106

A STUDY OF QUANTIFICATION TECHNIQUES

by

BETTY POCHEN HSU WANG

B.S., Kansas State University, 1972

A MASTER'S REPORT

Submitted in partial fulfillment of

the requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas
1978

Approved by:

Minoru Siotani
Major Professor

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

There have been some approaches to investigate scientifically the social, psychological, and biological phenomena by quantifying the qualitative data. The aim of quantification is to synthesize the numerical representation of qualitative data, not optionally but on the basis of a theoretical and statistical point of view, in order to withdraw the useful information from them to solve the individual problems. So quantification should be done only depending on the purpose of the problem under investigation: that is, "quantification should be made from the best point of view and by the most reasonable means that may answer our purpose, as we wish either to acquire some reasonable knowledge on something or to make reasonable, effective, and positive criteria how we have to act or behave ourselves in managing some affairs" Hayashi (1950). Thus it is not an arbitrary assignment of numerical values, but rather, it is an attempt to give them operationally and functionally to seemingly related qualitative data in order to utilize those data to solve the problem more efficiently and informatively.

Quantification is dependent upon the quality and number of adopted factors, the varieties of population and methods of treatment or experimental procedures.

Generally speaking, quantification can be applied to any kind of data, as long as the data can be categorized.

Hayashi (1975) has given the following two tables summarizing the general ideas of quantification methods according to the situations encountered in the actual problems:

Table 1.1 Pattern of Quantification When Outside Criteria Are Given

Numerical outside variable	One-dimensional case ···	Efficiency of prediction is the correlation coefficient (an application of regression analysis)	
	Multi-dimensional case ...	Efficiency of prediction is the vector correlation coefficient	
Categorical outside variable	Dichotomous case	Classification is based on absolute criteria	Correlation ratio (an application of discriminant analysis)
			Success rate
		Classification is based on the judgment by comparison ...	Paired comparison (Guttman's quantification)
	Classification number ≥ 3	Classification is based on the absolute criteria	Case of one dimensional classification
			... Correlation ratio (an application of discriminant analysis
		Classification is based on the judgment by comparison	Paired comparison
			Simultaneous comparison

Table 1.2 Pattern of Quantification When Outside Criteria Are Not Given

On the basis of response pattern	Association between 2 items	... Maximization of correlation ratio or correlation coefficient	
	Association for 3 or more items	... Corresponding to B factor analysis	
On the basis of the relation among elements	Numerical Case	Relation between 2 items	e_{ij}-type
			K-L type
		Relation among 3 or more items	e_{ijk}, e_{ijkl}-types, etc.
			Relation between i and j via k (Torgerson's multi-dimensional scaling)
	Non-numerical Case	Rank order or ordered groups are given	Shepard method, Kruskal method, the smallest space analysis (Guttman)
			Minimum dimensional analysis (Hayashi)
		Results of paired comparison are given	

In this report some special methods of quantification will be discussed with illustrative numerical examples. We shall mainly review only the one-dimensional case.

In Chapter II, we shall treat the quantification when the judgments are obtained by paired comparisons. Two cases are considered: the case of ordinary comparison (Section 2.1) where things compared may be items or objects themselves, and the case (Section 2.2) where the comparisons are made on combinations of items or objects.

Chapter III contains the quantification methods when an outside criterion is given. When an outside variable is numerical (Section 3.1), the quantification will be on the basis of the ideas of prediction or regression, while for the case where an outside variable is categorical (Section 3.2), the quantification will be done by applying the idea in the discriminant or classification analysis. For each case some artificial numerical example will be given to illustrate how to compute the desired numerical values.

Finally we shall consider, in Chapter IV, a quantification method of giving numerical values to types of persons and factors through their association.

As seen in Tables 1.1 and 1.2, there are various quantification methods according to the actual situations or purposes of investigation. The present reporter wishes to continue the study on other quantification methods, in particular, the multi-dimensional quantification methods.

It should be noted that, although the numerical examples given below are artificial and based on a small numer of observations, the method is essentially for the case of a large number of observations; so in the actual application, we need to keep this in mind.

CHAPTER II

THEORY OF QUANTIFICATION WHEN THE JUDGMENTS ARE OBTAINED BY PAIRED COMPARISON

2.1 Introduction

The problem of paired comparison arises when it is desired to obtain numerical values for a set of n things with respect to one characteristic such that these values will represent the judgments of population of N individuals.

It is noted that in comparing two things at a time, inconsistencies may be allowed to appear within the judgment of an individual, while it is sometimes harder in practice for people to judge n things simultaneously than to compare them two at a time, hence in this case a paired comparison method is applied.

The judgment varies from person to person and the problem is to determine a set of numerical values for the things compared so that they will, in some sense, best represent or average the judgment of the whole population.

Now let us define:

(a) Ordinary comparison is for the case where the things compared may be single items or objects,

(b) Comparison of combination of things or objects is for the case where the things being compared may be a combination of items or objects.

This section is devoted to the presentation of quantifying comparisons or rank orders with applications to the ordinary comparison and to the comparison of combination of two things. An example of a major practical use of this approach was given by Guttman (1946) on the demobilization score card of the United States Army. The problem was to determine the number of points to assign each of the variables on the score card according to the opinions of soldiers themselves. In a survey of enlisted men throughout the world by means of a questionnaire administered by field teams of the Research Branch, there were five variables to be considered on the score card in order to determine order of demobilization. They were:

1. length of time in the Army,
2. length of time overseas,
3. amount of combat,
4. age,
5. number of children.

Thus the problem there was to determine how much weight to give each of these variables in obtaining total scores. For this case the ordinary paired comparisons are not suitable. For example, one may ask, "Who should get out first after the war: a man who has two children or a man who has been in two battles?" But respondents certainly refuse to judge such a comparison because there is insufficient basis for judgment; the battle experience of the first man is not specified, and the number of children of the second man is not given.

Therefore, in actual research on this example, judgments were based on the comparison of combinations of items in the following fashion:

"Here are three men of the same age, all overseas the same length of time. Check the one you would want to have let out first:

____ A single man . . . through two campaigns of combat

____ A married man with no children . . . through one campaign of combat

____ A married man with two children . . . not in combat."

In this section, we however shall discuss on the quantification both for the case of the ordinary paired comparisons and for the case of comparison of combinations of two things. The basic principle in deriving numerical values for things being compared requires that the values of things a given person judges higher than other things should be as different as possible from the values of the things he judges to be lower than other things; in other words, our principle calls for minimizing the variation within individuals compared with that within the group as a whole.

2.2 Ordinary comparisons

Let $O_1, O_2, \cdots, O_n$ be n things to be compared. Each of N individuals is asked to make judgments of the form that

O_j is higher (or lower) than O_k $(j \neq K)$. We assume that judgments of equality are excluded and that all people compare all the pairs. Hence there are N sets of $n(n-1)/2$ comparisons. Let

$$\ell_{jK}^{(i)} = \begin{cases} 1, \text{ if individual i judges } O_j > O_K \\ 0, \text{ if individual i judges } O_j < O_K \\ 0, \text{ for } j = K. \end{cases} \qquad (2.1)$$

for $i = 1, \cdots, N$ and $j, k = 1, \cdots, n$. Then it is obvious that

$$\begin{aligned} &\ell_{jK}^{(i)} = 1 \quad \Rightarrow \quad \ell_{kj}^{(i)} = 0 \\ &\ell_{jK}^{(i)} + \ell_{kj}^{(i)} = 1 \qquad (j \neq K). \end{aligned} \qquad (2.2)$$

Let now $f_j^{(i)}$ be the number of things such that the individual i judged to be lower than O_j, and let $g_j^{(i)}$ be the number of things such that he judged to be higher than O_j. Then

$$f_j^{(i)} = \sum_k \ell_{jk}^{(i)}, \qquad g_j^{(i)} = \sum_k \ell_{kj}^{(i)} \qquad (2.3)$$

and

$$f_j^{(i)} + g_j^{(i)} = \sum_{k \neq j} (\ell_{jK}^{(i)} + \ell_{Kj}^{(i)}) = n-1. \qquad (2.4)$$

Let

$$F = \frac{1}{2}n(n-1) = \text{the total number of comparisons made by each person}$$

$$= \sum_k f_k^{(i)} = \sum_k g_k^{(i)} \qquad (2.5)$$

c = the number of times each O_j was judged in the whole experiment,

$$= N(n-1) = \sum_i (f_j^{(i)} + g_j^{(i)}) \qquad (2.6)$$

C = the total number of judgments in the experiment

$$= nc = Nn(n-1). \qquad (2.7)$$

Now then, let x_j be the numerical value to be given for O_j on the basis of the comparisons. In order to calculate the sum of squares B between individuals and the sum of squares W within individuals, let

$$t^{(i)} \equiv \frac{1}{F} \sum_k x_k f_k^{(i)}$$ = the mean of the x values of the things individual i ranked higher than the other things (2.8)

$$y^{(i)} \equiv \sum_k (x_k - t^{(i)})^2 f_k^{(i)} = \sum_k x_k^2 f_k^{(i)} - Ft^{(i)^2} \qquad (2.9)$$

and similarly let

$$u^{(i)} \equiv \frac{1}{F} \sum_k x_k g_k^{(i)}$$ = the mean of the x values of the things individual i ranked lower than the other things, (2.10)

$$z^{(i)} \equiv \sum_k (x_k - u^{(i)})^2 = \sum_k x_k^2 g_k^{(i)} - Fu^{(i)^2}. \qquad (2.11)$$

Let V be the mean of all the x-values in the experiment:

$$V = \frac{1}{C} \sum_k x_k c = \frac{1}{n} \sum_k x_k. \qquad (2.12)$$

Then the total sum of squares T for the experiment is defined by

$$T = \sum_k (x_k - V)^2 c = c\sum_k x_k^2 - V^2 C \tag{2.13}$$

which is the sum of B and W; T = B + W, where

$$B = \sum_i [(t^{(i)}-V)^2 + (u^{(i)}-V)^2]F$$
$$= F\sum_i (t^{(i)^2} + u^{(i)^2}) - V^2 C \tag{2.14}$$

$$W = \sum (y^{(i)} + z^{(i)}) = T - B \tag{2.15}$$

Now our principle is to quantify the judgments by obtaining the x-values such that they make W as small as possible compared with T, which is equivalent to making B as large as possible compared with T. Thus if we define the correlation ratio η by

$$\eta^2 = \frac{B}{T} = 1 - \frac{W}{T}, \tag{2.16}$$

then the problem is to determine the x_j that will maximize η^2.

Since η^2 is invariant with respect to translations of the x-values, we can without loss of generality set

$$V = 0; \tag{2.17}$$

hence B and T are expressed as

$$B = F\sum_i (t^{(i)^2} + u^{(i)^2}), \qquad T = c\sum_k x_k^2 \tag{2.18}$$

Now let us find the maximizing x_j for η^2 by $d\ ^2/dx_j = 0$, which gives us

$$\frac{\partial B}{\partial x_j} = \eta^2 \frac{\partial T}{\partial x_j}, \quad j=b, \cdots, n \tag{2.19}$$

From (2.18), we obtain

$$\frac{\partial B}{\partial x_j} = \frac{2}{F} \sum_i [t^{(i)} \frac{\partial t^{(i)}}{\partial x_j} + u^{(i)} \frac{\partial u^{(i)}}{\partial dx_j}]$$

$$= \frac{2}{F} \sum_i [(\sum_k x_k f_k^{(i)}) f_j^{(i)} + (\sum_k x_k g_k^{(i)}) g_j^{(i)}]$$

$$= \frac{2}{F} \sum_k x_k \sum_i (f_j^{(i)} f_k^{(i)} + g_j^{(i)} g_k^{(i)})$$

$$\frac{\partial T}{\partial x_j} = 2cx_j.$$

Let

$$h_{jk} = \frac{1}{cF} \sum_i (f_j^{(i)} f_k^{(i)} + g_j^{(i)} g_k^{(i)}); \tag{2.20}$$

then the equation (2.19) can now be written as

$$\sum_k h_{jk} x_k = \eta^2 x_j, \quad j=1, \cdots, n \tag{2.21}$$

which are the equations to be solved numerically for the maximizing x_j.

Note 1. Summing both members of (2.21) over $j=1, \cdots, n$ and using (2.20), (2.5), and (2.6), we easily obtain

$$\sum_k x_k = \eta^2 \sum_j x_j$$

or

$$(1-\eta^2)V = 0$$

Therefore if $\eta^2 \neq 1$, we must have $V = 0$. Since a perfect correlation ratio will not occur in practice, condition (2.17) will in general be satisfied by a solution of (2.21).

Let

$$\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix}, \quad \underset{\sim}{H} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix};$$

(2.22)

then the equation (2.21) can be written in the metric form

$$\underset{\sim}{H}\underset{\sim}{x} = \eta^2 \underset{\sim}{x} \qquad (2.23)$$

The non-trivial solution $\underset{\sim}{x}$ is a characteristic vector corresponding to a characteristic root η^2 of H. Since we want the largest possible correlation ratio, our final solution $\underset{\sim}{x}_m$ is the characteristic vector corresponding to the largest root η_m^2 of the equation $|\underset{\sim}{H}-\eta^2 I_n| = 0$.

