ED 356 251 TM 019 675

AUTHOR SenGupta, Saumitra
TITLE Detecting Point Diffusion Patterns in Two-Dimensional Maps: A Simulation Study.
SPONS AGENCY National Inst. of Mental Health (DHHS), Bethesda, Md.
PUB DATE Nov 92
CONTRACT R01MH46712-01A1
NOTE 12p.; Paper presented at the Annual Meeting of the American Evaluation Association (Seattle, WA, November 5-7, 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Simulation; *Maps; Matrices; *Multidimensional Scaling; *Profiles; Qualitative Research; *Scores
IDENTIFIERS *Graphic Representation; Pattern Matching; *Point Diffusion Patterns; Similarity (Concept)

ABSTRACT

A way of identifying non-random patterns of effects on a group of individuals as a result of some intervention when a sample of participants is arrayed according to some indices of similarity is presented. The principle of proximal similarity and the concept of pattern matching provide the background for this effort. Major advantages are the treatment of an individual as a unit of analysis, visual representation of the outcome across individuals, and the ability to use the representation in pattern matching. A non-metric multidimensional scaling method is used to generate a person map in which each individual is represented by an individual point (profile analysis). From the developed rectangular profile matrix a square matrix can be constructed to be used as the input for multidimensional scaling. Once the person map is generated with the overlaying of gain scores, the next step is to determine the presence of prototypical individuals on the map. A simulated data set representing the ideal pattern is compared to the map to determine point diffusion patterns. Although more work is needed to establish this method as an alternative to conventional statistical methods, it appears to provide a practical blend of qualitative thinking and quantitative methods. Eight figures illustrate the process. (SLD)

Detecting Point Diffusion Patterns in Two-Dimensional Maps: A Simulation Study

Saumitra SenGupta
Department of Human Service Studies
N132 MVR Hall
Cornell University

Paper presented at the Annual Conference of the American Evaluation Association
Seattle, WA
November 5 - 7, 1992

Intr⁄duction

The purpose of this paper is to identify non-random pattern of effects on a group of individuals as a result of some intervention, if such exists, when a sample of the participating individuals is arrayed according to some indices of similarity. The arraying is done here in a visually comprehensible manner by using multidimensional scaling. Two key concepts, namely, principle of proximal similarity (Campbell, 1986) and pattern matching (Trochim, 1989; Campbell, 1991) provide the background for such an endeavor. The principle of proximal similarity states that a given program should affect similar constructs or persons more similarly than dissimilar ones (Campbell, 1986; Trochim, 1989). The concept of pattern matching states that if there exists a theoretical pattern of expected outcomes, then a matching pattern in observed outcomes provides a strong evidence of internal and construct validity. The more detailed and well laid out the patterns are, the stronger such evidence is if a match exists. The major advantages of this approach include: a) treatment of an individual as the unit of analysis as opposed to most commonly used statistical models where the outcome is typically averaged across individuals, b) the visual representation of outcome across individuals which makes it easier to detect whether principle of proximal similarity is obeyed or not, and c) option for use in pattern matching.

The idea behind this approach is to identify a prototype of individuals who might benefit from a similar program in future. For instance, in the case of a clinical drug trial the researcher may be interested in identifying individual patients who benefited the most from the drug trial and from the demographic and pretest profile of such individuals be able to predict what type of individuals may benefit the most from future application of the trial medicine. Similar applications may be made in evaluative efforts in other fields such as education.

