STABILITY OF SIMILARITIES AND PREFERENCE OVER TIME

Junying Huang

I. Introduction

This study is an attempt to deal with one of the important and unsolved problems involved in the use of nonmetric multidimensional scaling solutions for predictive purposes, i.e., invariance of similarties/preference judgments over time.

Multidimensional scaling has been in existence for more than thirty years since the earliest work of Richardson[1] in 1938 based on the theoretical work of Young and Householder[2]. The major break-through in multidimensional scaling came in 1962 with the development of the nonmetric multidimensional scaling method by Shepard.[3]

Prior to the publication of Shepard's paper on nonmetric scaling, the existing procedures were either fully metric or fully nonmetric. Fully metric methods yield metric output, but they require metric input data to begin with. Fully nonmetric methods require only nonmetric input data to begin with, but they also yield nonmetric output. It was Shepard who developed nonmetric scaling which combines the best of both previous approaches, i.e., nonmetric input and metric output. Given only a rank of "psychological distance" data, the objective of nonmetric multidimensional scaling is to find a configuration whose rank order of ratio-scaled distances best reproduces the original input ranks.

Nonmetric multidimensional scaling is a mathematical and geometrical methodology that attempts to graphically represent a set of data with respect to similarities and preference. It provides a configuration of nonmetric data in a specific number of dimensions, with measures of distortion (stress) for each number of dimensions. The axes of the configuration are referred to as dimensions of the data, whereas the geometric distance is interpreted as the degree of similarity.

Since Shepard's work was published in 1962, progress in algorithm development has been rapid. Several useful refinements and extensions have been made. Substantial efforts have been made to examine the utility of nonmetric scaling in a variety of areas. A number of studies have been reported applying nonmetric scaling to the measurement of perceptions and preference of many stimuli. The results of these empirical studies seem to support the suggestion that nonmetric scaling methods can provide us with valuable information concerning such subjective mental phenomena as perception and preference.

The techniques of nonmetric multidimensional scaling require as input data a rank order of similarities/preference judgments. "Similarities" refers to perceived distances between stimuli, and "preference" refers to perceived distances of stimuli from an ideal point. A problem of interest in applying nonmetric scaling to prediction decisions concerns the effect of change in time on individual's similarities and preference judgments. If the individual's perceptions and preference judgments oscillate from time to time, then the optimal strategies worked out at the planning stage may no longer be the optimal ones when they are implemented since there is always a time gap between planning and execution. The similarities/and preference judgments must be relatively stable over change in time if we are ever to be able to use the nonmetric scaling solutions for various predictive purposes. However, little is known about the invariance of similarities/preference judgments over time. The purpose of this study is therefore to examine the stability problem. Two issues to be covered are:

(1) Stability of similarities and preference judgments over time.

(2) The effect of the length of time gap on stability of similarities/preference judgments.

II. Experimental Design

1. Subjects:

A laboratory experiment with sixty randomly selected students of a U. S. university as subjects was designed and implemented. The sixty students were randomly assigned to three experimental groups of equal size.

2. Stimulus and dimension sets:

The stimulus set includes six "imaginary" brands of soft drink. These imaginary brands were presented to each subject graphically in terms of their scale values on two prespecified dimensions (calorie and cola flavor) on cards.

In order to disguise the purpose of this experiment, the six imaginary brands were identified as A, B, C, D, E and F at the first session (t_0) and as G, H, I, J, K and L at the second session (t_1). The subjects were led to believe that they were asked to judge twelve different "real" brands of soft drink in both sessions.

3. Time intervals:

Three time intervals (2 days, 5 days and 22 days) were randomly applied to different subject groups. They were randomly selected from a treatment pool of one to thirty days.

4. Cover story:

The purpose of the experiment was disguised as a comparison of perceptions of soft drinks between different sex and age groups and between students and housewives. For disguise purpose some demographic data were collected including sex and age.

5. Data collection:

Similarities ranks were collected by a multi-stage card-sorting procedure. Preference orderings were collected by asking the subject to sort the six cards (brands) in terms of their preference. To overcome the possible effect of card position in the deck, each card occupied the same position in the deck at all time for all subjects.