<u>Note 2</u>. H is singular, since $\sum_k h_{jk} = 1$ for $j=1, \cdots, n$ or $\sum_k h_{jk} = 1$ for $k=1, \cdots, n$. It is seen from this that $\eta^2=1$ is always the characteristic root of $\underset{\sim}{H}$, which gives us the trivial solution and should be excluded here.

Illustrative Numerical Example. Suppose there are four objects to be compared and judged by fifteen persons. Table 2 is an artificial result obtained by these people. Let us follow the procedures which are mentioned in Section 2.2. From equations (2.1) - (2.4) and Table 2, we calculate

$$f_1^{(i)},\quad f_2^{(i)},\quad f_3^{(i)},\quad f_4^{(i)} \text{ and } g_1^{(i)},\quad g_2^{(i)},\quad g_3^{(i)},\quad g_4^{(i)}$$

For example, the values of the first person are

$$f_1^{(1)}=e_{11}^{(1)}+e_{12}^{(1)}+e_{13}^{(1)}+e_{14}=3 \qquad g_1^{(1)}=e_{11}^{(1)}+e_{21}^{(1)}+e_{31}^{(1)}+e_{41}^{(1)}=0$$

$$f_2^{(1)}=e_{21}^{(i)}+e_{22}^{(i)}+e_{23}^{(i)}+e_{24}=1 \qquad g_2^{(1)}=e_{12}^{(1)}+e_{22}^{(1)}+e_{32}^{(1)}+e_{42}^{(1)}=2$$

$$f_3^{(1)}=e_{31}^{(i)}+e_{32}^{(i)}+e_{33}^{(i)}+e_{34}=1 \qquad g_3^{(1)}=e_{13}^{(1)}+e_{23}^{(1)}+e_{33}^{(1)}+e_{43}^{(1)}=2$$

$$f_4^{(1)}=e_{41}^{(i)}+e_{42}^{(i)}+e_{43}^{(i)}+e_{44}=1 \qquad g_4^{(1)}=e_{14}^{(1)}+e_{24}^{(1)}+e_{34}^{(1)}+e_{44}^{(1)}=2$$

Others are given in the following figure:

Person (i)	$f_1^{(i)}$	$f_2^{(i)}$	$f_3^{(i)}$	$f_4^{(i)}$	$g_1^{(i)}$	$g_2^{(i)}$	$g_3^{(i)}$	$g_4^{(i)}$
1	3	1	1	1	0	2	2	2
2	3	1	0	2	0	2	3	1
3	1	3	2	0	2	0	1	3
4	3	0	2	1	0	3	1	2
5	3	0	2	1	0	3	1	2
6	1	2	2	1	2	1	1	2
7	1	0	3	2	2	3	0	1
8	2	1	3	0	1	2	0	3
9	3	0	2	1	0	3	1	2
10	2	2	2	0	1	1	1	3
11	2	3	0	1	1	0	3	2
12	2	1	2	1	1	2	1	2
13	3	1	2	0	0	2	1	3
14	1	1	3	1	2	2	0	2
15	1	2	1	2	2	1	2	1

$F = n(n-1)/2 = 4(3)/2 = 6$ $\qquad$ $c = N(n-1) = 15(3) = 45$

$$C = Nn(n-1) = 15(4)(3) = 180$$

Table 2

Person	O_1 O_2	O_1 O_3	O_1 O_4	O_2 O_3	O_2 O_4	O_3 O_4
1	>	>	>	>	<	>
2	>	>	>	>	<	<
3	<	<	>	>	>	>
4	>	>	>	<	<	>
5	>	>	>	<	<	>
6	>	<	<	>	>	>
7	>	<	<	<	<	>
8	>	<	>	<	>	>
9	>	>	>	<	<	>
10	>	>	>	<	>	>
11	<	>	>	>	>	<
12	>	>	<	<	>	>
13	>	>	>	<	>	>
14	<	<	>	<	<	>
15	<	>	<	<	>	<

According to the equations (2.20), (2.21), (2.22), we have

$$\underset{\sim}{H} = \begin{bmatrix} 0.37 & 0.18 & 0.24 & 0.21 \\ 0.18 & 0.37 & 0.2 & 0.25 \\ 0.24 & 0.2 & 0.35 & 0.21 \\ 0.21 & 0.25 & 0.21 & 0.33 \end{bmatrix}$$

Now we calculate the characteristic root η^2 and characteristic vector $\underset{\sim}{x}$. According to Notes 1 and 2, our solution is the characteristic vector corresponding to the largest non-trivial characteristic roots ($\eta^2 \neq 1$) of $|\underset{\sim}{H} - \eta^2{}_4| = 0$ where $\underset{\sim}{H}$ has been given above. The result is $\eta^2 = 0.210368$ and the corresponding vector is

$$\underset{\sim}{x} = \begin{bmatrix} \frac{-0.6}{-0.5} \\ \frac{0.66}{-0.5} \\ \frac{-0.35}{-0.5} \\ \frac{0.29}{-0.5} \end{bmatrix} = \begin{bmatrix} 1.2 \\ -1.32 \\ .7 \\ -.58 \end{bmatrix}$$

We then find that x_j has been weighted as in this order

$$x_1 > x_3 > x_4 > x_2$$

2.3 Comparing combinations of two things

Consider a set of n things or items, the jth of which has m_j categories. Let O_{jx} $(k=1, \cdots, m_j;\ j=1, \cdots, n)$ be the αth category of the jth item. Each of N individuals is asked to make judgments of the form that the combination $(O_{j\alpha}, O_{k\beta})$ is greater than (or less than) the combination $(O_{j\gamma}, O_{k\delta})$. Here the jth and the kth are combined. As in the case of ordinary comparisons, we assume that all people compare each of the pairs of combinations and that the judgments of equality are excluded.

Let

$$\ell^{(i)}_{jk|\alpha\beta,\gamma\delta} = \begin{cases} 1, & \text{if the individual i judges } (O_{j\alpha}, O_{k\beta}) \quad (O_{j\gamma}, O_{k\delta}) \\ 0, & \text{otherwise.} \end{cases} \tag{2.24}$$

Definition (2.24) implies that

$$\ell^{(i)}_{jk|\alpha\beta,\gamma\delta} = \ell^{(i)}_{kj|\beta\alpha,\delta\gamma} \qquad \text{(symmetry)} \tag{2.25}$$

and that

$$\ell^{(i)}_{jk|\alpha\beta,\gamma\delta} + \ell^{(i)}_{jk|\gamma\delta,\alpha\beta} = \begin{cases} 0, & \text{if individual i omits the comparison of } (O_{j\alpha}, O_{k\beta}) \text{ with } (O_{j\gamma}, O_{k\delta}) \\ 1, & \text{if he judges these two combinations to be unequal.} \end{cases} \tag{2.26}$$

The following notations and definitions are used:

$$\left.\begin{aligned} a^{(i)}_{jk|\alpha\beta} &= a^{(i)}_{kj|\beta\alpha} = \sum_{\gamma\delta}\sum e^{(i)}_{jk|\alpha\beta,\gamma\delta} \\ &= \text{the number of combinations individual } i \text{ judged to be lower than } (O_{j\alpha}, O_{k\beta}) \\ b^{(i)}_{jk|\alpha\beta} &= b^{(i)}_{kj|\beta\alpha} = \sum_{\gamma\delta}\sum e^{(i)}_{jk|\gamma\delta,\alpha\beta} \\ &= \text{the number of combinations individual } i \text{ judged to be higher than } (O_{j\alpha}, O_{k\beta}). \end{aligned}\right\} \quad (2.25)$$

$$c_{jk|\alpha\beta} = \sum_i (a^{(i)}_{jk|\alpha\beta} + b^{(i)}_{jk|\alpha\beta}) = c_{kj|\beta\alpha} \quad (2.26)$$

= the number of comparisons for all individuals involving $(O_{j\alpha}, O_{k\beta})$.

$$f^{(i)}_{j\alpha} = \sum_{k\beta}\sum a^{(i)}_{jk|\alpha\beta}, \qquad g^{(i)}_{j\alpha} = \sum_{k\beta}\sum b^{(i)}_{jk|\alpha\beta} \quad (2.27)$$

$$C_{j\alpha} = \sum_{k\beta}\sum c_{jk|\alpha\beta} = \sum_i (f^{(i)}_{j\alpha} + g^{(i)}_{j\alpha}) \quad (2.28)$$

= the total number of times in the entire experiment that $O_{j\alpha}$ was involved.

$$F = \sum_{j\alpha}\sum f^{(i)}_{j\alpha} = \sum_{j\alpha}\sum g^{(i)}_{j\alpha} \quad (2.29)$$

= the total number of comparisons made by each person.

$$C = \sum_{j\alpha}\sum C_{j\alpha} = 2NF. \quad (2.30)$$

= total number of judgments in the whole experiment.

By using these notations and definitions, we now consider the determination of the x_{jp}-values to be given to O_{jp} from the judgments. To do so, we obtain, as in the case of ordinary comparisons, the sum of squares B <u>between</u> individuals for the experiment and the sum of squares W <u>within</u> individuals. First of all, let

$$t^{(i)} \equiv \frac{1}{F} \sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta}) a^{(i)}_{jk|\alpha\beta}$$

$$= \frac{2}{F} \sum_{k\beta}\sum x_{k\beta} f^{(i)}_{k\beta} \tag{2.31}$$

= the mean of the x-values of the combinations individual i judged to be higher than other combinations,

$$u^{(i)} \equiv \frac{1}{F} \sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta}) b^{(i)}_{jk|\alpha\beta}$$

$$= \frac{2}{F} \sum_{k\beta}\sum x_{k\beta} g^{(i)}_{k\beta} \tag{2.32}$$

= the mean of the x-values of the combinations individual i judged to be lower than other combinations,

$$y^{(i)} \equiv \sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta} - t^{(i)})^2 a^{(i)}_{jk|\alpha\beta}$$

$$= \sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta})^2 a^{(i)}_{jk|\alpha\beta} - t^{(i)^2} F \tag{2.33}$$

$$z^{(i)} \equiv \sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta} - u^{(i)})^2 b^{(i)}_{jk|\alpha\beta}$$

$$= \sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta})^2 b^{(i)}_{jk|\alpha\beta} - u^{(i)^2} F \tag{2.34}$$

Then the total sum of squares T for the experiment can now be expressed as

$$T = \sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta} - V)^2 c_{hj|\alpha\beta}$$

$$= \sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta})^2 c_{jk|\alpha\beta} - V^2 C \tag{2.35}$$

where

$$V = \frac{1}{C}\sum_{jk\alpha\beta}\sum\sum\sum (x_{j\alpha} + x_{k\beta}) c_{jk|\alpha\beta} \tag{2.36}$$

$$= \frac{2}{C} \sum_{k\beta}\sum x_{k\beta} C_{k\beta}$$

is the grand mean of all x-values in the entire experiment, and B, W as

$$B = \sum_i [(t^{(i)} - V)^2 + (u^{(i)} - V)^2] F$$

$$= F \sum_i ({t^{(i)}}^2 + {u^{(i)}}^2) - V^2 C, \tag{2.37}$$

$$W = \sum_i (y^{(i)} + z^{(i)}) = T - B. \tag{2.38}$$

We again use the square of correlation ratio η

$$\eta^2 = \frac{B}{T} = 1 - \frac{W}{T} \tag{2.39}$$

as our criterion for determining the $x_{j\alpha}$; that is, we wish to determine the $x_{j\alpha}$ that will maximize η^2.

Exactly the same as in the previous case, we can put $V = 0$ without any loss of generality, so that

$$B = F\sum[t^{(i)^2}+u^{(i)^2}], \qquad T = \sum\sum\sum\sum(x_{j\alpha}+x_{k\beta})^2 c_{jk|\alpha\beta} \tag{2.40}$$

2.3.1 The unrestricted case

In this case, the computation of the x-values maximizing η^2 defined by (2.39) is carried out in the exact same way as in the case of ordinary comparisons, once we get the B, W, and T. The stationary equations are

$$\frac{\partial B}{\partial x_{j\alpha}} = \eta^2 \frac{\partial T}{\partial x_{j\alpha}}, \qquad \alpha=1, \cdots, m_j;\ j=1, \cdots, n \tag{2.41}$$

where

$$\frac{\partial B}{\partial x_{j\alpha}} = \frac{8}{F} \sum_k\sum x_{k\beta}\sum_i(f_j^{(i)} f_k^{(i)}+g_j^{(i)} g_k^{(i)}) \tag{2.42}$$

$$\frac{\partial T}{\partial x_{j\alpha}} = 4[x_{j\alpha}C_{j\alpha}+\sum_{k\beta}\sum x_{k\beta}c_{jk|\alpha\beta}]. \tag{2.43}$$

If we let

$$h_{jk|\alpha\beta} = \frac{1}{F}\sum_i(f_{j\alpha}^{(i)} f_{k\beta}^{(i)}+g_{j\alpha}^{(i)} g_{k\beta}^{(i)}), \tag{2.44}$$

then the simultaneous equations to be solved is

$$\sum_{k\beta}\sum x_{k\beta}h_{jk|\alpha\beta} = \frac{1}{2}\eta^2\{x_{j\alpha}C_{j\alpha}+\sum_{k\beta}\sum x_{k\beta}c_{jk|\alpha\beta}\} \tag{2.45}$$

for $\alpha=1, \cdots, m_j;\ j=1, \cdots, n$.

