DECISION SCIENCES

IN THE

PUBLIC AND PRIVATE SECTORS:

THEORY AND APPLICATIONS

PROCEEDINGS OF THE
TWELFTH ANNUAL CONFERENCE
OF THE
SOUTHWEST REGION
AMERICAN INSTITUTION FOR DECISION SCIENCES

EDITED BY
BETTY YANTIS
JOHN E. NIXON
COLLEGE OF BUSINESS AND ECONOMICS
UNIVERSITY OF NEVADA, LAS VEGAS
LAS VEGAS, NEVADA

NEW ORLEANS, LOUISIANA MARCH 5-8, 1981

MCA AND LOGIT: A COMPARISON*

Patricia M. Shields,
Southwest Texas State University

ABSTRACT

In this paper a technique is developed which shows how both logit and MCA can be adapted for comparison purposes. The technique is developed for the special case of dichotomous dependent variables. The method was empirically tested and the data reveal that the logit probabilities, with few exceptions, were consistent with those generated by the MCA technique. The author advocates the use of this technique to confirm the MCA coefficients. Furthermore, when MCA and logit yield consistent results MCA coefficients should be used for presentation and discussion purposes.

Introduction

Complex, rigorous, multivariate statistical techniques often yeild results which are difficult for individuals not trained in these techniques to interpret. Policy scholarship necessarily seeks an audience outside the academic community. There is a need to use multivariate methods which are statistically sound, yet easy to understand and interpret. Multiple Classification Analysis (MCA), a form of multiple regression analysis is such a technique [1]. The ability to read a simple table of means and understand the concept of "controls" is all that is needed to interpret MCA coefficients. MCA is particularly useful for describing the results of models employing dichotomous variables. In this instance the MCA coefficients can be interpreted as a "likelihood" or "probability." Unfortunately, dummy dependent variables violate several assumptions of the linear regression model. Logit, a log-linear, maximum likelihood, multivariate technique is correctly employed to explain relationships between dichotomous dependent variables and one or more predictor variables. [2, p. 292].

In this paper a technique is developed which shows how both logit and MCA can be adapted for comparison purposes. The technique is tested using data from the National Longitudinal Surveys data of young men 14 to 24 in 1966. Two dichotomous dependent variables are employed to test the technique and the data revealed that the logit probabilities, with few exceptions, were consistent with those generated by MCA. The author advocates the use of this technique to confirm the MCA coefficients. Furthermore, when MCA and logit yield consistent results MCA coefficients should be used for presentation and discussion purposes.

Multiple Classification Analysis

Multiple Classification analysis is a multivariate statistical technique for exploring the interrelationships between explanatory variables and a dependent variable in the context of an additive model [1,p. 1]. It is a version of multiple regression, without the intercept term, where all predictor variables are expressed in categorical form. The grand mean or mean of the dependent variable is roughly equivalent to the constant term in multiple regression. The MCA technique permits one to calculate the mean value of the dependent variable for each category of a particular explanatory variable "adjusted" for the effects of all other variables in the model. Differences in these values among the several categories of a given variable may be interpreted as indicating the "net" effect of that variable upon the dependent measure.

There are many advantages associated with the use of the MCA technique. Andrews et al identify MCA's ability to overcome the problem of correlated predictors as one of its chief strengths. For example, if the explanatory variables are positively correlated the increase in explanatory power provided by two independent variables would be less than the sum of the increase in explanatory power provided by each separately. The unadjusted and adjusted coefficients provide information about the combined and independent influence of the independent variables. The raw effect of the predictor variable is represented through the unadjusted mean. The adjusted figure takes into account or controls for the influence of the other explanatory variables in the model [1, p. 3].

Aside from its ability to deal with the problem of correlated explanatory variables, MCA is appropriate for analyzing qualitative data (nominal or ordinal scale) [1, p. 3]. Policy research often uses individuals as the unit of analysis. Many important variables such as attitudes and attributes cannot be placed on an interval or ratio scale. Interval or ratio data correctly fit the assumptions of classical regression analysis. In contrast, the MCA independent variables are always treated in categorical form. It makes no difference whether a particular category represents a nominal, ordinal, interval or ratio scale [1, p. 5].