III. Analysis of Results

1. Stability of Similarities Over Time

The degree of association between ranks at t_o and t_l are measured using Kendall's tau cofficient[4],

$$tau = 1 - \frac{2(\text{number of inversions})}{\text{number of pairs of objects}} \qquad [1]$$

When N(number of observation) is larger than 10, Kendall's tau can be considered to be normally distributed. Therefore, the test of significance for tau statistics is given by $Z=(tau/\sigma_{tau})$. When N is less than or equal to 10, Kendall's tau may not be considered as a normal distribution. Thus, different procedure of test of significance for tau is followed.[5]

For each of the 60 subjects the Kendall's tau coefficient between similarities ranks at t_o and t_l was computed. The result indicates that for 59 (98.33%) out of 60 subjects their inter-stimulus simularities ranks at t_o and t_l were not independent of each other at .05 level of significance. The results on the individual level was also used to derive the over-all patterns of results over all subjects. The results reject the hypothesis that similarities judgments between t_o and t_l are certainly unstable on the aggregate level at at least .0001 level of significance.

The average observed tau value for all 60 subjects is .79174 and the 95% confidence interval for tau in population is estimated as between .74305 and .84043.

Defining "relative stability" as tau value equal to or greater than .75, we found that 45 subjects (75%) were relatively stable over time in their similarities judgements.

It should be noted that the results concerning the stability of similarities over time are limited to the relatively short time span employed in this study (two, five and 22 days). The individual's perceptions of the stimuli are partly determined by the so-called "stimulus factors", i. e., the nature of the physical stimulus itself. However, perceptions may change over time even though the stimulus itself remains the same. As Kassarjian and Robertson note, numerous studies clearly indicate that perception of realitv is in part determined by the "individual's needs, drives and past experiences; by what he had learned; by his motives and personality; and by his social and geographic environment."[6] As we know, these so-called "personal factors" are subject to such environmental influences as culture, income, family, reference group, social class, physical condition and many others, and their dynamic interaction. While some of these influences are relatively stable over time, the others may fluctuate from time to time. Consequently, the "personal factors" of perception are likely to change with time. The longer the time span is, the more likely the "personal factors", and then the perceptions, are to change. Thus, it seems logical to expect longer time intervals then employed in this study to make a difference in the similarities judgments. Though we don't know exactly how much longer the time span must be, we may expect that the longer the time interval, the less stable the similarities judgments will be. We will deal with this aspect of the problem later.

2. Stability of Preference Over Time

Again, Kendall's tau coefficient was used as a measure of the degree of association between preference judgments at t_o and t_i. The result shows that for 43 (71.67%) out of 60 subjects the preference orderings at t_o and t_i are not mutually independent of each other at .05 level. The over-all patterns of results also reject the hypothesis that preference judgments between t_o and t_i are certainly unstable at at least .0001 level. The average observed tau value for all subjects is .77777. Also, it was found that we are 95 percent confident that the interval [.68798, .86756] includes tau for population.

As mentioned earlier, tau values of .75 or larger are viewed as an indication of "relative stability" by our definition. Thirty three subjects (55%) were found relatively stable in their preference judgments over time.

The result concerning the stability of preference judgments over time is, however, limited to the relatively short time span employed. Individual's preference for the stimuli is partly determined by his needs, drive, motive and many other "personal factors". As time goes by, these "personal factors" are likely to change, which in turn may affect the preference. Thus we expect that the longer the time span, the less stable are the individual's preference judgments. As will be shown later, our expectation concerning time interval vs. stability of preference seems to be justified.

3. Length of Time Interval vs. Stability of Similarities

As mentioned earlier, the subjects were randomly divided into three groups of 20 subjects each. Each subject group is subject to one of the three treatments: 2-day gap (Group I), 5-day gap (Group II) and 22-day gap (Group III). In the case of similarities judgment, the average number of inversions is 9.65 for Group I, 10.05 for Group II and 13.10 for Group III. As shown in equation [1], there is an inverse relationship between tau value and number of inversions. The tau value seemed the largest for Group I, followed by Group II and Group III. In other words, the similarities judgments appeared to be the most stable for Group I, followed by Group II and Group III as would be expected. A simple randomized analysis of variance was conducted to see whether or not there would be significant differences among these three average numbers of inversions. On the basis of the result of F-ratio test, we concluded that there was no significant differences among similarities judgments for the three subject groups at .05 level. In other words, for the relatively short time spans employed, the length of time interval did not make a significant difference so far as the stbility of similarities judgments is concerned.

4. Length of Time Interval vs. Stability of Preference

In the case of preference judgments the average number of inversions is 1.10 for Group I, 1.55 for Group II and 2.35 for Group III. The preference orderings seemed most stable for Group I, followed by Group II and Group III as would be expected. The result of F-ratio test indicates that the over-all F-ratio is significant at .05 level.

Since we have found evidence for over-all significance among the three experimental groups, the next logical step was to evaluate comparisons among means via the use of Sheffe method.[7] The Scheffe method is a device for testing the significance of post-hoc comparisions. The Scheffe method is chosen over the other methods because of its simplicity and relative insensitivity to departures from normality and homogeneity of variance.

