A Class of Linear Multiple Regression Techniques for Ordered Attributes

By

Hidekiyo ITAKURA* and Yoshikazu NISHIKAWA*

(Received May 28, 1980)

Abstract

A class of linear multiple regression techniques is discussed, in which the order of magnitude is constrained among regression coefficients. Each of the predictor variables is a qualitative variate having some categories which are on an ordinal scale. The criterion variable is quantitative. The problem is to solve a quadratic programming problem, in which the objective function is the residual sum of the squares of regression, and the constraints are linear ones imposed on the regression coefficients. Under some conditions for the observed data, the problem can be solved numerically. This technique works effectively for some types of regression analysis.

1. Introduction

Techniques of linear multiple regression are very useful for multivariate analysis methods, but conventional techniques are sometimes unsuitable for analyzing some types of statistical data. Among them is a type of data concerned with qualitative variates of ordered attributes. It is sometimes found that results from a formal application of the conventional techniques to the data bewilder us in trying to interpret them. One of the bewildering points is that the order of values of regression coefficients given to the categories of each attribute seems unnatural, at least from the viewpoint of deriving a meaning of the regression under study.

This paper is concerned with a class of linear multiple regression techniques. In order to avoid such unnaturalness, the order of magnitude is constrained among the values of the regression coefficients. Let the categories of an attribute be placed on an ordinal scale. Depending on the properties of the attributes and categories in the problem under study, although there may be a variety of order relations, one of the relatively general types is considered in this paper.

---

* Department of Electrical Engineering

In the next section, a conventional linear multiple regression technique for qualitative variates is reviewed. Section 3, which is the principal part of this paper, is concerned with a procedure of multiple regression modified for ordered attributes. The significance of the presented modified multiple regression is then explained complementarily, using trivial artificial data. Finally, an equational formulation for the regression is made, with a description of its systematic computational procedure. Section 4 illustrates an application of the method to real poll data associated with some areas' inhabitants' evaluation of their local cultural environment.

2. Multiple Regression for Qualitative Variates

This section reviews the technique of linear multiple regression for qualitative variates.[1,2] The statistical data to be analyzed are compiled observations of people's attitudes or opinions derived from a questionnaire polling, or measurements of some kinds of subjective evaluations.

It is assumed that the n attributes are under study in such a type of measurement, and that each one of them consists of k_j categories ($k_j \geqq 2; j=1, 2, \cdots, n$). Each one of the N individuals is to respond to one of the categories in each item. For convenience, this response is expressed by the symbol

$$\delta_i(j, k)=\begin{cases} 1, & \text{if the individual } i \text{ responds to the category } k \text{ of the item } j \\ 0, & \text{otherwise} \end{cases}$$

for $k=1, 2, \cdots, k_j$; $j=1, 2, \cdots, n$; and $i=1, 2, \cdots, N$. Then

$$\sum_{k=1}^{k_j} \delta_i(j, k)=1$$

holds for every i and j. Moreover, it is assumed that each individual has a quantitatively measured value, Y_i, as a criterion variable.

By using the response $\delta_i(j, k)$ as predictor variables, the linear multiple regression

$$y_i=\sum_{j=1}^{n} \sum_{k=1}^{k_j} \delta_i(j, k)x_{jk} \tag{1}$$

is considered. The unknown regression coefficients x_{jk} having quantitative values, called the category score for the category k of the attribute j, are to be determined in such a way that they minimize the residual sum of squares

$$J_0=\frac{1}{2}\sum_{i=1}^{N}(Y_i-y_i)^2 \tag{2}$$

The result gives the best regression in terms of minimizing J_0.

