest level of fertility. Finally, they found that consumption is highly competitive to children as well as quality of children is competitive to quantity of children. Freedman and Thornton (1982) used a longitudinal data to face planning fertility in 1962 with realized fertility in 1977. The analyses did not confirm strong and

[*] Department of Quantitative Methods in Management, Nowy Sacz School of Business – National-Louis University.

persistent dependence between husband's income and fertility. These outcomes were found for both a whole sample and for each order of birth. The authors stated that the households with higher expected income would not have more children in future. They argued that the economic conditions have changed between 1962 and 1977. Probably in that time, increased a quality of children instead of quantity of them but authors did not present any arguments for it. Rank (1989) analyzed the determinants of fertility of women on welfare. The author found that a probability being on welfare increases if woman was unmarried, with one child or with lower education. Besides, the results show that 11,5 percent of women on welfare gave a birth within 3 years. Finally, motherhood is determined by education (negatively), marital status (if married, higher probability of motherhood) and time being on welfare (negatively) significantly.

We analyzed the relationship between income and fertility using multiple classification analysis. A measure of income was income per adult equivalent (IAE) (Fuchs, 1986), whereas measure of fertility was number of live births between 1997 and 2001. The analyses were performed from a point of view age, marital status, education and number of children ever born (Brewster *et al.*, 1998; Mosher *et al.*, 1986). We analyzed variability in the level of fertility among households across the above listed variables. Marital status was consider as a length of first marriage (in years). In case of never married women or if a length of marriage was less than a year zero was attributed (Thornton, 1979). The source of dataset was the National Survey of Family Growth (NSFG), Cycle 6, 2002.

II. DESCRIPTION OF THE MODEL

Multiple classification model is given by the following equation (Nagpaul, 2001):

$$Y_{ij...n} = \bar{y} + a_i + b_j + ... + e_{ij...n}, \quad (1)$$

where: $Y_{ij...n}$ – value of function basing on a number of individuals n belonging to ith category of independent variable A, jth category of independent variable B, etc., $\bar{y}$ – mean value of dependent variable, a_i – effect of ith category of independent variable A on dependent variable, b_j – effect of jth category of independent variable B on dependent variable, $e_{ij...n}$ – random error.

Minimization of sum of squared error is used to estimate the parameters of model. The estimation is also possible by solving a set of normal equations (e.g. for three independent variables) (Nagpaul, 2001):

$$
\begin{aligned}
a_i &= A_i - Y - (1/W_i)SW_{ij}b_j - (1/W_i)SW_{ik}c_k\,, \\
b_j &= B_j - Y - (1/W_j)SW_{ij}b_j - (1/W_j)SW_{ik}c_k\,, \\
c_k &= -Y - (1/W_k)SW_{ij}b_j - (1/W_k)SW_{ik}c_k\,,
\end{aligned} \tag{2}
$$

where: A_i – mean value of dependent variable belonging to ith category of independent variable A, B_j – mean value of dependent variable belonging to jth category of independent variable B, C_k – mean value of depedent variable belonging to kth of independent variable C.

The significance of variance of dependent variable explained by the independent variables is tested by the statistic as follows (Nagpaul, 2001):

$$
F = \frac{E/(C-P)}{Z/(N-C+P-1)} \tag{3}
$$

with $(C - P)$ and $(N - C + P - 1)$ degrees of freedom; N – sample size, C – total number of categories of predictors, P – number of predictors, and E is given by the equation (Nagpaul, 2001):

$$
E = \sum_i \sum_j \ldots \sum_n W_{ij\ldots n}(a_i + b_j \ldots)^2\,. \tag{4}
$$

There are four characteristics using to measure an influence of independent variables on dependent one. *Eta statistic* (η_I) measures an influence of independent variable Ith on dependent one. This characteristic is also called a ‘correlation ratio’. It measures a relationship between the variables before adjustment. The *eta statistic* is given by the formula (Andrews *et al.,* 1973; Bachman, 1970; Nagpaul, 2001):

$$
\eta_I = \sqrt{\frac{U_I}{T}}\,, \tag{5}
$$

where:

$$U_I = \sum_j (\sum_k w_{Ijk})(\overline{Y}_{Ij} - \overline{Y})^2 , \tag{6}$$

$$T = \sum_k w_k (Y_k)^2 - \frac{(\sum_k w_k Y_k)^2}{\sum_k w_k} , \tag{7}$$

where: w_k – weight of kth category of independent variable, Y_k – effect of kth category of independent variable on dependent variable, $\overline{Y}_{Ij}$ – mean value of dependent variable of jth category of independent variable I.