The method presented in this paper can be divided into two distinct steps. First, the generation of patterns or, as will be seen later, the process of person mapping (SenGupta, 1991), and second, devising a technique to identify such patterns. While there are several types of patterns that can exist, the one which is commensurate with the stated purpose of this paper is what will be termed as "point diffusion pattern". In the ideal case of a point diffusion pattern there will be a single individual who would benefit the most from a

---

This research was supported in part by NIMH grant no. RO1MH46712-01A1, William M.K. Trochim, Principal Investigator. The author wishes to thank Prof. William M.K. Trochim for his help.

given program or intervention. There will be a regular gradient along which the other individuals' indices of benefit or gain scores will gradually decline. In real life of course, such a pattern is unlikely to occur. Instead, it is likely that there will be several individuals who will benefit the most and they will not have identical demographic or pretest profiles. However, the aim of this paper is to devise a way in which point diffusion pattern, if such exist, can be detected. In order to do that a simulated data set will be used, which approximate the ideal situation of point diffusion pattern.

Person Mapping

A non-metric multidimensional scaling method is used to generate a person map where each individual is represented by individual points. Shepard (1972) terms this kind of multidimensional scaling as profile analysis. Typically, in profile analysis a rectangular matrix is set up where the columns refer to the individuals to be scaled and the rows refer to the pretest characteristics of the individuals. In such a matrix each cell would contain the pretest score of an individual on a specific test. More specifically, a cell (i,j) situated at the intersection of i-th row and j-th column would contain the pretest score of the j-th person on the i-th test. It may be noted here that some or all of the pretests can actually be some demographic characteristics depending on the aim of the researcher and provided some numerical indices can be constructed of such characteristics. SenGupta (1992) discusses how some categorical data such as race, religion etc. can be suitably represented for such purpose. However, Shepard (1972) kept his discussion limited to test scores obtained from psychological and educational tests.

From this rectangular profile matrix a square matrix can be constructed where both rows and columns refer to the individuals and each cell contains an index of similarity (or dissimilarity) between the corresponding pair of individuals. While various indices can be constructed for this purpose, the commonly used ones are Euclidean distances (index of dissimilarity) and correlation coefficients (index of similarity). In either case, the computation involves comparing each pair of columns (or individuals) from the original profile matrix.

The square matrix is then used as the input for multidimensional scaling. Figure 1 provides an example of the output from profile analysis of one hundred individuals based upon a combination of pretest scores from twenty two diagnostic tests for anxiety related disorders and four demographic characteristics. Figure 2 contains the same map after overlaying the gain scores (difference of posttest and pretest scores) for each individual in one of the diagnostic tests (Clinician Rated Anxiety Scale).

Figure 1 - Map of 100 individuals after MDS

Figure 2 - CRAS gain-scores overlaid on the map of 100

Detection of Point Diffusion Patterns

Once the person map is generated complete with the overlaying of gain scores the next step is to devise a technique that will detect the presence of proto-typical individual(s) on the map. As mentioned earlier, in the ideal case of a point diffusion pattern one individual with the highest gain score will constitute a focal point on the map and the gain scores of other individuals will decline in a gradient from this focal point (referred to as <u>point peak pattern</u> to distinguish from other point diffusion patterns presented later). The final map (figure 3) is taken from real life data and as can be seen, there is no apparent pattern on this map. Therefore, any such technique to detect point diffusion pattern can not be applied here unless the technique is tested on an ideal pattern and determined to be functioning properly. For this purpose a simulated data set representing the ideal pattern is used here. Figure 3 represents the map generated from simulated data.