Equation (1) is substituted into Eq. (2) and the symbol J is used for J_0, from which the constant terms independent of x_{jk} are removed. The resulting J can be written in a vector-matrix form as

$$J=\frac{1}{2}x^T Fx-g^T x \tag{3}$$

where

$$x=(x_1{}^T, x_2{}^T, \cdots, x_n{}^T)^T:\ M\text{-vector}$$

$$x_j=(x_{j1}, x_{j2}, \cdots, x_{jk_j})^T:\ k_j\text{-vector } (j=1, 2, \cdots, n)$$

$$F=\begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ F_{n1} & F_{n2} & \cdots & F_{nn} \end{bmatrix}:\ M\times M\text{-symmetric matrix}$$

$$F_{jj'}=F_{j'j}{}^T=\begin{bmatrix} f(j, 1; j', 1) & \cdots & f(j, 1; j', k_j) \\ \cdots & \cdots & \cdots \\ f(j, k_j; j', 1) & \cdots & f(j, k_j; j', k_{j'}) \end{bmatrix}:\ k_j\times k_{j'}\text{-matrix } (j\neq j') \tag{4}$$

$$F_{jj}=\text{diag}(f(j, 1; j, 1), \cdots, f(j, k_j; j, k_j)):\ k_j\times k_j\text{-diagonal matrix}$$

$$g=(g_1{}^T, g_2{}^T, \cdots, g_n{}^T)^T:\ M\text{-vector}$$

$$g_j=\left(\sum_{i=1}^{N}\delta_i(j, 1)Y_i, \cdots, \sum_{i=1}^{N}\delta_i(j, k_j)Y_i\right)^T:\ k_j\text{-vector}$$

M equals the total number of categories, while $f(j, k; j', k')$ represents the total number of individuals responding simultaneously to the category k of the attribute j and the category k' of the attribute j':

$$M=\sum_{j=1}^{n} k_j$$

$$f(j, k; j', k')=\sum_{j=1}^{N}\delta_i(j, k)\delta_i(j', k')$$

The superscript T denotes the transpose of a vector or a matrix. As seen from Eq. (1), x_{jk} minimizing Eq. (2) is not unique. That is, any one category score ($n-1$ scores in all) in each of the arbitrary $n-1$ attributes is not independent. Therefore, the equation $Fx=g$ (which, in order to determine x_{jk} minimizing J, results from differentiating Eq. (3) with respect to x and putting it to zero) is a set of simultaneous equations having indeterminate solutions. Thus, without any loss of generality, x_{j1} is put to zero ($j=2, 3, \cdots, n$) and removed from Eq. (3); and

$$\tilde{J}=\frac{1}{2}\tilde{x}^T\tilde{F}\tilde{x}-\tilde{g}^T\tilde{x} \tag{5}$$

is defined, where

$$\tilde{x}=(x_1{}^T, \tilde{x}_2{}^T, \cdots, \tilde{x}_n{}^T)^T:\ \tilde{M}\text{-vector}$$

$$\tilde{x}_j=(x_{j2}, x_{j3}, \cdots, x_{jk_j})^T:\ (k_j-1)\text{-vector } (j=2, 3, \cdots, n)$$

$$\tilde{F}=\begin{bmatrix} F_{11} & \tilde{F}_{12} & \cdots & \tilde{F}_{1n} \\ \tilde{F}_{21} & \tilde{F}_{22} & \cdots & \tilde{F}_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \tilde{F}_{n1} & \tilde{F}_{n2} & \cdots & \tilde{F}_{nn} \end{bmatrix} : \ \tilde{M}\times\tilde{M}\text{-matrix}$$

$\tilde{F}_{1j}=\tilde{F}_{j1}{}^T=(F_{1j}$ whose first column is removed)

$\tilde{F}_{jj'}=\tilde{F}_{j'j}{}^T=(F_{jj'}$ whose first column and row are removed)

$$(j, j'=2, 3, \cdots, n)$$

$\tilde{g}=(g_1{}^T, \tilde{g}_2{}^T, \cdots, \tilde{g}_n{}^T)^T$

$\tilde{g}_j=(g_j$ whose first component is removed)

$\tilde{M}=M-(n-1)$

Consequently, the problem is reduced to the minimization of $\tilde{J}$ with respect to $\tilde{x}$, and x_{jk} minimizing $\tilde{J}$ satisfies the equation

$$\frac{\partial\tilde{J}}{\partial\tilde{x}}=\tilde{F}\tilde{x}-\tilde{g}=0 \tag{6}$$

For the individual i, let us define the M-vector

$$a_i=(a_{i1}{}^T, a_{i2}{}^T, \cdots, a_{in}{}^T)^T$$
$$a_{ij}=(\delta_i(j, 1), \cdots, \delta_i(j, k_j))^T$$

It is easily seen that, if at least the $\tilde{M}$ vectors among $a_1, a_2, \cdots, a_N$ are independent, $\tilde{F}$ is positive-definite. Then, the solution of the linear algebraic equation (6) is unique.