It is shown that as in the case of ordinary comparisons, the condition V = 0 will in general be satisfied by a solution of (2.45).

The system of the simultaneous equations can be expressed in matric form in the following way:

$$\underset{\sim}{x} = (x_{11}, \cdots, x_{1m_1};\ x_{21}, \cdots, x_{2m_2};\ \cdots,\ x_{n1}, \cdots, x_{nm_n})$$

$$\underset{\sim}{K} = \begin{array}{cc} \begin{pmatrix} \underset{\sim}{H}_{11} & \underset{\sim}{H}_{12} & \cdots & \underset{\sim}{H}_{1n} \\ \underset{\sim}{H}_{21} & \underset{\sim}{H}_{22} & \cdots & \underset{\sim}{H}_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \underset{\sim}{H}_{n1} & \underset{\sim}{H}_{n2} & \cdots & \underset{\sim}{H}_{nn} \end{pmatrix} & \begin{matrix} m_1 \\ m_2 \\ \cdots \\ \cdots \\ m_n \end{matrix} \\ \begin{matrix} m_1 & \quad m_2 & \quad \cdots & \quad m_n \end{matrix} & \end{array}$$

where $\underset{\sim}{H}_{kj}$ are $m_k \times m_j$ submatrices;

$$\underset{\sim}{H}_{kj} = \begin{pmatrix} h_{jk|11} & h_{jk|21} & \cdots & h_{jk|m_j1} \\ h_{jk|12} & h_{jk|22} & \cdots & h_{jk|m_j2} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ h_{jk|1m_k} & h_{jk|2m_k} & \cdots & h_{jk|m_jm_k} \end{pmatrix},$$

$$
\underset{\sim}{D} = \begin{pmatrix}
c_{11} & & & & & & & \\
& \ddots & & & & & 0 & \\
& & c_{1m_1} & & & & & \\
& & & c_{21} & & & & \\
& & & & \ddots & & & \\
& & & & & c_{2m_2} & & \\
& & & & & & \ddots & \\
& & & & & & & c_{n1} \\
& 0 & & & & & & & \ddots \\
& & & & & & & & & c_{nm_n}
\end{pmatrix}
$$

$$
\underset{\sim}{F}_{kj} = \begin{pmatrix}
c_{jk|11} & c_{jk|21} & \cdots & c_{jk|m_j1} \\
c_{jk|12} & c_{jk|22} & \cdots & c_{jk|m_j2} \\
\cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
c_{jk|m_k} & c_{jk|2m_k} & \cdots & c_{jk|m_jm_k}
\end{pmatrix}
$$

$$
\underset{\sim}{G} = \begin{array}{c c}
\begin{pmatrix}
\underset{\sim}{F}_{11} & \underset{\sim}{F}_{12} & \cdots & \underset{\sim}{F}_{1n} \\
\underset{\sim}{F}_{21} & \underset{\sim}{F}_{22} & \cdots & \underset{\sim}{F}_{2n} \\
\cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
\underset{\sim}{F}_{n1} & \underset{\sim}{F}_{n2} & \cdots & \underset{\sim}{F}_{nn}
\end{pmatrix} & \begin{array}{c} m_1 \\ m_2 \\ \cdot\cdot \\ \cdot\cdot \\ m_n \end{array} \\
\begin{array}{cccc} m_k & m_2 & \cdots & m_n \end{array} &
\end{array}
$$

It is noted here that K and G are both symmetric matrices of order $m=m_1+\cdots M_n$. Then (2.45) can now be written as

$$\underset{\sim}{x}\underset{\sim}{K} = \lambda(\underset{\sim}{x}\underset{\sim}{D}+\underset{\sim}{x}\underset{\sim}{G})$$

$$= \lambda\underset{\sim}{x}(\underset{\sim}{D}+\underset{\sim}{G}) \tag{2.46}$$

where $\lambda=\frac{1}{2}\eta^2$. Since $(\underset{\sim}{D}+\underset{\sim}{G})$ is generally non-singular, (2.46) becomes

$$\underset{\sim}{x}\underset{\sim}{K}(\underset{\sim}{D}+\underset{\sim}{G})^{-1} = \lambda\underset{\sim}{x}. \tag{2.47}$$

This shows that λ is a characteristic root of $\underset{\sim}{K}(\underset{\sim}{D}+\underset{\sim}{G})^{-1}$ and $\underset{\sim}{x}$ is the characteristic vector corresponding to λ. Since we want the largest possible correlation ratio, the desired numerical solution $\underset{\sim}{x}_m$ can be obtained by computing the characteristic vector corresponding to the largest root λ_m of the matrix $\underset{\sim}{K}(\underset{\sim}{D}+\underset{\sim}{G})^{-1}$.

2.3.2 The restricted case

For some problems, the $O_{j\alpha}$ may be quantitative and it may be desired within each item <u>to keep the distances between x_{jp} proportionate to the distances between the $O_{j\alpha}$</u>. This was the case for the score card, where a linear system of weighting had to be used to be practicable for the army. It was necessary to derive a constant multiplier for length of service, a constant multiplier for time overseas, etc., even though there were curvilinearities in the judgments.

Thus we set the x-values in the form

$$x_{j\alpha} = \xi_j + \alpha\zeta_j,\ \alpha=1, \cdots, m_j;\ j=1, \cdots, n; \tag{2.48}$$

hence the ξ_j and ζ_j are now the basic unknowns to be solved for maximizing the correlation ratio η. It is noted that

$$x_{j\alpha} - x_{j\beta} = (\alpha-\beta)\zeta_j \tag{2.49}$$

which is equivalent to the above statement that $(O_{j\alpha} - O_{j\beta})$ is proportional to $(\alpha-\beta)$ within the jth item.

Under these linear restrictions, the stationary equations for maximizing η^2 are now obtained by

$$\frac{\partial B}{\partial \xi_j} = \eta^2 \frac{\partial T}{\partial \xi_j},\quad \frac{\partial B}{\partial \zeta_j} = \eta^2 \frac{\partial T}{\partial \zeta_j}. \tag{2.50}$$

Let us introduce the following notations:

$$P_{0,jk} \equiv \frac{1}{F}\sum_i [(\sum_\alpha f_{j\alpha}^{(i)})(\sum_\beta f_{k\beta}^{(i)}) + (\sum_\alpha g_{j\alpha}^{(i)})(\sum_\beta g_{k\beta}^{(i)})], \tag{2.51}$$

$$P_{1,jk} \equiv \frac{1}{F}\sum_i [(\sum_\alpha \alpha f_{j\alpha}^{(i)})(\sum_\beta f_{k\beta}^{(i)}) + (\sum_\alpha \alpha g_{j\alpha}^{(i)})(\sum_\beta g_{k\beta}^{(i)})], \tag{2.52}$$

$$P_{2,jk} \equiv \frac{1}{F}\sum_i [(\sum_\alpha \alpha f_{j\alpha}^{(i)})(\sum_\beta \beta f_{k\beta}^{(i)}) + (\sum_\alpha \alpha g_{j\alpha}^{(i)})(\sum_\beta \beta g_{k\beta}^{(i)})], \tag{2.53}$$

$$d_{r,jk} \equiv \sum_{\alpha\beta}\sum \alpha^r c_{jk|\alpha\beta},\quad d_{11,jk} \equiv \sum_{\alpha\beta}\sum \alpha\beta c_{jk|\alpha\beta} \tag{2.54}$$

Expressing B and T in terms of ξ_j and ζ_j and differentiating them with respect to ξ_j and ζ_j, the stationary equations in (2.50) can now be written as

$$\sum_k(\xi_k p_{0,jk}+\zeta_k p_{1,kj}) = \frac{1}{2}\eta^2\sum_k[(\xi_j d_{0,jk}+\zeta_j d_{1,kj})$$

$$+ (\xi_k d_{0,jk}+\zeta_k d_{1,jk})] \tag{2.55}$$

$$\sum_k(\xi_k p_{1,jk}+\zeta_k p'_{2,jk}) = \frac{1}{2}\eta^2\sum_k[(\xi_j d_{1,jk}+\zeta_j d_{2,jk})$$

$$+ (\xi_k d_{1,jk}+\zeta_k d_{11,jk})]. \tag{2.56}$$

As in the previous case, we can show that a solution of (2.55) and (2.1.56) will satisfy V = 0. First of all, V of (2.36) is expressed as follows:

$$V=\frac{2}{C}\sum_k[\xi_k(\sum_j d_{1,jk})+\zeta_k(\sum_j d_{1,jk})]. \tag{2.57}$$

Summing the both sides of (2.55) with respect to j shows that

$$(1-\eta^2)\sum_k[\xi(\sum_j d_{0,jk})+\zeta_k(\sum_j d_{1,jk})] = 0$$

so that

$$(1-\eta^2)V = 0$$

Thus if $\eta^2 \neq 1$, V = 0.

We shall finally express the system of the stationary equations (2.55) and (2.56) in matric form. Let

$$\underset{\sim}{y} = (\xi_1,\ \xi_2,\ \cdots,\ \xi_n : \zeta_1, \zeta_2,\ \cdots,\ \zeta_n), \tag{2.58}$$

$$\underset{2n\times 2n}{P} = \begin{bmatrix} p_{0,11} & p_{0,12} & \cdots & p_{0,1n} & p_{1,11} & p_{1,21} & \cdots & p_{1,n1} \\ p_{0,21} & p_{0,22} & \cdots & p_{0,2n} & p_{1,12} & p_{1,22} & \cdots & p_{1,n2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{0,n1} & p_{0,n2} & \cdots & p_{0,nn} & p_{1,1n} & p_{1,2n} & \cdots & p_{1,nn} \\ p_{1,11} & p_{1,12} & \cdots & p_{1,1n} & p_{2,11} & p_{2,12} & \cdots & p_{2,1n} \\ p_{1,21} & p_{1,22} & \cdots & p_{1,2n} & p_{2,21} & p_{2,22} & \cdots & p_{2,2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{1,n1} & p_{1,n2} & \cdots & p_{1,nn} & p_{2,n1} & p_{2,n2} & \cdots & p_{2,nn} \end{bmatrix} \tag{2.59}$$

(It must be noted here that $p_{0,jk} = p_{0,kh}$, $p_{2,jk} = p_{2,kj}$, while $p_{1,jk} \neq p_{1,kj}$.)

$$\sum_k d_{m,jk} = D_{m,j}, \quad m = 0,1,2 \tag{2.60}$$

$$\underset{2n\times 2n}{\underset{\sim}{Q}} = \begin{pmatrix} D_{0,1}+d_{0,11} & d_{0,12} & \cdots & d_{0,1n} & D_{1,1}+d_{1,11} & d_{1,21} & \cdots & d_{1,n1} \\ d_{0,21} & D_{0,2}+d_{0,22} & \cdots & d_{0,2n} & d_{1,12} & D_{1,2}+d_{1,22} & \cdots & d_{1,n2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ d_{0,n1} & d_{0,n2} & \cdots & D_{0,n}+d_{0,nn} & d_{1,1n} & d_{1,2n} & \cdots & D_{1,n}+d_{1,nn} \\ D_{1,1}+d_{1,11} & d_{1,12} & \cdots & d_{1,1n} & D_{2,1}+d_{11,11} & d_{11,12} & \cdots & d_{11,1n} \\ d_{1,21} & D_{1,2}+d_{1,22} & \cdots & d_{1,2n} & d_{11,21} & D_{2,2}+d_{11,22} & \cdots & d_{11,2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ d_{1,n1} & d_{1,n2} & \cdots & D_{1,n}+d_{1,nn} & d_{11,n1} & d_{11,n2} & \cdots & D_{2,n}+d_{11,nn} \end{pmatrix} \tag{2.61}$$

Then our stationary equations can be written as

$$\underset{\sim}{y}\underset{\sim}{P} = \lambda \underset{\sim}{y}\underset{\sim}{Q} \tag{2.62}$$

or

$$\underset{\sim}{y}\underset{\sim}{P}\underset{\sim}{Q}^{-1} = \lambda \underset{\sim}{y}, \qquad (\lambda = \tfrac{1}{2}\eta^2) \tag{2.63}$$

since $\underset{\sim}{Q}$ is in general non-singular. Thus λ is a characteristic root of $\underset{\sim}{P}\underset{\sim}{Q}^{-1}$ and $\underset{\sim}{y}$ is a characteristic vector corresponding to a λ. Since we want the largest correlation ratio, our desired vector y_m is obtained as the characteristic vector corresponding to the largest root of λ_m of $\underset{\sim}{P}\underset{\sim}{Q}^{-1}$.

CHAPTER III

QUANTIFICATION OF QUALITATIVE DATA WHEN AN OUTSIDE CRITERION IS GIVEN

3.1 The case where the outside variable is numerical: prediction of an outside variable from a response pattern.

We draw a random sample of size N from a population. Suppose that each person is asked to respond to the questionnaires in the following manner: the questionnaires consist of M items, $I_1, I_2, \cdots, I_M$; each I_j has the k_j subcategories $C_{j1}, C_{j2}, \cdots, C_{jk}$, $(j=1, \cdots, M)$. Each person is asked to check in only one subcategory for each item which he thinks to be most appropriate as his response. Suppose that a numerical value is given to each person as an outside criterion from another survey. Thus the response patterns with numerical values of an outside variable Y of N persons are given, for example, as in the table in the next page.