Nominal and ordinal scale predictors when transformed to dummy variables are correctly employed in the regression equation. The regression coefficients, however, are difficult to interpret. For each set of dummy variables there must be an omitted catetory. Without an omitted category the regression model would suffer from pure or perfect multicollinerity. These regression coefficients represent deviations from the omitted category. The grand mean is unknown and the intercept term in regression is a composite sum of the means of the omitted category [1, p. 52]. Adjusted MCA coefficients, in contrast, represent deviations from the grand mean controlling for all other variables in the model.

MCA is often employed using dichotomous dependent variables (for example, see [8], [9], [11], [14], [18]). Unfortunately, dichotomous dependent variables violate several statistical assumptions of the regression model. MCA,as a form of regression,also suffers from these problems. First, the dependent variable and the error term are not distributed normally. Secondly, heteroscedasticity pervades the model (variance not constant). And third, the expected value of the estimate can go outside the (0, 1) range [4], [20]. Hence, it is questionable whether MCA is an appropriate technique to analyze models using dichotomous dependent variables.

MCA and Logit: A Method of Comparison

The problems associated with the use of dichotomous dependent variables can be overcome through the use of the logistic transformation. Logit analysis is employed to examine the relationship of a dummy dependent variable to one or more predictor variables. Logit modesl are maximum likelihood, log linear models. They are analogous to regression models for which the expected value of a continuous dependent variable is a linear function of one or more explanatory variables [7, p. 292]. After the logistic transformation, the range of the predicted value of the dichotomous dependent variable is no longer constrained to values between 0 and 1. Values between negative infinity and infinity are legitimate. Additionally, the problem of heteroscadesticity is overcome.

*The research reported is part of the National Longitudinal Surveys being conducted by the Ohio State University Center for Human Resource Research under contract with the Employment and Training Administration U.S. DOL. I wish to thank Tim Carr, Gilbert Nestel and Jane Tally.

The purpose of this section is to indicate the similarity between the MCA coefficients and those obtained through logit analysis. In order to make the comparison it is necessary to transform the coefficients (or "p" in the case of logit) obtained through the two different techniques. The method of transforming the coefficients is discussed in the remainder of this section. The logistic model is of the following form:

$$\ln(P_t/_{1-P_t}) = X_t'B \qquad (1)$$

where:

B = vector of parameter coefficients

p_t = probability for the t-th individual

x_{jt} = vector of independent variables for the t-th individual

From (1) one can derive

$$p_t = [1 - e^{-X_t'B}]^{-1} \qquad (2)$$

P_t is also referred to as the log odds [7, p. 292] and is often interpreted as a probability. We are interested in a transformed P_t since it most closely resembles the adjusted MCA percentages.

Differentiating P_t with respect to X_{jt}

$$\frac{\partial p_t}{\partial X_{jt}} = B_j \; [1 - e^{-X_t'B}] \quad {}^{-2}e^{-X_t'B} \qquad (3a)$$

$$= B_j p_t (1-p_t) \qquad (3b)$$

where:

B_j = the jth element of B

X_{jt} = the jth element of X_t

Both B_j and p are unknown parameters and must be estimated. The B_j's are simply estimated using coefficients from the logit analysis. The method of estimating p is a somewhat more complex function of the B_j's.

The MCA technique permits one to calculate the mean value of the dependent variable for all categories of predictor variables. Each coefficient measures the "net" effect of that category (variable) upon the dependent variable after controlling for the effects of all other predictor variables. For example, when the dependent variable is dichotomous, the MCA technique allows one to calculate for each independent variable category, the proportion of that category which would have taken the value "1" on the dependent measure had the members of that category been "average" in terms of all other variables entering into the analysis.

Hence, it was necessary to evaluate the predicted proportions from the logit analysis in a way which would hold "average" all categories. To do this p* was estimated by evaluating it at the mean of each independent variable.

$$p^* = [1 - e^{-\bar{X}'b}]^{-1} \qquad (4)$$

where b is the vector of estimated logit coefficients and X is a vector consisting of the sample means of all independent variables.

The logit analysis are calculated substituting the estimates for the parameters of (3b) and evaluating $b_j p^* (1-p^*)$ where p* is found by evaluating (4).