Useful information can be obtained from the results of the best regression determined. For x_{jk} obtained in the above way, the deviation of x_{jk} from the average, weighted by the number of individuals for every attribute

$$x_{jk}'=x_{jk}-\frac{1}{N}\sum_{l=1}^{k_j} N_{jl}x_{jl}$$

is computed, where N_{jl} is the total number of individuals responding to the category l of the attribute j:

$$N_{jl}=\sum_{i=1}^{N}\delta_i(j, l)$$

The quantity x_{jk}' expresses the relative degree of influence of the category k in the attribute j on the criterion variable. Further, the correlation among

$$a_{ij}=\sum_{k=1}^{k_j}\delta_i(j, k)x_{jk}'$$
$$a_{i,n+1}=Y_i \qquad\qquad (j=1, 2, \cdots, n;\ i=1, 2, \cdots, N)$$

is obtained:

$$r_{jj'}=\frac{\sum_{i=1}^{N} a_{ij}a_{ij'}}{\sum_{i=1}^{N} a_{ij}^2 \cdot \sum_{i=1}^{N} a_{ij'}^2}$$

The partial correlation coefficient between the attribute j and the criterion variable can be calculated:

$$r_{Yj}^{*}=-\frac{r^{n+1,j}}{\sqrt{r^{n+1,n+1}r^{jj}}}$$

where r^{ij} is the (i, j) element of the inverse of the $(n+1)\times(n+1)$-matrix $R=\{r_{jj'}\}$. The value of r_{Yj}^{*} reflects the degree of the relation of the attribute j directly to the criterion variable, and the relation involves no indirect relations through the attributes other than j.

3. Constrained Multiple Regression for Ordered Attributes

The category scores x_{jk} are determined formally by minimizing the residual sum of the squares (5) in the last section. Accordingly, the order of magnitude among the resulting x_{jk} is unconstrained. However, in a case where the categories of an attribute are on an ordinal scale, e.g., good or bad, necessary or unnecessary, satisfactory or unsatisfactory, sufficient or insufficient, etc., it is sometimes natural to presuppose some ordinal relation among the scores x_{jk} for every j. Although such a presupposition is a very strong one, it is sometimes reasonable and it is often due to some type of transitivity in human judgment.

As an illustrative example, let us consider the trivial artificial data given in Table 1, in which the symbol ○ means an individual's response to one of three categories in each of two attributes. The categories of each attribute are on an ordinal scale, for example, category 1 indicating "good", 2 indicating "neutral", and 3 "bad". The criterion

Table 1. Artificial data of responses.

Individual i	Attribute 1			Attribute 2			Criterion variable Y_i	
	Category			Category				
	1	2	3	1	2	3	Case (i)	Case (ii)
1	○			○			1.0	1.0
2	○				○		2.0	2.0
3		○				○	3.0	3.0
4		○		○			2.0	3.0
5			○		○		3.0	3.0
6			○			○	4.0	3.0

Table 2. Category scores obtained for the data of Table 1.

Case	Attribute 1			Attribute 2		
	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}
(i)	$\frac{4}{3}$	$\frac{11}{6}$	$\frac{5}{2}$	0	$\frac{2}{3}$	$\frac{4}{3}$
(ii)	$\frac{4}{3}$	$\frac{17}{6}$	$\frac{5}{2}$	0	$\frac{2}{3}$	$\frac{1}{3}$

variable is such that, the smaller the value the higher the general goodness becomes. Thus, the condition should be satisfied whereby the value of y_i becomes larger for an individual responding to category 2 or 3 rather than for one responding to category 1 or 2, respectively.