The problem considered here is <u>to predict the outside variable from a known response pattern of a person</u>, and to establish an appropriate formula for quantifying the response pattern to do so.

Let X_j^* $(j=1, \cdots, m)$ be the random variable representing the j-th item and let

$$P(X_j^*=C_{j\alpha}) \equiv p_{j\alpha}, \quad P(X_j^*=c_{j\alpha}, X_k^*=c_{k\beta}) = p_{jk|\alpha\beta} \quad (3.1)$$

TABLE 3.1 The Response Patterns With the Numerical Values of the Outside Variable

Items	I_1				I_2				$\cdots$	I_M				
Variables	X_1				X_2				$\cdots$	X_M				Outside
Subcategories	c_{11}	c_{12}	$\cdots$	c_{1k_1}	c_{21}	c_{22}	$\cdots$	c_{2k_2}	$\cdots$	c_{M1}	c_{M2}	$\cdots$	c_{Mk_M}	Variable
Values giving	x_{11}	x_{12}	$\cdots$	x_{1k_1}	x_{21}	x_{22}	$\cdots$	x_{2k_2}	$\cdots$	x_{M1}	x_{M2}	$\cdots$	x_{Mk_M}	Y
Persons														
1		V	$\cdots$				$\cdots$	V	$\cdots$	V		$\cdots$		y_1
2	V		$\cdots$			V	$\cdots$		$\cdots$			$\cdots$	V	y_2
3			$\cdots$	V	V		$\cdots$		$\cdots$		V	$\cdots$		y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i		V	$\cdots$			V	$\cdots$		$\cdots$				V	y_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	V		$\cdots$				$\cdots$	V	$\cdots$	V		$\cdots$		y_N
No. of Responses	n_{11}	n_{12}	$\cdots$	n_{1k_1}	n_{21}	n_{22}	$\cdots$	n_{2k_2}	$\cdots$	n_{M1}	n_{M2}	$\cdots$	n_{Mk_M}	

V is the sign of response.

where $\sum_{\alpha} p_{j\alpha}=1$ for all j, $\sum_{\alpha\beta}\sum p_{jk|\alpha\beta}=1$ for all $j\neq k$, and $p_{jj|\alpha\beta}=0$ for all $\alpha\neq\beta$, $p_{jj|\alpha\alpha}= p_{j\alpha}$.

Now let $x_{j\alpha}$, d=1, $\cdots$, k_j; j=1, $\cdots$, M be the numerical values to be given to $c_{j\alpha}$ so as to predict the outside variable from a response pattern as efficiently as possible. Let X_j (j=1, $\cdots$, m) be the corresponding random variable to X_j^* taking values over $(x_{j1}, \cdots, x_{jk_j})$. Then

$$P(X_j=x_{j\alpha}) = P(X_j^*=c_{j\alpha}) = P_{j\alpha}$$

$$P(X_j=x_{j\alpha}, X_k=x_{k\beta}) = P(X_j^*=c_{j\alpha}, X_k^*= c_{k\beta}) = p_{jk|\alpha\beta} \quad (3.2)$$

We define the score to be given to a person by

$$S = X_1 + X_2 + \cdots + X_M \quad (3.3)$$

as the first approximation.

Note 1. (3.3) can be interpreted in another way; that is, let

$$Z_{j\alpha} = \begin{cases} 1, \text{ if a person responds to } c_{j\alpha} \\ 0, \text{ otherwise.} \end{cases} \quad (3.4)$$

Then the score S can be expressed as

$$S = x_{11}Z_{11}+ \cdots + x_{1k_1}Z_{1k_1} + x_{21}Z_{21}+ \cdots + x_{2k_2}Z_{2k_2}$$

$$+ \cdots + x_{M1}Z_{M1}+ \cdots + x_{Mk_M}Z_{Mk_M}. \quad (3.5)$$

and here $P(Z_{j\alpha}= 1) = p_{j\alpha}$. Hence $x_{j\alpha}$ are weight attached to the $Z_{j\alpha}$ or subcategories $c_{j\alpha}$.

We wish then to quantify the response pattern or to determine the values of $x_{j\alpha}$ so that the correlation coefficient P_{YS} between Y and S is maximum or equivalently the mean square error of the prediction $E(Y-S)^2$ is minimum. By using the expression (3.5), P_{YS} is interpreted as the multiple correlation coefficient between Y and $(Z_{11}, \cdots, Z_{1k_1}, \cdots, Z_{M1}, \cdots, Z_{Mk_M})$ and $E(Y-S)^2$ is the residual variance after we determined the optimum values of $x_{j\alpha}$'s.

$$E(Y-S)^2 = E(Y^2)-2E(YS) + E(S^2)$$
$$= E(Y^2)-2\sum_{j=1}^{M} E(YX_j)+\sum_{j=1}^{M} E(X_j^2)+\sum_{j\neq k} E(X_jX_k) \qquad (3.6)$$

Each term in the right side of (3.6) can be expressed in terms of $x_{j\alpha}$ as follows

$$E(YX_j) = \sum_{\alpha=1}^{k_j} \{E(YX_j|X_j=x_{j\alpha})\}\cdot P(X_j=x_{j\alpha})$$
$$= \sum_{\alpha=1}^{k_j} x_{j\alpha}E(Y|X_j=x_{j\alpha})\cdot p_{j\alpha}$$
$$= \sum_{\alpha=1}^{k_j} x_{j\alpha}\mu_{j\alpha}p_{j\alpha} \qquad (j=1, \cdots, M) \qquad (3.7)$$

where $E(Y|X_j=x_j) \equiv \mu_{j\alpha}$,

$$E(X_j^2) = \sum_{\alpha=1}^{k_j} x_{j\alpha}^2 P(X_j=x_{j\alpha}) = \sum_{\alpha=1}^{k_j} x_{j\alpha}^2 p_{j\alpha}, \quad (j=1, \cdots, M) \qquad (3.8)$$

$$E(X_jX_k) = \sum_{\alpha=1}^{k_j} \sum_{\beta=1}^{k_k} x_{j\alpha}x_{k\beta}p_{jk|\alpha\beta}. \qquad (j\neq k) \qquad (3.9)$$

Hence

$$E(Y-S)^2 = E(Y^2) - 2\sum_{j=1}^{M}\sum_{\alpha=1}^{k_j} x_{j\alpha}\mu_{j\alpha}p_{j\alpha} + \sum_{j=1}^{M}\sum_{\alpha=1}^{k_j} x_{j\alpha}^2 p_{j\alpha}$$

$$+ \sum_{\substack{j=1 \\ (j\neq k)}}^{M}\sum_{k=1}^{M}\sum_{\alpha=1}^{k_j}\sum_{\beta=1}^{k_k} x_{j\alpha}x_{k\beta}p_{jk|\alpha\beta}. \quad (3.10)$$

Now then we differentiate (3.10) with respect to $x_{j\alpha}$ and equate the resultant to zero, which gives us

$$p_{j\alpha}x_{j\alpha} + \sum_{\substack{k \\ (\neq j)}}\sum_{\beta} p_{jk|\alpha\beta}x_{k\beta} = p_{j\alpha}\mu_{j\alpha} \quad (3.11)$$

$$\alpha=1, \cdots, k_j, \ j=1, \cdots, M$$

If we know $\mu_{j\alpha}$, $p_{j\alpha}$ and $p_{jk|\alpha\beta}$, then the desired values of $x_{j\alpha}$'s are obtained by solving (3.11). Since this is not the case in a practical situation, we need to use their estimates based on the sample. Before this, we calculate the minimum mean square errors: that is, denoting the solution of (3.11) by $\hat{x}_{j\alpha}$, we obtain

$$E(Y\hat{S}) = \sum_j E(Y\hat{X}_j) = \sum_j\sum_\alpha \hat{x}_{j\alpha}\mu_{j\alpha}p_{j\alpha}$$

$$= \sum_j\sum_\alpha \hat{x}_{j\alpha}^2[p_{j\alpha}\hat{x}_{j\alpha} + \sum_{\substack{k \\ (\neq j)}}\sum_\beta \hat{x}_{k\beta}p_{jk|\alpha\beta}] \quad [\because (3.12)]$$

$$= \sum_j\sum_\alpha \hat{x}_{j\alpha}^2 p_{j\alpha} + \sum_{j\neq k}\sum \sum_{\alpha}\sum_{\beta} \hat{x}_{j\alpha}\hat{x}_{k\beta}p_{jk\Delta\alpha\beta}$$

$$= \sum_j E(\hat{X}_j^2) + \sum_{j\neq k}\sum E(\hat{X}_j\hat{X}_k)$$

$$= E(\hat{S}^2) \quad (3.13)$$

where $\hat{S}$, $\hat{X}_j$ are random variables when $\hat{x}_{j\alpha}$ are used. Thus we have

$$\text{Min}\ \{E(Y-S)^2\} = E(Y^2)-E(\hat{S}^2)$$

$$= E(Y^2)-E(Y\hat{S}) \qquad (3.14)$$

From this, it is seen that $E(YS) \geq 0$, and $E(YS)$ has been maximized by minimizing $E(Y-S)^2$.

To obtain the estimates of $\mu_{j\alpha}$, $p_{j\alpha}$, and $p_{jk|\alpha\beta}$, let

$n_{j\alpha}$ = the number of persons (or marks "v") responding to $c_{j\alpha}$ among N persons, so that $\sum_\alpha n_{j\alpha} = N$ for all j.

$n_{jk|\alpha\beta}$ = the number of persons (or marks "i") responding simultaneously to $c_{j\alpha}$ and $c_{k\beta}$ among N persons, so that

$n_{jk|\alpha\beta} = n_{kj|\beta\alpha}$,

$\sum_{\alpha\beta}\sum n_{jk|\alpha\beta} = N$ for all j and k, $\sum_\beta n_{jk|\alpha\beta} = n_{j\alpha}$ for all j and α

$\sum_\alpha n_{jk|\alpha\beta} = n_{k\beta}$ for all k and β,

$t_{j\alpha}$ = the sum of Y-values of persons (or for the mark "V") responding to $c_{j\alpha}$.

Then the parameters or population quantities involved in the equations (3.11) can be estimated by

$$\hat{p}_{j\alpha} = n_{j\alpha}/N,$$

$$\hat{p}_{jk|\alpha\beta} = n_{jk|\alpha\beta}/N,$$

$$\hat{\mu}_{j\alpha} = t_{j\alpha}/n_{j\alpha}. \qquad (3.15)$$

Using these estimates, the simultaneous equations of $x_{j\alpha}$ are now replaced by

$$n_{j\alpha}x_{j\alpha} + \sum_{k(\neq j)} \sum_{\beta} n_{jk|\alpha\beta}x_{k\beta} = t_{j\alpha} \tag{3.16}$$

$$\alpha=1, \cdots, k_j; \; j=1, \cdots, M$$

In the actual computation, they are the equations to be solved numerically.

To have a matric form, let

$$\underset{\sim}{x} = \begin{pmatrix} x_{11} \\ \cdot \\ \cdot \\ \cdot \\ x_{1k_1} \\ x_{21} \\ \cdot \\ \cdot \\ \cdot \\ x_{2k_2} \\ \cdot \\ \cdot \\ \cdot \\ x_{M1} \\ \cdot \\ \cdot \\ \cdot \\ x_{Mk_M} \end{pmatrix} \qquad \underset{\sim}{t} = \begin{pmatrix} t_{11} \\ \cdot \\ \cdot \\ \cdot \\ t_{1k_1} \\ t_{21} \\ \cdot \\ \cdot \\ \cdot \\ t_{2k_2} \\ \cdot \\ \cdot \\ \cdot \\ t_{M1} \\ \cdot \\ \cdot \\ \cdot \\ t_{Mk_M} \end{pmatrix},$$

$$\underset{\sim}{A} = \left(\begin{array}{c|c|c|c}
\begin{matrix} n_{11} & & 0 \\ & \ddots & \\ 0 & & n_{1k_1} \end{matrix} &
\begin{matrix} n_{12/11} & \cdots & n_{12/1k_2} \\ \vdots & & \vdots \\ n_{12/k_1 1} & \cdots & n_{12/k_1 k_2} \end{matrix} &
\cdots &
\begin{matrix} n_{1M/11} & \cdots & n_{1M/1k_M} \\ \vdots & & \vdots \\ n_{1M/k_1 1} & \cdots & n_{1M/k_1 k_M} \end{matrix} \\ \hline
\begin{matrix} n_{21/11} & \cdots & n_{21/1k_1} \\ \vdots & & \vdots \\ n_{21/k_2 1} & \cdots & n_{21/k_2 k_1} \end{matrix} &
\begin{matrix} n_{21} & & 0 \\ & \ddots & \\ 0 & & n_{2k_2} \end{matrix} &
\cdots &
\begin{matrix} n_{2M/11} & \cdots & n_{2M/1k_M} \\ \vdots & & \vdots \\ n_{2M/k_2 1} & & n_{2M/k_2 k_M} \end{matrix} \\ \hline
\vdots & \vdots & \cdots & \vdots \\ \hline
\begin{matrix} n_{M1/11} & \cdots & n_{M1/1k_1} \\ \vdots & & \vdots \\ n_{M1/k_M 1} & \cdots & n_{M1/k_M k_1} \end{matrix} &
\begin{matrix} n_{M2/11} & \cdots & n_{M2/1k_2} \\ \vdots & & \vdots \\ n_{M2/k_M 1} & \cdots & n_{M2/k_M k_2} \end{matrix} &
\cdots &
\begin{matrix} n_{M1} & & 0 \\ & \ddots & \\ 0 & & n_{Mk_M} \end{matrix}
\end{array}\right)$$