The logit program requires that one of the categories be omitted when the predictor variable is described by a series of dichotomous variables. In contrast, the MCA procedure does not require that one of the categories for each predictor variable be omitted. Hence, it was necessary to transform the MCA adjusted proportions into deviations from the omitted category in order to make them consistent with the logit results. This was done using the following formula:

$$C_{ij} - C_{i\ \text{omitted}} = C_{ij*}$$

where:

C_{ij} = the adjusted percentage of category $_j$ of predictor variable i

$C_{i\ \text{omitted}}$ = the adjusted percentage of the category of predictor variable i which corresponds to the omitted logit

C_{ij*} = the transformed MCA coefficient which corresponds to C_{ij}

The Empirical Test

The data used to test the method developed in the previous section are based upon information collected in the National Longitudinal Surveys NLS) of the labor market and educational experiences of young men. The entire sample contains 5225 respondents age 14 to 24 when first interviewed in the fall of 1966. The young men were interviewed annually through 1971 and less frequently thereafter.

The author was interested in exploring several aspects of the determinants of military service throughout the Vietnam era. The sample is well suited for this purpose since about 30 percent of the young men interviewed entered the armed forces.

Two dichotomous dependent variables were employed: 1) the likelihood of enlisting, and 2) the likelihood of serving. Independent variables included education, health, fatherhood status, region of residence, socioeconomic status, wages, unemployment, mental ability, and draft pressure. [17]

In Table 1 MCA and logit coefficients are compared using the technique discussed earlier. All in all, the logit and MCA technique yielded similar coefficients. Differences in the magnitude of the predicted results are rare. Approximately 96% of the 108 possible comparisons were within 5 percentage points of one another. Perhaps more importantly, the few differences greater than 5 percentage points did not change the interpretation of the results. Through this method a scholar can feel confident of the MCA results. In addition, the advantages of MCA remain.

TABLE 1

Comparison of the Net Relationships Between the Likelihood of Service and Enlisting and Categories of the Explanatory Variables Using Logit and MCA Analysis

	Likelihood of Service				Enlistment			
Explanatory	Whites		Blacks		Whites		Blacks	
Variables	Logit	MCA[a]	Logit	MCA[a]	Logit	MCA[b]	Logit	MCA[b]
Education								
0-8	omit	omit	omit	omit	omit	omit	omit	omit
9-11	.07	.06	.18	.13	.01	.01	.12	.10
12	.13	.13	.27	.22	.08	.10	.14	.12
13-15	.13	.14	.20	.15	.07	.09	.11	.12
16	.16	.17	.06	.06	.10	.12	.01	.00
17+	.18	.19	c	c	.03	.09	c	c
Dependents								
none	omit	omit	omit	omit	omit	omit	omit	omit
some	-.30	-.20	-.19	0.14	-.20	-.13	-.98	-.06
Draft Pressure								
low	omit	omit	omit	omit	omit	omit	omit	omit
high	.21	.23	.18	.17	.11	.14	.05	.06
Residence								
N.E.Central City	-.06	-.06	.16	.13	.00	.00	.06	.03
N.E.Other	-.06	-.05	.04	.04	-.01	-.01	.00	.00
N.Central City	-.09	-.08	.21	.18	-.05	-.05	.10	.06
N.Central Other	-.04	-.04	.16	.14	-.02	-.02	-.07	-.04
South Urban	-.03	-.03	.21	.16	.01	-.01	.20	.16
South Rural	-.11	-.11	.20	.15	-.05	-.07	.21	.18
West	omit	omit	omit	omit	omit	omit	omit	omit
Health Problems								
no	omit	omit	omit	omit	omit	omit	omit	omit
yes	-.35	-.25	-.14	-.12	-.25	-.18	0.16	-.11
Ability								
Low	omit	omit	omit	omit	omit	omit	omit	omit
Medium	.02	.02	.03	.04	.01	.01	.08	.10
High	.02	.02	c	c	.02	.02	c	c
N/A	-.03	-.02	.09	.10	-.02	-.02	-.02	-.02
Potential Wage								
low	omit	omit	omit	omit	omit	omit	omit	omit
medium	-.08	-.10	.00	-.02	-.03	0.07	.04	.05
high	-.32	-.30	-.15	0.15	-.20	0.22	.04	.03
N/A	-.10	-.12	-.00	-.03	0.16	-.19	.07	.08
Unemployment								
low	-.12	-.10	-.13	-.09	-.08	0.07	-.01	0.01
medium	omit	omit	omit	omit	omit	omit	omit	omit
high	.03	.04	.05	.05	.01	-.01	-.07	-.06
N/A	-.07	-.06	.00	.01	-.05	-.05	.02	.02
SES								
low	omit	omit	omit	omit	omit	omit	omit	omit
medium	.00	.00	.01	.00	.00	.00	-.01	-.02
high	.00	.00	-.03	-.03	.01	.02	-.07	-.07
N/A	.00	-.03	-.04	-.03	-.05	-.05	.10	.11

[a] These figures represent deviations from the omitted categories using the adjusted percentages foun in [18].