A pairwise comparison of mean numbers of inversions (preference) was made. The pairwise differences between mean numbers of inversions are provided in Table 1. In Table 1, one of the absolute differences (between Groups I and III) is greater than the required value of 1.2158 and this pairwise comparison is significant at .05 level.[8] We

Table 1: Pairwise Differences in Number of Inversions

mean / mean	Group II 1.55	Group III 2.35
Group I 1.10	-0.45	-1.25*
Group II 1.55		-0.80

* significant at .05 level.

could say that the difference in mean number of inversions between Groups I and III contributes to the over-all significance of F-ratio. In other words, the preference judgments for those subjects in Group III who were subject to a 22-day time gap were on the average less stable than the preference judgments for those subjects in Group I who were subject to a 2-day time gap. As to the preference judgments between Groups I and II and between Group II and Group III, no significant differences in the degree of stability were found.

IV. Summary

The study focuses on two issues concerning (1) stability of similarities and preference judgments over time, and (2) effect of time interval on the stability of similarities and preference judgments. It was found that

1. For most of the subjects, their similarities and preference judgments were "relatively stable" over a relatively short time span (2 to 22 days). Kendall's tau values of .75 or larger are viewed as "relatively stable" by our definition.

2. Three different treatments (2-, 5-, and 22-day intervals) were randomly applied to three experimental groups. No significant difference in the degree of stability of similarities judgments were found among groups. However, the preference judgments of subjects given a 22-day time gap were found less stable than those given a 2-day time interval.

It should be kept in mind, when drawing any inference from the findings, that the study is subject to two major limitations arising from the relatively short time spans involved and the type of subjects employed in the experiment.

Appendix: Comparison by Sheffe method

Given any comparison g made on the data after a significant F has been found for the relevant factor, the significance of the comparison value $\hat{\psi}_g$ may be found by use of the following confidence interval:

$$\hat{\psi}_g - S\sqrt{V(\hat{\psi}_g)} \leqq \psi_g \leqq \hat{\psi}_g + S\sqrt{V(\hat{\psi}_g)}$$

where $\sqrt{V(\hat{\psi}_g)} = \sqrt{(MS_W)W_g}$

$S = \sqrt{(J-1)F_\alpha}$

$W_g = \sum(c_j^2/n_j)$

c_j: Weights

n_j: Number of subjects in group j

F_α: The value required for significance at α level, with J-1 and N-J degrees of freedom.

For any α, this gives the 100 $(1-\alpha)$ per cent confidence interval for ψ_g, the true value of the comparison. When the confidence interval fails to cover zero, the comparison is said to be significant, and identified as one possible contributor to the over-all significance of F. For this interval to exclude zero for any of the differences, the obtained difference would have to be greater than $S\sqrt{V(\hat{\psi}_g)}$ in absolute magnitude.

In this case,

$$W_g = (1/20)(1+1) = .1$$

$$\sqrt{V(\hat{\psi}_g)} = \sqrt{2.33859649 \times .1} = .4836$$

$$S\sqrt{V(\hat{\psi}_g)} = \sqrt{(3-1) \times 3.16} \times .4836 = 1.2158$$

REFERENCES

* The author would like to thank Professor David J. Curry of the University of Iowa, U.S.A., for his assistance in the research for this study.

1. M. W. Richardson, "Multidimensional Psychophysics", Psychological Bulletin, 35 (1938) pp. 659-660.

2. G. Young and A. Householder, "Discussion of a Set of Point in Terms of Their Mutual Distances," Psychometrika, 3 (1938), pp. 19-22.

3. R. N. Shepard, "The Analysis of Proximities: Multidimensional Scaling with An Unknown Distance Function, I & II," Psychometrika, 27 (June, 1962), pp. 125-139; and 27 (Sept. 1962) pp. 219-246.

4. W. L. Hay and R. L. Winkler, Statistics: Probability, Inference and Decision (New York: Holt, Rinehart and Winston, Inc., 1970), Vol. II, Chapter 12.

5. For a detailed description of this procedure, please refer to S. Siegel, Nonparametric Statistics for the Behavioral Science (New York: McGraw-Hill Co., 1956), chapter 9.

6. H. H. Kassarjian and T. S. Robertson. (eds.), Perspectives in Consumer Behavior (Glenview, Ill.,: Scott and Foresman, 1973), P. 2.

7. For a detailed description of Schffe method, please refer to W. L. Hays, Statistics (Holt, Rinehart and Winston, Inc., 1963), chapter 14.

8. See Appendix for an explanation of this comparison by Sheffe method.