Then (3.16) can be rewritten as

$$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{t} \tag{3.17}$$

Since

$$\sum_{\alpha} n_{j\alpha} + \sum_{k(\neq j)} \sum_{\alpha\beta}\sum n_{jk|\alpha\beta} = MN \text{ for all } j=1, \cdots, M, \tag{3.18}$$

then $\underset{\sim}{A}$ is singular and has the rank equal to

$$(k_1 + k_2 + \cdots + k_M) - (M - 1).$$

Hence to solve the equation (3.17), it is convenient to reduce (3.17) to a non-singular equation by putting, for example,

$$x_{1k_1} = 0,\ x_{2k_2} = 0,\ \cdots,\ x_{M-1,\ k_{M-1}} = 0 \tag{3.19}$$

and deleting the k_1-st, k_2-nd, $\cdots$, k_{M-1}th columns and rows from A and $t_{k_1}, t_{k_2}, \cdots, t_{k_{M-1}}$ from t. Let $\underset{\sim}{x}^*$, $\underset{\sim}{t}^*$ and A* be the resultant shrinked vectors and matrix; then we have

$$\underset{\sim}{A}^*\underset{\sim}{x}^* = \underset{\sim}{t}^* \tag{3.20}$$

and hence

$$\underset{\sim}{x}^* = \underset{\sim}{A}^{*-1}\underset{\sim}{t}^* \tag{3.21}$$

where

$$\underset{\sim}{x}^* = (x_{11}, \cdots, x_{1,k_1-1};\ x_{21}, \cdots, x_{2,k_2-1};\ \cdots; x_{M-1,1}$$

$$\cdots, x_{M-1,k_{M-1}-1}; x_{M1}, \cdots, x_{M,k_M}).$$

Thus we obtain as our solution ("^" is attached)

$$x_j\colon\ (\hat{x}_{j1}, \hat{x}_{j2}, \cdots, \hat{x}_{j,k_j-1}, 0),\quad j=1, \cdots, M-1$$

$$X_M\colon\ (\hat{x}_{M1}, \hat{x}_{M2}, \cdots, \hat{x}_{Mk_M}). \tag{3.22}$$

Since we have assumed that $Y = S + \varepsilon = X_1 + \cdots + X_M + \varepsilon$ and $E(Y) = E(X_1) + \cdots + E(X_M)$, $(E(\varepsilon) = 0)$,

$$Y - E(Y) = [X_1 - E(X_1)] + \cdots + [X_M - E(X_M)] + \varepsilon.$$

or

$$Y_D = X_{1D} + \cdots + X_{MD} + \varepsilon \tag{3.23}$$

where $Y_D = Y - E(Y)$, $X_{jD} = X_j - E(X_j)$. So if we consider the problem in this form, the numerical calculation is carried out by the following adjustments: calculate first

$$y_{iD}=y_i-\bar{y}, \quad \bar{y}=\frac{1}{N}\sum_{i=1}^{N} y_i, \quad i=1, \cdots, N \tag{3.24}$$

and $t_{j\alpha}$ based on these adjusted values y_{iD}; then solve the equation (3.20) to obtain the solution (3.22). We finally calculate

$$\bar{x}_j=\frac{1}{N}\sum_{\alpha=1}^{k_j} n_{j\alpha}x_{j\alpha} \tag{3.25}$$

to obtain the adjusted values of $\hat{x}_j$

$$\hat{x}_{j\alpha,D}=\hat{x}_{j\alpha}-\bar{x}_j, \quad \alpha=1, \cdots, k_j; \; j=1, \cdots, M. \tag{3.26}$$

Illustrative numerical example

The following example is prepared just for the illustration of the numerical calculation. Data are responses made by N = 20 persons on the 3 items; I_1: income, I_2: occupation, and I_3: habit of buying new production. Each person has one response in each item. The outside variable is the expenditure for clothing per month. Based on these given responses and outside variables, we wish to find suitable numerical values $\{x_{j\alpha}\}$, so that we can predict the expenditure for clothing on the basis of the information on the items. The response pattern is given in Table 3.2.

Now let us carry out the calculation according to the procedure explained above. The simultaneous equations of $x_{j\alpha}$ in a matric form are as follows:

$$\underset{\sim}{A}^*\underset{\sim}{x}^*=\underset{\sim}{t}^*$$

where

$$\underset{\sim}{A}^* = \left(\begin{array}{cc|ccc|cc} 7 & 0 & 2 & 1 & 3 & 3 & 4 \\ 0 & 8 & 4 & 1 & 2 & 5 & 3 \\ \hline 2 & 4 & 8 & 0 & 0 & 5 & 3 \\ 1 & 1 & 0 & 4 & 0 & 2 & 2 \\ 3 & 2 & 0 & 0 & 5 & 2 & 3 \\ \hline 3 & 5 & 5 & 2 & 2 & 11 & 0 \\ 4 & 3 & 3 & 2 & 3 & 0 & 9 \end{array}\right) \qquad \underset{\sim}{t}^* = \left(\begin{array}{c} t_{11} \\ t_{12} \\ \hline t_{21} \\ t_{22} \\ t_{23} \\ \hline t_{31} \\ t_{32} \end{array}\right) = \left(\begin{array}{r} -195.8 \\ 8.8 \\ \hline 40.8 \\ 94.4 \\ -145.0 \\ \hline 54.6 \\ -54.6 \end{array}\right)$$

Table 3.2

Items I_j	Income I_1			Occupation I_2				Habit buying New production I_3		Expenditure for Clothing Y	$Y-\overline{Y}$
Subcategories / Persons	C_{11} ~600	C_{12} 600 ~1000	C_{13} 1000~	C_{21} Salary	C_{22} Self Employed	C_{23} Labor	C_{24} Others	C_{31} Yes	C_{32} No		
1	✓			✓				✓		40	-27.4
2	✓					✓			✓	32	-35.4
3		✓		✓			✓	✓		72	4.6
4			✓				✓		✓	120	52.6
5	✓							✓		60	-7.4
6		✓				✓			✓	48	-19.4
7		✓		✓				✓		88	20.6
8	✓					✓		✓		48	-19.4
9			✓		✓				✓	100	32.6
10	✓			✓					✓	32	-35.4
11			✓	✓				✓		80	12.6
12		✓				✓		✓		40	-27.4
13		✓		✓					✓	56	-11.4
14	✓					✓			✓	24	-43.4
15			✓		✓			✓		112	44.6
16		✓					✓	✓		32	-35.4
17	✓				✓				✓	40	-27.4
18		✓		✓					✓	100	32.6
19			✓	✓				✓		112	44.6
20		✓			✓			✓		112	44.6
$n_{j\alpha}$	n_{11} 7	n_{12} 8	n_{13} 5	n_{21} 8	n_{22} 4	n_{23} 5	n_{24} 3	n_{31} 11	n_{32} 9	1348	0

$$A^{*-1} = \left(\begin{array}{cc|ccc|cc}
0.4132 & 0.2461 & -0.055 & 0.0593 & -0.1197 & -0.2110 & -0.2372 \\
0.2461 & 0.3660 & -0.0409 & 0.0495 & -0.0925 & -0.2071 & -0.1979 \\
\hline
-0.0055 & -0.0409 & 0.4652 & 0.3316 & 0.3409 & -0.3137 & -0.3263 \\
0.0593 & 0.0495 & 0.3316 & 0.5986 & 0.3242 & -0.3572 & -0.3945 \\
-0.1197 & -0.0925 & 0.3409 & 0.3242 & 0.5850 & -0.2456 & -0.2967 \\
\hline
-0.2110 & -0.2071 & -0.3137 & -0.3572 & -0.2456 & 0.4948 & 0.4286 \\
-0.2372 & -0.1979 & -0.3263 & -0.3945 & -0.2967 & 0.4286 & 0.5778
\end{array}\right)$$

t_{ij}'s are calculated based on the adjusted values of $Y_D = Y_i - \overline{Y}$. For example,

$$t_{11} = -27.4 -35.4 -7.4 -19.4 - 35.4 -43.4 - 27.4 = 195.8,$$

$$t_{12} = 4.6 - 19.4 + 20.6 - 27.4 - 11.4 - 35.4 +32.6 -44.6 = 8.8$$

Thus $\underset{\sim}{x}^* = \underset{\sim}{A}^{*-1}\underset{\sim}{t}$

$$\underset{\sim}{x}^* = \left(\begin{array}{cc|ccc|cc}
0.4132 & 0.2461 & -0.055 & 0.0593 & -0.1197 & -0.2110 & -0.2372 \\
0.2461 & 0.3660 & -0.0409 & 0.0495 & -0.0925 & -0.2071 & -0.1979 \\
\hline
-0.0055 & -0.0409 & 0.4652 & 0.3316 & 0.3409 & -0.3137 & -0.3263 \\
0.0593 & 0.0495 & 0.3316 & 0.5986 & 0.3242 & -0.3572 & -0.3945 \\
-0.1197 & -0.0925 & 0.3409 & 0.3242 & 0.5850 & -0.2456 & -0.2967 \\
\hline
-0.2110 & -0.2071 & -0.3137 & -0.3572 & -0.2456 & 0.4948 & 0.4286 \\
-0.2372 & -0.1979 & -0.3263 & -0.3945 & -0.2967 & 0.4286 & 0.5778
\end{array}\right)$$

$$\cdot \begin{pmatrix} -195.8 \\ 8.8 \\ 40.8 \\ 94.4 \\ -145.0 \\ 54.6 \\ -54.6 \end{pmatrix} = \begin{pmatrix} -54.06 \\ -29.06 \\ 2.26 \\ 13.90 \\ -14.92 \\ 32.22 \\ 29.02 \end{pmatrix}$$

Based on these $\underset{\sim}{x}^*$ values, we find $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}_3$, which enables us to find $\underset{\sim}{x}_D$.

$$\bar{x}_1 = \frac{1}{20}\,(n_{11}\hat{x}_{11} + n_{12}\hat{x}_{12}) = \frac{1}{20}\,\{7(-546) + 8(-29.06)\}$$

$$= -30.734$$

$$(\hat{x}_{11,D},\ \hat{x}_{12,D},\ \hat{x}_{13,D}) = (-54.6 + 30.73,\ =29.06 + 30.73,\ 30.73)$$

$$= (-23.871,\ 1.67,\ 30.73)$$

$$\bar{x}_2 = \frac{1}{20}\,(n_{21}\hat{x}_{21} + n_{22}\hat{x}_{22} + n_{23}\hat{x}_{23}) = \frac{1}{20}\ 18\ (2.26)$$

$$+ 4(12.9) + 5(-14.92) = 0.046$$

$$(\hat{x}_{21,D},\ \hat{x}_{22,D},\ \hat{x}_{23,D},\ \hat{x}_{24,D}) = (2.26 + 0.046,\ 13.9 + 0.046,$$

$$-\ 14.92 + 0.046,\ 0.046)$$

$$= (2.31,\ 13.95,\ -14.87,\ 0.05)$$

$$\bar{x}_3 = \frac{1}{20}\,(n_{31},\ \hat{x}_{31} + n_{32}\hat{x}_{32}) = \frac{1}{20}\,[11\ (32.22) + 9(29.02]$$

$$= 30.78$$

$$(\hat{x}_{31,D},\ \hat{x}_{32,D}) = (32.22 - 30.78,\ 29.02 - 30.78) = (1.44,\ -1.76)$$

$$\underset{\sim}{x}_D = \begin{pmatrix} \hat{x}_{11,D} \\ \hat{x}_{12,D} \\ \hat{x}_{13,D} \\ \cdots \\ \hat{x}_{21,D} \\ \hat{x}_{22,D} \\ \hat{x}_{23,D} \\ \hat{x}_{24,D} \\ \cdots \\ \hat{x}_{31,D} \\ \hat{x}_{32,D} \end{pmatrix} = \begin{pmatrix} -23.87 \\ 1.67 \\ 30.73 \\ \cdots\cdots \\ 2.31 \\ 13.95 \\ -14.87 \\ 0.05 \\ \cdots\cdots \\ 1.44 \\ -1.76 \end{pmatrix}$$