Table 2 shows the category scores x_{jk} obtained by the procedure of the last section for the two cases regarding the values of the criterion variable. Although the above condition is certainly satisfied in Case (i), it is not true for Case (ii). For example, the value of y_i is $4/3+2/3=2$ in responding to category 1 of attribute 1 and category 2 of attribute 2, and $4/3+1/3=5/3$ in responding to category 1 of attribute 1 and category 3 of attribute 2. This is inconsistent with the condition. Therefore, it is reasonable that x_{jk} be determined under the constraint $x_{j1} \leqq x_{j2} \leqq x_{j3}$. The authors have sometimes experienced this type of inconsistency, when they were analyzing real data.

There could be various types of order relation among the category scores, depending on the properties of attributes and categories in the problem under study. Here, let us consider the constraints

$$x_{j1} \leqq x_{j2} \leqq \cdots \leqq x_{jk_j} \qquad (j=1, 2, \cdots, n) \tag{7}$$

which might be comparatively general, and also could be a basic form of other various types of constraints.

A possibly less sophisticated way of obtaining the category scores satisfying the condition (7) may be as follows: First, x_{jk} are computed by the procedure of the last section, discarding the constraints. Next, two neighboring categories whose scores do not satisfy the constraints (7) are merged into one category. Then, the new problem with some merged categories is dealt with so as to obtain x_{jk}. Again, the constraints (7) are checked for the new x_{jk}. The same process is repeated until all the constraints become satisfied. This is somewhat cumbersome. A method is being considered for obtaining the category scores satisfying the constraints in one step, not needing to go through a cumbersome process with many steps.

The problem is to determine $\tilde{x}$ minimizing $\tilde{J}$ of the equation (5) subject to the constraints (7). This is just one of the quadratic programming problems, and can be solved numerically under the condition, with respect to the data, that $\tilde{F}$ is positive-definite, as stated in the last section. However, making an inequality constraint of a vector-matrix form directly from the inequality (7) leads to having the constraint matrix with a number of zero-elements. This causes a considerable waste of computation time. Accordingly, the following procedure is done.

By introducing the new variables ξ_{jk}, x_{jk} are replaced as

$$
\begin{aligned}
&x_{12}=x_{11}+\xi_{12}\\
&x_{13}=x_{12}+\xi_{13}=x_{11}+\xi_{12}+\xi_{13}\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdot\\
&x_{1k_1}=x_{1,k_1-1}+\xi_{1k_1}=x_{11}+\xi_{12}+\xi_{13}+\cdots+\xi_{1k_1}\\
&x_{j3}=x_{j2}+\xi_{j3}\\
&x_{j4}=x_{j3}+\xi_{j4}=x_{j2}+\xi_{j3}+\xi_{j4}\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdot\\
&x_{jk_j}=x_{j,k_j-1}+\xi_{jk_j}=x_{j2}+\xi_{j3}+\xi_{j4}+\cdots+\xi_{jk_j} \qquad (j=2,\ 3,\ \cdots,\ n)
\end{aligned}
$$

Equivalently these are

$$\tilde{x}=S\xi$$

in the vector-matrix form, where

$$
\begin{aligned}
&\xi=(\xi_1{}^T,\ \xi_2{}^T,\ \cdots,\ \xi_n{}^T)^T:\ \tilde{M}\text{-vector}\\
&\xi_1=(x_{11},\ \xi_{12},\ \cdots,\ \xi_{1k_1})^T\\
&\xi_j=(x_{j2},\ \xi_{j3},\ \cdots,\ \xi_{jk_j})^T \qquad (j=2,\ 3,\ \cdots,\ n)\\
&S=\mathrm{diag}(S_1,\ S_2,\ \cdots,\ S_n):\ \tilde{M}\times\tilde{M}\text{-matrix}\\
&S_j=\begin{bmatrix} 1 & 0 & \cdots & 0\\ 1 & 1 & \ddots & \vdots\\ \vdots & & \ddots & 0\\ 1 & 1 & \cdots & 1 \end{bmatrix} \qquad (j=1,\ 2,\ \cdots,\ n)
\end{aligned}
$$

and S_1 is the $k_1\times k_1$-matrix and S_j is the $(k_j-1)\times(k_j-1)$-matrix for $j=2,\ 3,\ \cdots,\ n$. From this transformation, the constraints (7) are written simply as

$$\tilde{\xi}\geqq 0 \tag{8}$$

where

$$
\begin{aligned}
&\tilde{\xi}=(\tilde{\xi}_1{}^T,\ \xi_2{}^T,\ \cdots,\ \xi_n{}^T)^T\\
&\tilde{\xi}_1=(\xi_{12},\ \xi_{13},\ \cdots,\ \xi_{1k_1})^T
\end{aligned}
$$