Eta square statistic describes a portion of variance explained by the independent variable I before adjustment. This measure is given by the equation (Andrews *et al.,* 1973; Bachman, 1970; Nagpaul, 2001):

$$eta^2_{adj.} = 1 - \left[\frac{(T-E)/(N-C+P-1)}{T/(N-1)} \right] . \tag{8}$$

Beta statistic (β_I) measures the influence of independent variable I on dependent one. This characteristic is also called a ‘correlation ratio’. It measures the relationship between the variables after adjustment. It is described by the formula (Andrews *et al.,* 1973; Bachman, 1970; Nagpaul, 2001):

$$\beta_I = \sqrt{\frac{D_I}{T}} , \tag{9}$$

where:

$$D_I = \sum_j (\sum_k w_{Ijk})(a_{ij})^2 , \tag{10}$$

where: a_{ij} – adjusted deviance of jth category of independent variable I on final iteration.

The last statistic is *beta square statistic*. It measures a portion of variance explained by the independent variable I after adjustment. This measure is expressed by the formula (Andrews *et al.,* 1973; Bachman, 1970; Nagpaul, 2001):

$$beta_{adj.}^2 = 1-(1-beta^2)A \tag{11}$$

where:

$$A = \frac{N-1}{N+P-C-1}. \tag{12}$$

The detailed description of the MCA model is presented in papers of F. Adrews *et al.* (1973), N. H. Nie *et al.* (1976), J. G. Bachman (1970), R. D. Retherford, M. J. Choe (1993) and applications in A. Thornton (1979), C. Goldschneider and W. D. Mosher (1991), W. D. Mosher *et al.* (1992), J. van Ginneken, A. Razzaque (2003).

III. THE MCA ANALYSES

The analyses were performed for Hispanic origin and African-American households. For these two subgroups of households we found that fertility was determined by income significantly (not shown here). We included the variables as follows: age, education, length of first marriage and number of children ever born (CEB). We verified differences in fertility with respect to income regarding the above variables. Income was expressed by income per adult equivalent (IAE) whereas fertility by number of planning children. The MCA estimations were shown in table 1.

Table 1. The MCA for Hispanic origin and African-American households

IAE (in USD)	Number of households	Avarage number of planning children (observed)	Average number of planning children (adjusted)						
			A	M	EDU	B	A,B	M, B	A, M, B
Hispanic origin									
below 9722	621	0,75	0,86	0,83	0,70	1,07	1,03	1,04	1,02
9723–19643	476	0,90**	0,94	0,96	0,92	1,01	1,00	1,02	1,00
19644–32500	275	1,18**	1,09	1,11	1,22	0,87	0,92	0,90	0,92
above 32501	217	1,68**	1,39	1,41	1,71	0,91	0,98	0,94	0,99
African-American households									
below 9722	518	0,52	0,58	0,54	0,50	0,80	0,76	0,78	0,76
9723–19643	431	0,66	0,71	0,68	0,68	0,79	0,79	0,79	0,79
19644–32500	306	0,87**	0,90	0,89	0,89	0,73	0,78	0,74	0,78
above 32501	275	1,45**	1,22	1,34	1,43	0,88	0,90	0,89	0,90

Source: own calculations based on NSFG. A-age, M-length of first marriage, EDU-education, B-CEB. ** $p< 0,05$.

The analyses show that when CEB was controlled the difference between observed averages for Hispanic origin households dropped from 0,93 (1,68 and 0,75) to 0,16 (1,07 and 0,91) child. In case of African-American households, the differences were, respectively 0,93 (1,45 and 0,52) and 0,08 (0,88 and 0,80) (table 1). It indicates that CEB affecting planning fertility among households with different income, mostly. Two additional variables like age and length of first marriage explained only to some degree the differences in observed averages. When averages were adjusted by age the difference was reduced to 0,53 for African-American households, and 0,64 child for Hispanic origin households. When the length of first marriage was controlled, then the differences were, respectively 0,58 and 0,80 child. The results were improved when both age and CEB were controlled. The differences were 0,05 (Hispanic origin) and 0,14 (African-American) child. Finally, when all variables were controlled, the differences were, respectively 0,03 and 0,14 child. The level of education influenced the relationship between fertility and income very slightly. Details are shown in table 1.