From this $\underset{\sim}{x}_D$ we calculate the score for each person, i.e., $S_{i,D}$ as follows:

$$S_{1,D} = \hat{x}_{11,D} + \hat{x}_{21,D} + \hat{x}_{31,D} = 23.87 + 2.31 + 1.44$$

$$= 20.12,$$

$$S_{2,D} = \hat{x}_{11,D} + \hat{x}_{23,D} + \hat{x}_{32,D} = -23.87 - 14.87 - 1.76$$

$$= -40.50,$$

$$S_{3,D} = \hat{x}_{12,D} + \hat{x}_{21,D} + \hat{x}_{31,D} = 1.67 + 2.31 + 1.44$$

$$= 5.42.$$

Other scores are given as:

$$S_{4,D} = 29.02,\ S_{5,D} = -22.38,\ S_{6,D} = -14.96,\ S_{7,D} = 5.42$$

$$S_{8,D} = -3.73,\ S_{9,D} = 42.92,\ S_{10,D} = -23.32,\ S_{11,D} = 34.48$$

$$S_{12,D} = -11.76,\ S_{13,D} = 2.22,\ S_{14,D} = -40.5,\ S_{15,D} = 46.12$$

$$S_{16,D} = 3.16,\ S_{17,D} = 11.68,\ S_{18,D} = 2.22,\ S_{19,D} = 34.48$$

$$S_{20,D} = 17.06$$

From these results we obtain $\sum_{i=1}^{n} S_{i,D}^{\,2}$ and ρ.

$$\sum_{i=1}^{n} S_{i,D}^{\,2} = 14,178.5$$

$$\sum_{i=1}^{n} (Y_i - y)^2 = \sum Y_{i,D}^{\,2} = 20296.8,\quad \frac{1}{20}\sum_{i=1}^{n} Y_{i,D}^{\,2} = 1014.9$$

$$\sum Y_{i,D}^{\;2} - \sum S_{i,D}^{\;2} = 20296.8 - 14178.5336 = 6118.3$$

$$\rho_{YS} = \frac{\sum(Y_i-\bar{y})(S_i-\bar{s})}{\sqrt{\sum(Y_i-\bar{y})^2 S_{i,D}^{\;2}}} = \frac{\sum(Y_i-\bar{y})(S_{i,D})}{\sqrt{\sum(Y_i-\bar{y})^2\; S_{i,D}^{\;2}}} = \frac{14177.4}{\sqrt{(20296.8)(14178.5)}}$$

$$= 0.84$$

Thus the minimum of the average of the squared differences between Y-values and scores, which corresponds to (3.14), is

$$V^2 = \frac{1}{20}\,[\sum Y_{i,D}^{\;2} - \sum S_{i,D}^{\;2}] = 305.9 \text{ or } V = 17.5.$$

Hence the relative rate of the reduction of the variation of Y_i's is

$$(1 - \frac{V^2}{\frac{1}{20}\sum Y_{i,D}^{\;2}}) \times 100 = (\frac{305.9}{1014.9}) \times 100 = 69.86\ (\%).$$

$\rho_{YS} = 0.84$ shows us that the method of quantification is pretty good; that is, we can predict a value of the outside variable for a person from the score value given to him with a good accuracy.

It must be noted here that the numerical example given above is an artificial one. For the actual applications, sample size $N = 20$ is very small because the numerical values $\{x_{j\alpha}\}$ have not enough accuracy. To apply the method effectively, we need a sample large enough to guarantee the stable values of $\{x_{j\alpha}\}$. The sample size for this depends on the total number of categories which we intend to give the numerical values.

3.2 The case where the outside variable is categorical; an application of classification analysis.

When the outside variable Y is categorical, how categories $c_{j\alpha}$'s in the last section are quantified in order to predict the response of a person on Y as effectively as possible? Let $G_1, G_2, \cdots, G_s$ are s categories of Y. The response on Y by each person is also assumed to be only one of them. The quantification in this situation can be solved by regarding s categories $G_1, \cdots, G_s$ as s groups and by applying the concept and technique of classification analysis.

Let π_ν be the probability that a randomly chosen person responds to G_ν, $(\nu=1, \cdots, s)$, so that $\sum_\nu \pi_\nu = 1$. Let the variables $X_{1(\nu)}, \cdots, X_{M(\nu)}$ be the item variables $X_1, \cdots, X_M$ restricted in the ν-th category (group) G_ν, and

$$Z = X_1 + \cdots + X_M$$

$$Z(\nu) = X_{(\nu)1} + \cdots + X_{(\nu)M}. \tag{3.27}$$

We use the same notations as in the last section. The values of $x_{j\alpha}$ to be determined are now those maximizing the correlation ratio $\eta^2 = \sigma_B^2/\sigma_T^2$ where σ_B^2 is the between variance for $G_1, \cdots, G_s$ and σ_T^2 is the total variance of Z. Here the role of discriminator are played by $Z = X_1 + \cdots + X_M$ or $Z_{(\nu)} = X_{(\nu)1} + \cdots + X_{(\nu)M}$.

Now we need to express σ_B^2 and σ_T^2 in terms of $x_{j\alpha}$'s.

$$\sigma_B^{\ 2} = \sum_{\nu=1}^{s} \pi_\nu \{E(Z_{(\nu)}) - E(Z)\}^2$$

$$= \sum_{\nu=1}^{s} \pi_\nu \{ \sum_{j=1}^{M} \sum_{\alpha=1}^{k_j} x_{j\alpha} p_{(\nu)j\alpha} - \sum_{j=1}^{M} \sum_{\alpha=1}^{k_j} x_{j\alpha} p_{j\alpha} \}^2$$

$$= \sum_{\nu=1}^{s} \pi_\nu \sum_{j=1}^{M} \sum_{k=1}^{M} \sum_{\alpha=1}^{k_j} \sum_{\beta=1}^{k_k} x_{j\alpha} x_{k\beta} (P_{(\nu)j} - p_{j\alpha})$$

$$(p_{(\nu)k\beta} - p_{k\beta}), \tag{3.28}$$

where

$$p_{(\nu)j\alpha} = P(X_{(\nu)j} = x_{j\alpha}) = P(X_j = x_{j\alpha} | Y = G_\nu)$$

= the probability that the category $C_{j\alpha}$ of the j-th item is responded by a person under the condition that his response on the outside variable Y belongs to G_ν.

$$\sigma_T^{\ 2} = E(Z^2) - E^2(Z) = \sum_{j=1}^{M} \sum_{k=1}^{M} E(X\ X\ ) - [\sum_{j=1}^{M} E(X_j)]^2$$

$$= \sum_{j=1}^{M} \sum_{k=1}^{M} \sum_{=1}^{k_j} \sum_{=1}^{k_k} x_{j\alpha} x_{k\beta} p_{jk|\alpha\beta} - [\sum_{j=1}^{M} \sum_{\alpha=1}^{k_j} x_{j\alpha} p_{j\alpha}]^2$$

$$= \sum_{j}\sum_{k}\sum_{\alpha}\sum_{\beta} \{p_{jk|\alpha\beta} - p_{j\alpha} p_{k\beta}\} x_{j\alpha} x_{k\beta} \tag{3.29}$$

The maximization of η^2 is equivalent to the maximization of $\sigma_B^{\ 2}$ under $\sigma_T^{\ 2} = 1$; that is, let

$$\phi = \sigma_B^{\ 2} - \lambda(\sigma_T^{\ 2} - 1) \tag{3.30}$$

where λ is the Lagrange multiplier, and we solve the equation $\partial\phi/\partial x_{j\alpha} = 0$ for $\alpha=1, \cdots, k_j$; $j=1, \cdots, M$. This gives us the following simultaneous equations:

$$\sum_{\nu=1}^{s} \sum_{k=k}^{M} \sum_{\beta=1}^{k_k} \pi_\nu [p_{(\nu)j\alpha} - p_{j\alpha}] \; [p_{(\nu)k\beta} - p_{k\beta}] x_{k\beta}$$

$$= \lambda \sum_{k=1}^{M} \sum_{\beta=1}^{k_k} [p_{jk|\alpha\beta} - p_{j\alpha} p_{k\beta}] x_{k\beta} \tag{3.31}$$

$$\alpha=1, \cdots, k_j; \; j=1, \cdots, M$$

Multiplying the both sides of (3.31) by $x_{j\alpha}$ and summing up with respect to j and α, we find that $\lambda = \eta^2$; so the desired values of $x_{j\alpha}$'s can be obtained by the solution $\{\hat{x}_{j\alpha}\}$ corresponding to the largest characteristic root of the determinental equation in λ obtained from (3.31).

Now consider the (3.31) obtained by substituting the estimates of population parameters. Let us define additional notations:

N_ν = the number of persons responding (belonging) to G_ν among N persons,

$n_{(\nu)j\alpha}$ = the number of persons responding to $c_{j\alpha}$ among N_ν persons belonging to G_ν.

Then the simultaneous equations (3.31) are replaced with

$$\sum_{k=k}^{M} \sum_{\beta=1}^{k_k} [\sum_{\nu=1}^{s} \frac{1}{N\nu} n_{(\nu)j\alpha} n_{(\nu)k\beta} - \frac{1}{N} n_{j\alpha} n_{k\beta}] x_{k\beta}$$

$$= \lambda \sum_{k=1}^{M} \sum_{\beta=1}^{k_k} [n_{jk|\alpha\beta} - \frac{1}{N} n_{j\alpha} n_{k\beta}] x_{k\beta} \tag{3.32}$$

for $\alpha=1, \cdots, k_j$ and $j=1, \cdots, M$, since

$$\hat{\pi}_\nu = N_\nu/N, \quad \hat{p}_{(\nu)j\alpha} = n_{(\nu)j\alpha}/N_\nu$$

and $\sum_{\nu} N_{\nu} = N$, $\sum_{\nu} n_{(\nu)j\alpha} = n_{j\alpha}$.

In order to express (3.32) in matric form, let

$$\sum_{y=1}^{s} \frac{1}{N_{\nu}} n_{(\nu)j\alpha} n_{(\nu)k\beta} - \frac{1}{N} n_{j\alpha} n_{k\beta} \equiv h_{jk|\alpha\underline{\underline{\beta}}}$$

$$n_{jk|\alpha\beta} - \frac{1}{N} n_{j\alpha} n_{k\beta} \equiv f_{jk|\alpha\beta} \qquad (3.32)$$

and

$$\underset{\sim}{x} = \begin{pmatrix} x_{11} \\ \cdot \\ \cdot \\ \cdot \\ x_{1k_1} \\ x_{21} \\ \cdot \\ \cdot \\ \cdot \\ x_{2k_2} \\ \cdot \\ \cdot \\ \cdot \\ x_{M1} \\ \cdot \\ \cdot \\ \cdot \\ x_{Mk_M} \end{pmatrix} \qquad \underset{\sim}{F} = \begin{pmatrix} \underset{\sim}{F}_{11} & \underset{\sim}{F}_{12} & \cdots & \underset{\sim}{F}_{1M} \\ \underset{\sim}{F}_{21} & \underset{\sim}{F}_{22} & \cdots & \underset{\sim}{F}_{2M} \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{F}_{M1} & \underset{\sim}{F}_{M2} & \cdots & \underset{\sim}{F}_{MM} \end{pmatrix}$$

$$\underset{\sim}{H} = \begin{pmatrix} \underset{\sim}{H}_{11} & \underset{\sim}{H}_{12} & \cdots & \underset{\sim}{H}_{1M} \\ \underset{\sim}{H}_{21} & \underset{\sim}{H}_{22} & \cdots & \underset{\sim}{H}_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \underset{\sim}{H}_{M1} & \underset{\sim}{H}_{M2} & \cdots & \underset{\sim}{H}_{MM} \end{pmatrix}$$

where

$$\underset{\sim}{F}_{jk} = \begin{pmatrix} f_{jk/11} & f_{jk/12} & \cdots & f_{jk/1k_k} \\ f_{jk/21} & f_{jk/22} & \cdots & f_{jk/2k_k} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ f_{jk/k_j1} & f_{jk/k_j2} & \cdots & f_{jk/k_jk_k} \end{pmatrix},$$

$$\underset{\sim}{H}_{jk} = \begin{pmatrix} h_{jk/11} & h_{jk/12} & \cdots & h_{jk/1k_k} \\ h_{jk/21} & h_{jk/22} & \cdots & h_{jk/2k_k} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ h_{jk/k_j1} & h_{jk/k_j2} & \cdots & h_{jk/k_jk_k} \end{pmatrix}$$

It is noted that $\underset{\sim}{H}$ and $\underset{\sim}{F}$ are both $(k_1 + \cdots + k_M) \times (k_1 + \cdots k_M)$ and symmetric matrices. Then (3.32) can now be written as

$$\underset{\sim}{H}x = \lambda \underset{\sim}{F}\underset{\sim}{x} \text{ or } (\underset{\sim}{H} - \lambda\underset{\sim}{F})\underset{\sim}{x} = \underset{\sim}{0}. \tag{3.34}$$

It is seen from this equation that $\underset{\sim}{x}$ should be the characteristic vector corresponding to a root λ of the equation

$$|\underset{\sim}{H} - \lambda\underset{\sim}{F}| = 0 \tag{3.35}$$

Since λ is found to be the square of the sample correlation ratio, our desired solution is the characteristic vector $\hat{\underset{\sim}{x}}_m$ corresponding to the largest root $\hat{\lambda}_m$ of the equation (3.35).