Moreover, for the convenience of the linear programming problem appearing in Wolfe's method, which will be described later, the values of criterion variable Y_i are transformed as

$$\tilde{Y}_i=Y_i-\max_{l=1,\cdots,N} Y_l$$

and g in Eq. (3) is replaced by

$$g^*=(g_1^{*T}, g_2^{*T}, \cdots, g_n^{*T})^T$$
$$g_j^*=\left(\sum_{i=1}^{N}\delta_i(j, 1)\tilde{Y}_i, \cdots, \sum_{i=1}^{N}\delta_i(j, k_j)\tilde{Y}_i\right)^T$$

Due to this transformation, the $\tilde{M}$ components in g^* are all non-positive. The values of x_{1k} ($k=1, 2, \cdots, k_1$), obtained by using this g^*, added to $\max_{l=1,\cdots,N} Y_l$ give the solution for the original Y_i.

From the above, J is rewritten as

$$J=\frac{1}{2}\xi^T A\xi+b^T\xi \tag{9}$$

where

$$A=S^T\tilde{F}S, \qquad b=-S^T\tilde{g}^*$$

and $\tilde{g}^*$ is g^* with the first component of g_j^* ($j=2, 3, \cdots, n$) removed. If $\tilde{F}$ is positive-definite, then so is A, because S is non-singular. The problem is to determine ξ minimizing J of Eq. (9) subject to the constraint (8).

The solution ξ of the problem ought to satifsy the Kuhn-Tucker condition. That is, by introducing the Lagrange multiplier u of the ($\tilde{M}-1$)-vector and defining

$$L=u^T\tilde{\xi}-J$$

the Kuhn-Tucker condition is given by

$$\frac{\partial L}{\partial \xi}=0$$

or

$$-A\begin{bmatrix} x_{11} \\ \tilde{\xi} \end{bmatrix}+\begin{bmatrix} 0 \\ u \end{bmatrix}=b \tag{10}$$

and

$$u^T\tilde{\xi}=0 \tag{11}$$
$$u\geqq 0, \qquad \tilde{\xi}\geqq 0 \tag{12}$$

Since the elements of A and the components of b are all non-negative, $x_{11}\leqq 0$ is observed from Eq. (10) together with (12). Therefore, by defining

$$z=\begin{bmatrix} x_{11}' \\ \tilde{\xi} \end{bmatrix}=\begin{bmatrix} -x_{11} \\ \tilde{\xi} \end{bmatrix}$$

Eqs. (10) and (12) are written as

$$Vz+\begin{bmatrix} 0 \\ u \end{bmatrix}=b, \quad z\geqq 0, \quad u\geqq 0, \quad V=(a, -\tilde{A})$$

where a is the first-column vector in A, and $\tilde{A}$ is A from which a is removed.

According to Wolfe's method[3] for the quadratic programming problem, a basic solution of Eq. (10) satisfying the conditions (11) and (12) minimizes J. Hence, by introducing a scalar artificial variable w, the problem is reduced to the linear programming problem of minimizing the objective function

$$\phi = w$$

subject to the constraint

$$Vz + v = b, \quad z \geqq 0, \quad v = \begin{bmatrix} w \\ u \end{bmatrix} \geqq 0$$

This can be solved by the usual simplex method, using the initial basic feasible solution $w=b$, $z=0$, $u=0$. However, since the condition (11) has to be satisfied, the basic change in the simplex method has to be executed in such a way that, if the lth component ($l \geqq 2$) of v is a basic variable, the lth component of z is a non-basic variable. Also, if the lth component ($l \geqq 2$) of z is a basic one, the lth component of v is a non-basic one. If the matrix A is positive-definite, it is assured that, unless degeneracy occurs, this execution will always give the solution upon termination.