We would like to underline that control variables explained the differences in fertility due to changes in the level of income. This is an evidence that a source of differences in planning fertility is age, marital status and CEB. It proves that in research on influence of income on fertility additional factors should be included. They could improve quality of research and give a full perspective on the analyzed relationship.

IV. CONCLUSIONS

The presented study shows that the influence of income on fertility was not consistent. This relation depends on how fertility and income are measured. We found that current fertility is proportional inversely but planning fertility is proportional directly to the family income. Moreover, race and Hispanic origin as well as control variables shoud be included in the analysis between income and fertility. We also found that planning fertility depends on the distribution of age, marital status and current fertility. The influence of income on fertility was slightly when those three variables were controlled.

REFERENCES

Andrews F., Morgan J. N., Sonquist J. A., Klem L. (1973), *Multiple Classification Analysis. A report on a computer program for multiple regression using categorical predictors*, Second Edition, The University of Michigan.

Becker G. S. (1960), An economic analysis of fertility, in: Becker G. S., ed., *Demographic and Economic Change in Developed Countries*, Universities-National Bureau Conference Series 11. Princeton University Press.

Becker G. S. (1965), A theory of the allocation of time, *The Economic Journal* 75(299), 493–517.
Brewster K. L., Cooksey E. C., Guilkey D. K., Rindfuss R. R. (1998), The changing impact of religion on the sexual and contraceptive bevavior of adolescent women in the United States, *Journal of Marriage and the Family* 60(2), pp. 493–504.
Cain G. G., Weininger A. (1973), Economic determinants of fertility: results from cross-sectional aggregate data, *Demography* 10(2), 205–223.
Easterlin R. A. (1969), Toward a socioeconomic theory of fertlity: A survey of recent research on economic factors in American fertility, in: Behram J. E. et al., eds., *Fertility and family planning: A World View*, University of Michigan Press.
Easterlin R. A. (1978), The economics and sociology of fertility: a synthesis, in: Tilly C., ed. *Historical Studies of Changing Fertility*, Princeton University Press.
Freedman D. S., Thorton A. (1982), Income and fertility: the elusive relationship, *Demography* 19(1), 65–78.
Fuchs V. R. (1986), Sex differences in economic well-being, *Science*, 23(4749), 459–464.
Goldschneider C., Mosher W. D. (1991), Patterns of contraceptive use in the United States: the importance of religious factors, *Studies in Family Planning* 22(2), 102–115.
Mosher W. D., Johnson D., Horn M. (1986), Religion and fertility in the United States: the importance of marriage pattern and hispanic origin, *Demography* 23(3), 367–379.
Mosher W. D., Williams L. B., Johnson D. P. (1992), Religion and fertility in the United States: new patterns, *Demography* 29(2), 199–214.
Rank, M. R. (1989), Fertility among women on welfare: incidence and determinants. *American Sociology Review* 54, 296–303.
Retherford R. D., Choe M. J. (1993), *Statistical models for causal analysis*, Wiley and Sons, New York.
Thornton A. (1979), Religion and fertility: the case of mormonism, *Journal of Marriage and the Family* 41(1), 131–142.
van Ginneken J., Razzaque A. (2003), Supply and Demand Factors in the Fertility Decline in Matlab. Bangladesh in 1977–1999, *European Journal of Population* 19(1), 29–45.

Aleksander Suseł

**MULTIPLE CLASSIFICATION ANALYSIS.
THEORY AND APPLICATION TO DEMOGRAPHY**

Model analizy klasyfikacji wielokrotnej (MCA) jest addytywnym modelem mającym szersze możliwości zastosowania niż, np. modele regresji liniowej. Przede wszystkim ze względu na to, gdyż zmienne w modelu MCA mogą pochodzić ze skal np. przedziałowej czy nominalnej. Poza tym, możliwe jest określenie stopnia wpływu zmiennych niezależnych zarówno przed jak i po uwzględnieniu zmiennych kontrolnych. Wreszcie, nie jest wymagane spełnienie założenia liniowej zależności pomiędzy rozważanymi zmiennymi. Artykuł przedstawia zastosowanie modelu MCA w analizie wpływu poziomu dochodu na planowaną płodność Amerykanek.