<u>Note 3.1</u>. Matrices $\underset{\sim}{H}_{jk}$ and $\underset{\sim}{F}_{jk}$ are singular, since

$$\sum_{\alpha=1}^{k_j} h_{jk|\alpha\beta} = \sum_{\nu=1}^{s}\Big(\sum_{\alpha=1}^{k_j} \frac{1}{N_\nu} n_{(\nu)j\alpha}\Big) n_{(\nu)k\beta} - \frac{1}{N}\Big(\sum_{\alpha=1}^{k_j} n_{j\alpha}\Big) n_{k\beta}$$

$$= \sum_{\nu=1}^{s} n_{(\nu)k\beta} - n_{k\beta} = 0, \qquad (j,k=1,\cdots,M;\ \beta=1,\cdots,k_k)$$

$$\sum_{\alpha=1}^{k_j} f_{jk|\alpha\beta} = \sum_{\alpha=1}^{k_j} n_{jk|\alpha\beta} - \Big(\frac{1}{N}\sum_{\alpha=1}^{k_j} n_{j\alpha}\Big) n_{k\beta} = n_{k\beta} - n_{k\beta} = 0$$

$$(j,\ k=1,\cdots,M;\ \beta=1,\cdots,k_k).$$

Consequently, in the solving of (3.34) and (3.35), we may, at the beginning, put, for example, $\hat{x}_{1k_1} = 0$, $\hat{x}_{2k_2} = 0$, $\cdots$, $\hat{x}_{Mk_m} = 0$ and correspondingly we may delete from $\underset{\sim}{H}$ and $\underset{\sim}{F}$ the k_1-column and row, k_2-column and row, $\cdots$, k_M-column and row.

Note 3.2. When s=2, i.e. in the case of two categories (groups), we have

$$\sigma_\beta{}^2 = \pi_1 \{E(Z_{(1)}) - E(Z)\}^2 + \pi_2 \{E(Z_{(2)}) - E(Z)\}^2$$

$$= \pi_1 \{E(Z_{(1)}) - \pi_1 E(Z_{(1)}) - \pi_2 E(Z_{(2)})\}^2$$

$$+ \pi_2 \{E(Z_{(2)}) - \pi_1 E(Z_{(1)}) - \pi_2 E(Z_{(2)})\}^2$$

$$= 2\pi_1\pi_2\{E(Z_{(1)}) - E(Z_{(2)})\}^2.$$

Hence the maximizing η^2 is equivalent to the maximizing

$$\frac{\{E(Z_{(1)}) - E(Z_{(2)})\}^2}{E(Z^2) - E^2(Z)}$$

<u>Illustrative Numerical Example</u>. In this example, let us use the same data sets that appeared in the last example. Thus, we have the same response pattern for each person except that the outside variables are not given as numerical values. Instead, the outside variable is now represented as the response of each person, and in this example, it is the response of renting an apartment or owning a house. Table 3.3 is the response pattern of this example.

Again we wish to find the suitable numerical values for $\{x_{j\alpha}\}$, so that we can predict the type of housing for each person on the basis of the given information. According to the Note 3.1, the calculation is carried out as follows:

The simultaneous equations in $x_{j\alpha}$ are expressed as:

$$\underset{\sim}{H}^*\underset{\sim}{x}^* = \lambda\underset{\sim}{F}^*\underset{\sim}{x}^* \text{ or } (\underset{\sim}{H}^*-\lambda\underset{\sim}{F}^*)\underset{\sim}{x}^* = 0; \qquad (3.36)$$

in other words, $\underset{\sim}{x}^*$ is the characteristic vector corresponding to the characteristic root λ of the following equation:

$$|\underset{\sim}{H}^* - \underset{\sim}{F}^*| = 0 \qquad (3.37)$$

where

$\underset{\sim}{x}^*$, $\underset{\sim}{H}^*$, and $\underset{\sim}{F}^*$ are vector and matrices formed by the deletion in Note 3.1, and

Table 3.3

Items	Income I_1			Occupation I_2				Habit of buying new production I_3		Outside Variable	
Subcategories / Persons	C_{11} ~600	C_{12} 600 ~1000	C_{13} 1000~	C_{21} salary	C_{22} self employed	C_{23} labor	C_{24} other	C_{31} yes	C_{32} no	Rent an Apt	Own a House
1	✓			✓				✓			✓
2	✓					✓			✓	✓	
3		✓		✓				✓		✓	
4			✓				✓		✓		✓
5	✓						✓	✓		✓	
6		✓				✓			✓	✓	
7		✓		✓				✓		✓	
8	✓					✓		✓			✓
9			✓		✓				✓	✓	
10	✓			✓					✓	✓	
11			✓	✓				✓		✓	
12		✓				✓		✓			✓
13		✓		✓					✓	✓	
14	✓					✓			✓		✓
15			✓		✓			✓		✓	
16		✓					✓	✓		✓	
17	✓				✓				✓	✓	
18		✓		✓					✓	✓	
19			✓	✓				✓			✓
20		✓			✓			✓		✓	
$n_{j\alpha}$	n_{11} 7	n_{12} 8	n_{13} 5	n_{21} 8	n_{22} 4	n_{23} 5	n_{24} 3	n_{31} 11	n_{32} 9	N_1 14	N_2 6

$$H^* = \left(\begin{array}{cc:ccc:c}
0.19 & -0.3 & -0.09 & -0.26 & 0.32 & 0.15 \\
-0.3 & 0.47 & 0.13 & 0.4 & -0.5 & -0.23 \\
\hdashline
-0.09 & 0.13 & 0.04 & 0.11 & -0.14 & -0.07 \\
-0.26 & 0.4 & 0.11 & 0.34 & 0.43 & -0.2 \\
0.32 & -0.5 & -0.14 & -0.43 & 0.54 & 0.25 \\
\hdashline
0.15 & -0.23 & -0.07 & -0.2 & 0.25 & 0.12
\end{array}\right)$$

$$F^* = \left(\begin{array}{cc:ccc:c}
4.55 & -2.8 & -0.8 & -0.4 & 1.25 & -0.85 \\
-2.8 & 4.8 & 0.8 & -0.6 & 0 & 0.6 \\
\hdashline
-0.8 & 0.8 & 4.8 & -1.6 & -2 & 0.6 \\
-0.4 & -0.6 & -1.6 & 3.2 & -1 & -0.2 \\
1.25 & 0 & -2 & -1 & 3.75 & -0.75 \\
\hdashline
-0.85 & 0.6 & 0.6 & -0.2 & -0.75 & 4.95
\end{array}\right)$$

$h_{jk|\alpha\beta}$ and $f_{jk|\alpha\beta}$ are calculated based on the equation (3.33); for example,

$$h^{11}/_{11} = \frac{1}{14}(4)(4) + \frac{1}{6}(3)(3) - \frac{1}{20}(7)(7) = 0.19$$

$$h^{11}/_{12} = \frac{1}{14}(4)(7) + \frac{1}{6}(3)(1) - \frac{1}{20}(7)(8) = -0.3$$

$$f^{11}/_{11} = 7 - \frac{1}{20}(7)(7) = 4.55$$

$$f^{11}/_{12} = 0 - \frac{1}{20}(7)(8) = -2.8$$

Others are given in the same way.

By solving equations (3.36)and (3.37), we find the largest characteristic root $\lambda = 0.401$ with corresponding characteristic vector $\hat{x}^*$

$$\hat{\underset{\sim}{x}}^* = \begin{bmatrix} 0.24 \\ 0.59 \\ 0.08 \\ 0.32 \\ -0.65 \\ -0.24 \end{bmatrix}$$

According to x*, we have $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, and x_D calculated as follows:

$$\bar{x}_1 = \frac{1}{20}(n_{11}\hat{x}_{11}+n_{12}\hat{x}_{12}) = \frac{1}{20}(7[0.24]+[0.59]) = 0.32$$

$$(\hat{x}_{11,D}, \hat{x}_{12,D}, \hat{x}_{12,D}) = (0.24{=}0.32,\ 0.59-0.32,\ -0.32)$$

$$= (-0.08,\ 0.27,\ -0.32)$$

$$\bar{x}_2 = \frac{1}{20}(8[0.08]+4[0.32]+5[-0.65]) = -0.0665$$

$$(\hat{x}_{21,D}, \hat{x}_{22,D}, \hat{x}_{23,D}, \hat{x}_{24,D}) = (0.08+0.0665,\ 0.32+0.0665,\ -0.65+0.0665,\ 0.0665)$$

$$= (0.14,\ 0.38,\ -0.57,\ 0.07)$$

$$\bar{x}_3 = \frac{1}{20}(11[-0.24]) = -0.132$$

$$(\hat{\alpha}_{31,D}, \hat{\alpha}_{32,D}) = (-0.24+0.132,\ 0.132) = (-0.108,\ 0.132)$$

$$\underset{\sim}{x}_D = \begin{pmatrix} -0.08 \\ 0.27 \\ -0.32 \\ 0.14 \\ 0.38 \\ -0.57 \\ 0.07 \\ -0.108 \\ 0.132 \end{pmatrix}$$

To see whether this method of quantification is effective or not, we can use the $\underset{\sim}{x}_D$ valve to find the correlation ratio which is

$$\eta^2 = \frac{2\pi_1\pi_2(E[Z_1]-E[Z_2])^2}{E(Z^2)-(E[Z])^2} = \frac{2(\frac{14}{20})(\frac{6}{20})(0.259)}{0.1441} = 0.7$$

From the observed μ^2, we see that this is a good method of quantification.

Finally the same comment on the sample size stated at the end of the last example is also necessary here.

CHAPTER IV. QUANTIFICATION THROUGH THE ASSOCIATION

Suppose there are M factors $F_1, \cdots, F_M$ to be made the response by each of N individuals or persons. Suppose each individual can choose or respond to any number of M factors. For example, the factors are M kinds of T.V. programs and each of N persons is asked to check a number of them which he likes. Individuals may be N species of a certain family of plants or animals and factors may be M qualitative characters. The responses made by N individuals will show various types. Suppose there are s types. Then the result of the experiment or survey will be summarized as in the table below:

TABLE 4.1 Observation Pattern

Types \ Factors			F_1	F_2	F_3	F_4	......	F_{M-1}	F_M
			y_1	y_2	y_3	y_4	......	y_{M-1}	y_M
T_1	N_1	x_1	V	V		V	......		V
T_2	N_2	x_2		V	V		......	V	
T_3	N_3	x_3	V		V		......	V	
T_4	N_4	x_4		V	V		......		V
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
T_{s-1}	N_{s-1}	x_{s-1}	V			V	......	V	V
T_3	N_s	x_s		V			......		V
Total	N								

In the table n_i $(i=1, \cdots, s)$ is the number of individuals showing the i-th response type, so that $\sum_{i=1}^{s} n_i=N$, and x_i's and y_j's are numerical values which we are going to determine under the consideration explained below.

First of all we must state the purpose of the analysis. We assume that the individuals having the similar natures choose or respond to the factors having the similar characteristics. We are interested in the information "what among M factors are chosen or responded by individuals of what types and vice versa." In other words, the purpose of the study is to have the classification method by which we may predict or classify the factors chosen or responded by the individuals by knowing what type they belong to and, at the same time, predict or classify the types of individuals by knowing what factors are chosen or responded by them.

To accomplish the simultaneous classification of this kind, the following procedure may be considered: we try to rearrange the types of individuals in such a way that if they have the similar response patterns in the choice of factors, then we put them near each other; on the other hand if their response patterns are dissimilar, then we put them far from each other. At the same time we try to rearrange the factors in wuch a way that if the factors are chosen by similar types of individuals, then we put them near each other; on the other hand, the factors, which are chosen by different types of individuals, are put far from each other.

For example, from the original pattern shown in Table 4.2a, we try to obtain the rearrangements of the types and factors shown in Table 4.2b. This means that we try to rearrange them so that "V's" gather along the main diagonal as near as possible.

TABLE 4.2a

Types \ Factors	1	2	3	4	5
1			V	V	
2	V		V		V
3		V		V	
4		V	V		
5	V				V
6			V		V

TABLE 4.2b.

Types \ Factors	1	5	3	4	2
5	V	V			
2	V	V	V		
6		V	V		
1			V	V	
4			V		V
3				V	V

If numbers of types and factors are so small, we may do this work just by inspection, but this in general is not the case in the many practical problems. To carry out this when those numbers are large, the following quantification technique is useful: Let x_i, $i=1, \cdots, x$ and y_j, $j=1, \cdots, M$ be the numerical values to be given to s types and M factors respectively in order to work out the aim explained above. Hence $a(x_i, y_j)$ corresponds to each response "V". The problem of gathering "V's" along the main diagonal is now translated into the problem of gathering the corresponding (x_i, y_j)'s along a line as near as possible. This is again equivalent to the problem of maximizing the correlation coefficient p_{xy} between a random variable x taking values $x_1, \cdots, x_s$ and a random variable Y taking values $y_1, \cdots, y_m$; that is, our quantification is now formulated as the determination of values of x_i's and y_j's so that

$$P_{XY} = C_{XY}/\sigma_X\sigma_Y \tag{4.1}$$

is maximum, where C_{XY} is the covariance of X and Y, and σ_X, σ_Y are standard deviations of X and Y, respectively.

Now let

$$\delta_i(j) = \begin{cases} 1, & \text{if an individual of the i-th type chooses the j-th factor} \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

Then

$$m_i = \sum_{j=1}^{M} \delta_i(j) \tag{4.3}$$

is the number of marks "V" chosen by the i-th type, and

$$N\bar{M} = \sum_{i=1}^{s} n_i m_i \tag{4.4}$$

is the total number of marks "V" chosen by N individuals.

$$P_{ij} = \frac{n_i}{N\bar{m}} \delta_i(u) \tag{4.5}$$

is the proportion of the marks "V" in the cell (T_i, F_j).