4. Application to Real Data

The data treated here are those resulting from the summary of a poll, which was conducted by the Association for Architecture Research[4] for the National Land Agency of Japan, with regard to inhabitants' evaluations of their local cultural environments. The purpose of the poll was to investigate how inhabitants evaluate their physical and human environments in local cities, preserving traditional manners and customs, and also the role that the local cultural environment plays in those general living environments. The poll was conducted in three local cities in Japan. 3057 families were chosen randomly, among which 2428 families responded, meaning a high 80% response.

Several inquiries were made in the questionnaire concerning the above aims. The principal part of the questionnaire consisted of thirty items, concerning physical environment mainly from the viewpoint of local cultures and traditional manners and customs. For each item, the people were asked to indicate the degree of their feelings of satisfaction, in one of five specified categories: category 1—very satisfactory, 2—satisfactory, 3—indifferent, 4—unsatisfactory, and 5—very unsatisfactory. The items used in this section were the following sixteen items among the thirty:

2. A religious institution like a temple, a shrine, or a church.
3. Precincts of a temple or a shrine.
4. A statue of a children's guardian or travelers' guardian by a roadside.
5. A sacred or aged tree, a stone monument, or a shrine gate by a roadside.

6. A historic building like an old house of a prominent person.
7. An old-style residence on a street.
10. Stone steps, a stone wall, or a stone pavement.
13. View of a temple or a pagoda from one's home or neighborhood.
14. Some annual celebrations of their family, like annual festivals in certain seasons, or the Festivals of Star Vega in July.
15. Their own community's celebration of a traditional local festival.
17. Town association's activities for their neighbors.
18. Decoration of their homes with a traditional drapery or a Japanese paper lantern on a festival.
19. An early morning fair or an evening fair.
20. A street named with a historic atmosphere.
21. A very curvy road preserved from old times.
28. General traditional heritages in their town or community life.

The first fifteen items were used as the attributes for predictor variables in the regression analysis, and the last item was used for the criterion variable. This analysis of regression aimed at observing the unconscious elements of traditional heritages.

Categories 1 and 5 in every item were merged with categories 2 and 4, respectively, in advance, because there were very few responses to category 5 (very unsatisfactory) in some items, less than one percent of the total responses. Further, although item 28, used for the criterion variable, is on an ordinal scale as well as the other items, this

Table 3. Average of observed values of the criterion variable from individuals responding to each category.

Item j	$\bar{Y}_{j2}$	$\bar{Y}_{j3}$	$\bar{Y}_{j4}$
2	2.657	2.792	3.115
3	2.648	2.795	3.006
4	2.562	2.779	3.053
5	2.554	2.785	2.983
6	2.621	2.767	2.983
7	2.587	2.783	3.066
10	2.616	2.736	2.973
13	2.556	2.797	3.000
14	2.667	2.770	3.071
15	2.676	2.768	3.011
17	2.636	2.798	2.987
18	2.608	2.772	3.022
19	2.572	2.733	2.983
20	2.565	2.806	3.079
21	2.550	2.799	2.959

criterion variable was regarded as taking the values 2.0 corresponding to categories 1 and 2, 3.0 to category 3, or 4.0 to categories 4 and 5. In this analysis, it is assumed that because every one of those fifteen items could be one of the factors constituting traditional heritages, the more satisfactory the item the smaller the value of the criterion variable should be. One of justifications of this assumption is described by Table 3. This table shows the value

$$\bar{Y}_{jk}=\frac{1}{N_{jk}}\sum_{i=1}^{N}\delta_i(j,k)Y_i$$

namely, the average of the observed values of the criterion variable from people who responded to ecah category of a predictor variable. From the table, it is seen that for every item, without exception, the average value corresponding to the category indicating a high degree of satisfaction is smaller than the average value corresponding to the category indicating a high degree of unsatisfaction. Therefore, on the average, the present assumption is true.

After the preliminary processing of the data, the original method in Section 2 and the present modified method in Section 3 were applied to the data of 1838 valid responses. (The remaining 590 responses gave no answer to any of the above sixteen items under study.) The computed values of the category scores are shown in Table 4. Some interesting points are observed from the table. It had been expected that $x_{j2}\leqq x_{j3}\leqq x_{j4}$. The results by the original method do not necessarily meet this expectation.