[b] These figures represent deviations from the omitted categories using the adjusted percentages found in [17].

[c] The results not shown when they represent fewer than 20 sample cases.

REFERENCES

[1] Andrews, F.,Morgan, J., Sonquist, Jr., and Klem, L., Multiple Classification Analysis: A Report on a Computer Program for Multiple Regression Using Categorical Data (Ann Arbor, Michigan: Survey Research Center, Institute for Social Research, The University of Michigan, 1967).

[2] Appelbaum, Richard P. and Follett, Ross., "Size, Growth, and Urban Life: A Study of Medium-Sized American Cities." Urban Affairs Quarterly, Vol. 14, No. 2 (Dec. 78), pp. 138-168.

[3] Atkinson, Michael, "Policy Interests of Provincial Backbenches and the Effects of Political Ambition" Legislative Studies Quarterly, Vo. 3, No. 4 (Nov. 78), pp. 629-645.

[4] Cox, D.R., Analysis of Binary Data. (London: Methuen, 1970).

[5] DuMouchel, William H., "On the Analogy Between Linear and Log-Linear Models." Technical Report No. 67, Department of Statistics, University of Michigan, 1978.

[6] Gundlach, James H. and Roberts, Alden E.,"Native American Indian Migration and Relocation: Success or Failure." Pacific Sociological Review, Vol. 21, Issue 1 (Jan. 78), pp. 117-128.

[7] Haberman, Shelby J., Analysis of Qualitative Data: Volume I Introductory Topics. (New York: Academic Press, 1978).

[8] Johnston, J. and Bachman, J.G., Youth in Transition Volume V: Young Men and Military Service. (Ann Arbor, Michigan: Institute for Social Research, 1972).

[9] Kendel, D.B., Treiman, D., Faust, R. and Single, E., "Adolescent Involvement in Legal and Illegal Drug Use." Social Forces, Vol. 55, No. 2, pp. 438-458, 1976.

[10] Knoke, David, "A Comparison of Log-Linear and Regression Models for Systems of Dichotomous Variables." Sociological Methods and Research 3, (1975), pp. 416-434.

[11] Kohen, Andrew I. and Shields, Patricia M., "Reaping the Spoils of Defeat: Labor Market Benefits of Vietnam Veterans." Strangers at Home: Vietnam Veterans Since the War. Ed.: C.R. Figley and S. Leventman. (New York: Praeger, 1980).

[12] McPherson, J. Miller., "Correlates of Social Participation: A Comparison of the Ethnic Community and Compensatory Theories." Sociological Quarterly, Vol. 18, No. 2., (Spring 1977),pp.197-208.

[13] Netter, John and Wasserman, William. Applied Linear Statistical Models. (Homewood, Illinois: Irwin, 1974).

[14] Parnes, Herbert S. and Nestel, Gilbert. "Early Retirement." The Pre-Retirement Years, Vol. IV. (Columbus, Ohio: The Center for Human Resources Research, 1974). pp. 153-196.

[15] Ridgeway, Cecilla L., "Predicting College Women's Aspirations from Evaluations of the Housewives Work Role." Sociological Quarterly Vol. 19, No.2, (Sp. 1978) pp. 281-291.

[16] Shammas, Carole, "The Determinants of Personal Wealth in 17th Century England and America," Journal of Economic History (Vol. 37, No. 3 (Sept. 1977), pp. 675-689.

[17] Shields, Patricia M., "Enlistment during the Vietname Era and the 'Representation' Issue of the AVF." Armed Forces and Society, forthcoming.

[18] Shields, Patricia M., The Determinants of Service in the Armed Forces During the Vietnam Era,(Columbus, Ohio: The Center for Human Resources Research, 1977).

[19] Suits, D.B. "Use of Dummy Variables in Regression Equations." Journal of the American Statistical Association, 1957, 52, 548-551.

[20] Theil, Henri, Principles of Econometrics (New York: John Wiley and Sons, 1971).