Hence

$$p_i = \sum_{j=1}^{M} p_{ij} = \frac{ni}{N\bar{m}} \sum_{j=1}^{M} \delta_i(j) = \frac{n_i m_i}{N\bar{m}}, \qquad (4.6)$$

$$q_j = \sum_{i=1}^{S} p_{ij} = \frac{1}{N\bar{m}} \sum_{i=1}^{S} n_i \delta_i(j) \qquad (4.7)$$

are the proportions of the marks "V" for T_i and for F_j, respectively. We can now regard (X,Y) as a bivariate random variate with

$$P(X=x_i,\ Y=y_j) = P_{ij}, \quad i=1, \cdots, s;\ j=1, \cdots, M. \qquad (4.8)$$

Then $e_{XY}, \sigma_X{}^2$ and $\sigma_Y{}^2$ are expressed as

$$e_{XY} = \sum_{i=1}^{S} \sum_{j=1}^{M} x_i y_j p_{ij} - \left(\sum_{i=1}^{S} x_i p_i\right)\left(\sum_{j=1}^{M} y_i q_j\right), \qquad (4.9)$$

$$\sigma_X{}^2 = \sum_{i=1}^{S} x_i{}^2 p_i - \left(\sum_{i=1}^{S} x_i p_i\right)^2, \qquad (4.10)$$

$$\sigma_Y{}^2 = \sum_{j=1}^{M} y_j{}^2 q_j - \left(\sum_{j=1}^{M} y_j q_j\right)^2. \qquad (4.11)$$

Now then we calculate $\{x_i\}$ and $\{y_i\}$ maximizing p_{XY}; that is, we calculate

$$\frac{\partial P_{XY}}{\partial x_i} = 0, \ \frac{\partial P_{XY}}{\partial y_j} = 0 \qquad (4.12)$$

$$i=1, \cdots, S;\ j=1, \cdots, M.$$

Since $\log P_{XY} = \log C_{XY} - \frac{1}{2} \log \sigma_X{}^2 - \frac{1}{2} \log \sigma_Y{}^2$, we have

$$\frac{\partial P_{XY}}{\partial x_i} = \frac{P_{XY}}{C_{XY}} \frac{\partial C_{XY}}{\partial x_i} - \frac{P_{XY}}{2\sigma_X^2} \frac{\partial \sigma_X^2}{\partial x_i} = 0$$

$$\frac{\partial P_{XY}}{\partial y_i} = \frac{P_{XY}}{C_{XY}} \frac{\partial C_{XY}}{\partial y_j} - \frac{P_{XY}}{2\sigma_Y^2} \frac{\partial \sigma_Y^2}{\partial y_j} = 0$$

and hence

$$\frac{\partial C_{XY}}{\partial x_i} = \frac{1}{2} \lambda \frac{\partial \sigma_X^2}{\partial x_j} \tag{4.13}$$

$$\frac{\partial C_{XY}}{\partial y_j} = \frac{1}{2} \lambda_2 \frac{\partial \sigma_Y^2}{\partial y_j} \tag{4.14}$$

where $\lambda_1 = P_{XY}\sigma_Y/\sigma_X$ and $\lambda_2 = P_{XY}\sigma_X/\sigma_Y$. Calculating the derivatives by using the expressions (4.9), (4.10), and (4.11), we obtain the simultaneous equations to be solved for x_i's and y_j's,

$$\sum_{j=1}^{M} (p_{ij}-p_i q_j) y_j = \lambda_1 (x_i - \sum_{k=1}^{S} x_k p_k) p_i \tag{4.15}$$

$$\sum_{i=1}^{S} (p_{ij}-p_i q_j) x_i = \lambda_2 (y_j - \sum_{\ell=1}^{M} y_\ell q_\ell) q_j \tag{4.16}$$

$$(j=1, \cdots, M)$$

Multiplying the both sides of (4.15) by $(P_{i\ell}/P_i)$ and the summing up with respect to i, we obtain

$$\sum_{j=1}^{M} \{ \sum_{i=1}^{S} \frac{p_{ij}p_i}{p_i} - q_\ell q_j \} \; y_j = \lambda_1 \; \{ \sum_{i=1}^{S} x_i p_{i\ell} = q_\ell \sum_{k=1}^{S} x_k p_k \}$$

$$= \lambda_1 \sum_{i=1}^{S} (p_{i\ell} - p_i q_\ell) x_i$$

Substituting this into (4.16),

$$\sum_{j=1}^{M} \{ \sum_{i=1}^{S} \frac{p_{ij}p_{i\ell}}{p_i} - q_j q_\ell \} y_j = p_{XY}^2 \; \{ y_\ell - \sum_{j=1}^{M} y_j q_j \} \; q_\ell$$

$$(\ell=1, \cdots, M) \qquad (4.17)$$

Similarly we have

$$\sum_{i=1}^{S} \{ \sum_{j=1}^{M} \frac{p_{kj}p_{ij}}{q_i} - p_i p_k \} \; x_i = \{ x_k - \sum_{i=1}^{S} x_i p_i \} \; p_k. \qquad (4.18)$$

$$(k=1, \cdots, s).$$

It is easily seen that the solutions $\{x_i\}$ of (4.18) and $\{y_j\}$ of (4.17) do not depend on the origin and hence we may put

$$\sum_{i=1}^{S} x_i p_i = 0, \qquad \sum_{j=1}^{M} y_j q_j = 0 \qquad (4.19)$$

Thus we can finally express the simultaneous equations to be solved as

$$\sum_{j=1}^{M} (\sum_{i=1}^{S} \frac{p_{ij}p_{i\ell}}{p_i}) y_j = p_{XY}^2 q_\ell y_\ell, \quad \ell=1, \cdots, M \qquad (4.20)$$

$$\sum_{i=1}^{S} (\sum_{j=1}^{M} \frac{p_{kj}p_{ij}}{q_j}) x_i = p_{XY}^2 p_k x_k, \quad =1, \cdots, S. \qquad (4.21)$$

It is noted here that, in the actual computation, we need not to solve two systems (4.20) and (4.21), but just solve one of them, say (4.20), because if we denote the solution of (4.20) by $\{\hat{y}_\ell\}$, then from (4.15) with conditions in (4.19) we immediately obtain the solution $\{\hat{x}_k\}$ by

$$\hat{x}_k = \frac{1}{\lambda_1}\frac{1}{k}\sum_{j=1}^{M} p_{kj}\hat{y}_j, \qquad k=1, \cdots, S$$

or putting $\lambda_1 = 1$,

$$\hat{x}_k = \frac{1}{k}\sum_{j=1}^{M} p_{kj}\hat{y}_j, \qquad k=1, \cdots, S. \tag{4.22}$$

Now (4.20) can be rewritten as

$$\sum_{j=1}^{M}\Big[\sum_{i=1}^{S}\frac{n_i}{m_i}\delta_i(\ell)\delta_i(j)\Big]y_j = p_{XY}^2\Big[\sum_{i=1}^{S} n_i\delta_i(\ell)\Big]y\ ,$$

$$\ell=1, \cdots, M. \tag{4.23}$$

So if we let

$$a_{\ell i} = \sum_{i=1}^{S}\frac{n_i}{m_i}\delta_i(\ell)\delta_i(j), \qquad d_\ell = \sum_{i=1}^{S} n_i\delta_i(\ell), \tag{4.24}$$

we have the following equation in the matric form:

$$\left.\begin{aligned} &\underset{\sim}{A}\underset{\sim}{y} = p_{XY}^2\underset{\sim}{D}\underset{\sim}{y} \\ \text{or}\quad & \\ &\underset{\sim}{D}^{-1}\underset{\sim}{A}\underset{\sim}{y} = p_{XY}^2 y \end{aligned}\right\} \tag{4.25}$$

where

$$\underset{\sim}{A} = (a_{\ell i}), \quad \underset{\sim}{D} = \mathrm{diag}(d_1, d_2, \cdots, d_M)$$

$$\underset{\sim}{y}' = (y_1, y_2, \cdots, y_M). \tag{4.26}$$

From (4.25) our solution $\hat{\underset{\sim}{y}}$ is obtained as the characteristic vector corresponding to the largest characteristic root of $\underset{\sim}{D}^{-1}\underset{\sim}{A}$.

We finally summarize the expressions needed to calculate the desired values of $\{\hat{y}_\ell\}$ and $\{\hat{\underset{\sim}{x}}_k\}$:

(i) Obtain the largest root p^2_{MAX} of the equations

$$|\underset{\sim}{A} - p^2 D| = 0,$$

(ii) Obtain the characteristic vector $\hat{\underset{\sim}{y}}$ corresponding p^2_{MAX},

(iii) Calculate

$$\bar{\hat{y}} = \frac{1}{N\bar{m}} \sum_{j=1}^{M} [\sum_{i=1}^{S} n_i \delta_i(j)] \hat{y}_j \tag{4.27}$$

and

$$\hat{y}_{j \cdot D} = \hat{y}_j - \bar{\hat{y}}, \; j=1, \cdots, M. \tag{4.28}$$

(iv) Calculate

$$\hat{x}_k = \frac{1}{m_k} \sum_{j=1}^{M} \delta_k(j) \hat{y}_{j \cdot D}, \tag{4.29}$$

(v) Calculate

$$\bar{\hat{x}} = \frac{1}{N\bar{m}} \sum_{k=1}^{S} n_k m_k \hat{x}_k \tag{4.30}$$

and

$$\hat{x}_{k \cdot D} = \hat{x}_k - \bar{\hat{x}}, \qquad k=1, \cdots, S \tag{4.31}$$

Our final numerical solutions are then $\{x_{k \cdot D}\}$ and $\{y_{j \cdot D}\}$.

REFERENCES

1. Fisher, R. A., Statistical Methods for Research Workers, Eighth Edition, Oliver and Boyd, 1941.

2. Guilford, J. P., Psychiatric Methods, McGraw Hill, 1936.

3. Guttman, L., "The quantification of a class of attributes: a theory and method of scale construction," in P. Horst, et. al., "The Prediction of Personal Adjustment," Soc. Sci. Res. Council, New York, 1941.

4. Guttman, L., "A basis for scaling qualitative data," Amer. Sociological Review, Vol. 9, 1944.

5. Guttman, L., "An Approach for Quantifying Paired Comparisons and Rank Order," The Annals of Mathematical Statistics, Vol. 17, No. 2, 1946.

6. Hayashi, C., "On the Quantification of Qualitative Data from the Mathematics-Statistical Point of View," Annals of the Institute of Statistical Mathematics, Vol. II, No. 1, 1950.

7. Hayashi, C., "On the Prediction of Phenomena from Qualitative Data and the Quantification of Qualitative Data from the Mathematicis-Statistical Point of View," Annals of the Institute of Statistical Mathematics, Vol. III, No. 2, 1952.

8. Hayashi, C., "Multidimensional Quantification," Annals of the Institute of Statistical Mathematics, Vol. V, No. 2, 1954.

9. Hayashi, C., "Examples of the theory of quantification," Proceedings of the Institute of Statistical Mathematics, Vol. 2, No. 1, 1954.

10. Hayashi, C., Methods of Quantification, Toyokeizai-Shinpo-Sha, Tokyo, 1975.

11. Horst, P., Wallin, P., Guttman, L., Wallin, F. B., Clausen, J.A., Reed, R. B., Richardson, M. W., Rosenthal, E., "The Prediction of Personal Adjustment," Social Science Research Council, 1941.

12. Ishida, M. D., "Fire Prediction," Research Memoir of the Institute of Statistical Mathematics, Vol. 5, 1949.

13. Krench, D. and Crutchfield, R. S., Theory and Problems of Social Psychology, McGraw Hill, 1948.

14. Mises, R. V., On the Classification of Observation Data in the Distinct Groups," Annals of Mathematical Statistics, Vol. 16, No. 1, 1945.

15. Ushijima, Y., Judging Method of Delinquent Juveniles in Early Stage, Kaneko Co., Tokyo, 1948.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my major professor, Dr. M. Siotani, for his guidance, inspiration and enthusiasm during the course of this study. Thanks are also due to committee members, Dr. A. D. Dayton and Dr. R. Nassar for their suggestions and discussions.

I wish to acknowledge the Department of Statistics for financial support during my graduate study.

Appreciations are accorded to my husband, Ruey-Hwa and my parents, Dr. and Mrs. C. J. Hsu, for their encouragement during my study of this report.

Finally, I want to express special thanks to Mrs. Endal who typed my report.

A STUDY OF QUANTIFICATION TECHNIQUES

by

BETTY POCHEN HSU WANG

B.S., Kansas State University, 1972

AN ABSTRACT OF A MASTER'S REPORT

Submitted in partial fulfillment of

the requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas
1978

ABSTRACT

Some theories and methods of quantification are presented and compared. Methods of quantification when the judgments are obtained by paired comparisons are considered for two cases: ordinary comparison and comparison of combination of items. Two cases of quantification methods are discussed when there is an outside criterion: a numerical criterion and a categorical criterion. Finally, methods of giving numerical values to types of persons and factors through their associations are discussed. Examples are presented which illustrate how to compute the desired numerical values.