Table 4. Category scores computed by the two methods—by using full valid data (1838 responses).

Item j	Original method					Modified method				
	x_{j2}'		x_{j3}'		x_{j4}'	x_{j2}'		x_{j3}'		x_{j4}'
2	−0.016		0.002		0.124	−0.014		−0.002		0.136
3	−0.004		−0.002		0.027	−0.002	=	−0.002		0.015
4	−0.045		0.003		0.115	−0.035		0.001		0.097
5	−0.061		0.024		0.031	−0.042		0.017	=	0.017
6	0.075	>	−0.027	>	−0.071	0.0	=	0.0	=	0.0
7	−0.067		0.022		0.120	−0.035		0.004		0.099
10	0.028	>	−0.020		0.042	−0.008	=	−0.008		0.045
13	−0.063		0.028		0.041	−0.051		0.022		0.038
14	−0.025		0.002		0.148	−0.018		−0.005		0.142
15	0.028	>	−0.046		0.018	−0.003	=	−0.003		0.023
17	−0.047		0.022		0.113	−0.038		0.013		0.109
18	−0.027		0.014		0.021	−0.022		0.011		0.020
19	−0.051		−0.018		0.116	−0.044		−0.020		0.114
20	−0.068		0.026		0.122	−0.062		0.021		0.121
21	−0.117		0.043		0.114	−0.112		0.042		0.107

Table 5. Category scores computed by the two methods—by using half valid data (919 responses).

Item j	Original method					Modified method				
	x_{j2}'		x_{j3}'		x_{j4}'	x_{j2}'		x_{j3}'		x_{j4}'
2	−0.008		−0.006		0.109	−0.008	=	−0.008		0.127
3	−0.016		0.017	>	0.011	−0.009		0.008	=	0.008
4	−0.035		−0.007		0.131	−0.024		−0.009		0.113
5	−0.025		0.012	>	0.006	−0.017		0.007		0.009
6	0.050	>	−0.030		0.021	−0.002	=	−0.002		0.016
7	−0.052		0.011		0.112	−0.033		−0.000		0.104
10	0.008	>	−0.009		0.024	−0.006	=	−0.006		0.033
13	−0.094		0.031		0.118	−0.087		0.027		0.116
14	−0.029		0.002		0.164	−0.020		−0.006		0.155
15	0.037	>	−0.048		−0.029	0.0	=	0.0	=	0.0
17	−0.038		0.006		0.126	−0.026		−0.003		0.108
18	−0.020		0.018	>	−0.011	−0.018		0.011	=	0.011
19	−0.080		0.007		0.083	−0.076		0.005		0.080
20	−0.060		0.026		0.108	−0.052		0.023		0.097
21	−0.124		0.037		0.156	−0.119		0.034		0.154

That is, there are pairs of two neighboring category scores where the order of values is reversed. (This case is indicated by the symbol > in the table.) By way of an extreme example, it is seen that the more satisfactory item 6 is, the more unsatisfactory is item

Table 6. Category scores computed by the two methods—by using the remaining half valid data not used in Table 5.

Item j	Original method					Modified method				
	x_{j2}'		x_{j3}'		x_{j4}'	x_{j2}'		x_{j3}'		x_{j4}'
2	−0.025		0.010		0.154	−0.021		0.007		0.147
3	0.000	>	−0.011		0.052	−0.003	=	−0.003		0.027
4	−0.045		0.009		0.088	−0.039		0.010		0.063
5	−0.091		0.038	>	0.033	−0.060		0.024	=	0.024
6	0.094	>	−0.021	>	−0.171	0.0	=	0.0	=	0.0
7	−0.079		0.028		0.140	−0.038		0.011		0.080
10	0.055	>	−0.030		0.052	−0.006	=	−0.006		0.037
13	−0.027		0.026	>	−0.045	−0.007		0.004	=	0.004
14	−0.027		0.011		0.118	−0.018		0.005		0.093
15	0.020	>	−0.050		0.088	−0.009	=	−0.009		0.083
17	−0.057		0.045		0.085	−0.049		0.030		0.102
18	−0.031		0.008		0.055	−0.024		0.005		0.049
19	−0.025	>	−0.045		0.175	−0.035	=	−0.035		0.155
20	−0.083		0.027		0.145	−0.072		0.019		0.140
21	−0.112		0.051		0.069	−0.105		0.051		0.057

Table 7. Root mean square and maximum value of the differences between the category scores.

