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, <u>our principle calls for minimizing the variation within individuals compared with that within the group as a whole</u>.

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

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

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

$$F = \tfrac{1}{2}n(n-1) = \text{the total number of comparisons made by each person}$$
$$= \sum_k f_k^{(i)} = \sum_k g_k^{(i)} \qquad (2.5)$$

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

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

C = the total number of judgments in the experiment

$$= nc = Nn(n-1). \tag{2.7}$$

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 <u>between</u> individuals and the sum of squares W <u>within</u> individuals, let

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

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

and similarly let

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

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

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

$$V = \frac{1}{C} \sum_k x_k c = \frac{1}{n} \sum_k x_k. \tag{2.12}$$

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

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

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

$$\underset{\sim}{H}\underset{\sim}{x} = \eta^2 \underset{\sim}{x} \tag{2.23}$$

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

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

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

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

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 j*th* of which has m_j categories. Let O_{jx} $(k=1, \cdots, m_j; \ j=1, \cdots, n)$ be the α*th* category of the j*th* 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 j*th* and the k*th* 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:

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

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

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

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

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

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

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

= the total number of comparisons made by each person.

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

= 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 between individuals for the experiment and the sum of squares W within individuals. First of all, let

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

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

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

We again use the square of correlation ratio η

$$\eta^2 = \frac{B}{T} = 1 - \frac{W}{T} \qquad (2.39)$$

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:

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

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

$$\qquad\qquad m_1 \qquad m_2 \qquad \cdots \qquad m_n$$

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

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

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

$$
\underset{\sim}{F}_{kj} = \begin{pmatrix}
c_{jk|11} & c_{jk|21} & \cdots\cdot & c_{jk|m_j 1} \\
c_{jk|12} & c_{jk|22} & \cdots\cdot & c_{jk|m_j 2} \\
\cdots\cdot & \cdots\cdot & \cdots\cdot & \cdots\cdot \\
\cdots\cdot & \cdots\cdot & \cdots\cdot & \cdots\cdot \\
c_{jk|m_k} & c_{jk|2m_k} & \cdots\cdot & c_{jk|m_j m_k}
\end{pmatrix}
$$

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

$$
\begin{matrix} m_k & m_2 & \cdots\cdot & m_n \end{matrix}
$$

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 j<u>th</u> 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

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

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

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

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

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:

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

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

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

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

(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

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

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

Items	I_1				I_2				$\cdots$	I_M				Outside Variable Y
Variables	X_1				X_2				$\cdots$	X_M				
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}	
Values giving Persons	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}	
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.