Two tables compared	Original method		Modified method	
	d	d_m	d	d_m
Tables 4 and 5	0.026	0.092	0.019	0.076
Tables 4 and 6	0.030	0.100	0.020	0.060

28. In contrast with these situations, the results by the modified method are reasonable. Also, it is interesting to note that in the case of the original method, every pair of two neighboring category scores having the same values (indicated by the symbol = in the table) or nearly same values, are in accord with the pairs of the scores between which the symbol $>$ appears.

Tables 5 and 6 show similar results. The 1838 valid responses were divided into two random groups of equal numbers. Table 5 uses one group and Table 6 uses the other. The universe of the data used in Tables 4, 5, and 6 is the same, but since the sets of samples used are different from each other, the results shown in these tables are not necessarily the same. Notwithstanding, it is desirable that these three results should be similar to each other as much as possible, due to those from the same population. In order to see the similarity, the root mean square, d, of the differences between the category scores in Tables 4 and 5 is calculated for each method. The same thing is done for the differences between the category scores in Tables 4 and 6. Further, the maximum value, d_m, of the differences is also examined. Table 7 is a summary of these values. It is to be desired that these should be small. Hence, it is observed that the modified method is better than the original one, since the values d and d_m of the modified method are smaller than those of the original method.

5. Concluding Remarks

A modified technique of linear multiple regression has been discussed, in which some linear constraints were imposed on the magnitude of category scores. The computational procedure was reduced to solving a simple problem of quadratic programming. The problem can be solved numerically by using Wolfe's method, under the condition that there is available a certain amount of observed data independent of each other.

The procedure here presented has been applied to real data. The problem has fifteen attributes, and each attribute indicates the degree of satisfaction about a certain matter. It also has three categories, which were reduced from the original five categories. The data had features suitable for the method. In particular, the average of the observed values of a criterion variable from people responding to the category indicating a high degree of satisfaction is always smaller than that indicating a high degree of unsatisfaction. As a consequence of this numerical computation, the expected results have been obtained: Pairs of two category scores, obtained by the original method, violating the presumed constraints are turned into the same or nearly the same values by the modified method. Further, when sets of data from different groups of people but from the same universe, have been analyzed, the dispersion of the category scores by the modified method is less than that by the original method.

The idea of the modified method is fairly simple and, if the method is used carefully, it will be very useful for a type of regression analysis. Although the procedure in Section 3 has been formulated for the same type of constraints for all the attributes, the formulation is also possible for some other types of constraints: for example, a case where category scores are constrained for some attributes but not for others.

Acknowledgments

The authors would like to sincerely acknowledge the kind consideration of the members of the Local Living Environmental Planning Laboratory, Department of Architecture, Kyoto University. Chaired by Professor Koji Nishikawa, it was these members who made available the polling data to the authors. Also, the authors would like to thank Mr. Hideo Yoshida of the Urban Planning Bureau, Kyoto City Government, for his helpful advice. They would also like to thank Mr. Kazuhiko Fukuda, presently with the Fuji Electric Co., Ltd., for his cooperation regarding computer programming codings while a student at Kyoto University.

References

1) C. Hayashi: On the Prediction of Phenomena from Qualitative Data and the Quantification of Qualitative Data from the Mathematico-Statistical Point of View, Annals of Institute of Statistical Mathematics, **3**, 69–98 (1958).
2) N. R. Draper and H. Smith: Applied Regression Analysis, John Wiley, New York (1966).
3) P. Wolfe: The Simple Method for Quadratic Programming, Econometrica, **27**, 382–398 (1959).
4) Association for Architecture Research: Report of Investigation on Utilization of Local Cultural Assets for Regional Development, Fundamental Investigation on Regional Development Planning, Nos. 50-11-1 and 2, Commissioned by National Land Agency (1976).