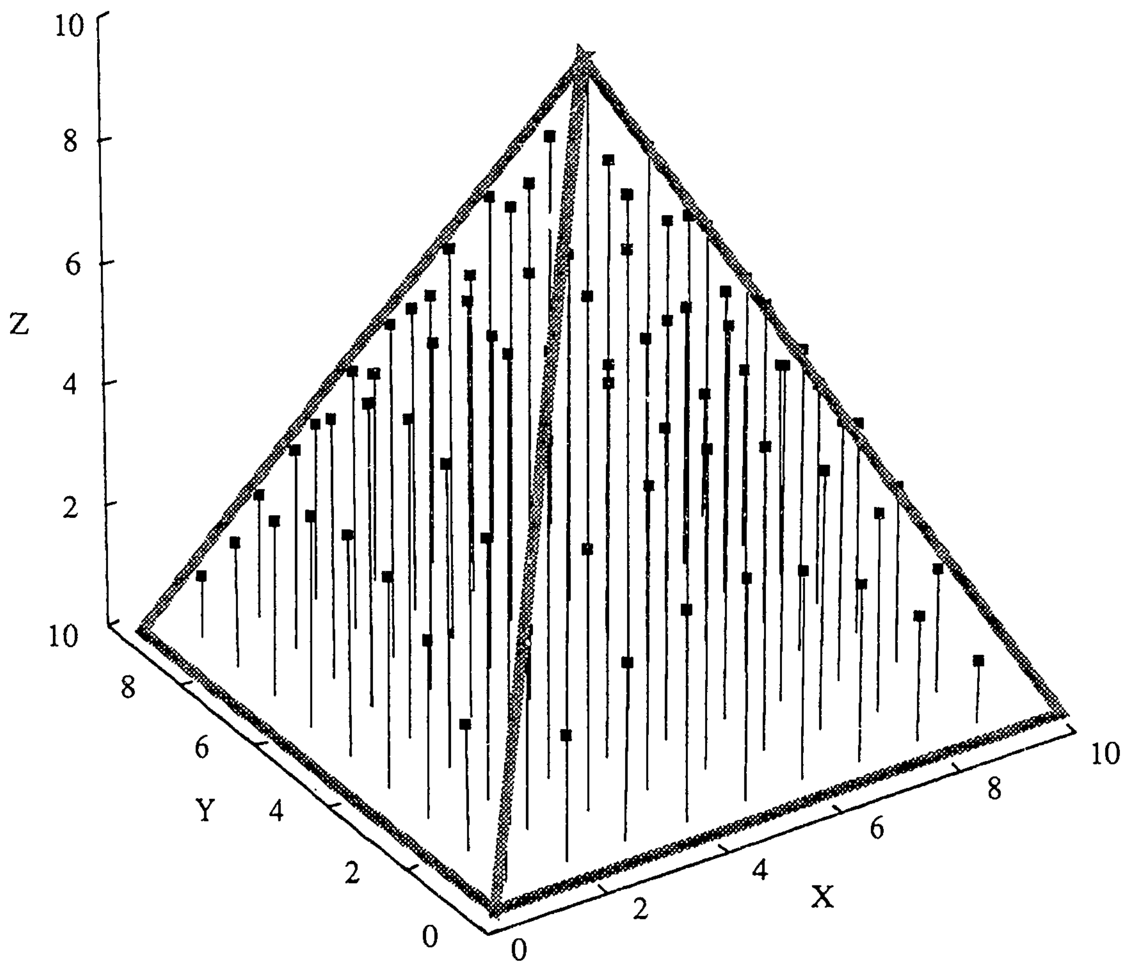

Fig. 3 - Simulated Point Peak Pattern

A Total of eighty one points are used here in a 9 by 9 square grid. The hypothetical individual in the center of this map is the one with the highest gain score while the gain scores of other individuals declines in a regular gradient on all sides creating a pyramid shaped map of individuals. A correlational approach is used here to detect the existence of the focal point representing the proto-typical individual. In this approach the difference between the gain scores (in this case the difference in height of spikes) between each pair of individuals is correlated with the Euclidean distance between the points representing the individuals. In other words, for each individual a set of two columns containing the difference in gain scores with every other individual and the distance from every other individual are set up and the correlation for each set of two columns are then computed.

A high positive correlation for any individual from such computation would indicate the existence of a point diffusion pattern in the map with that individual being situated either at or very close to the focal point. In descriptive terms this would mean that for the individual situated at the focal point, the difference in gain score with any other individual would be hi ier for longer distance from that individual and *vice versa*. The result from correlation analysis for the simulated data set is presented in figure 4.

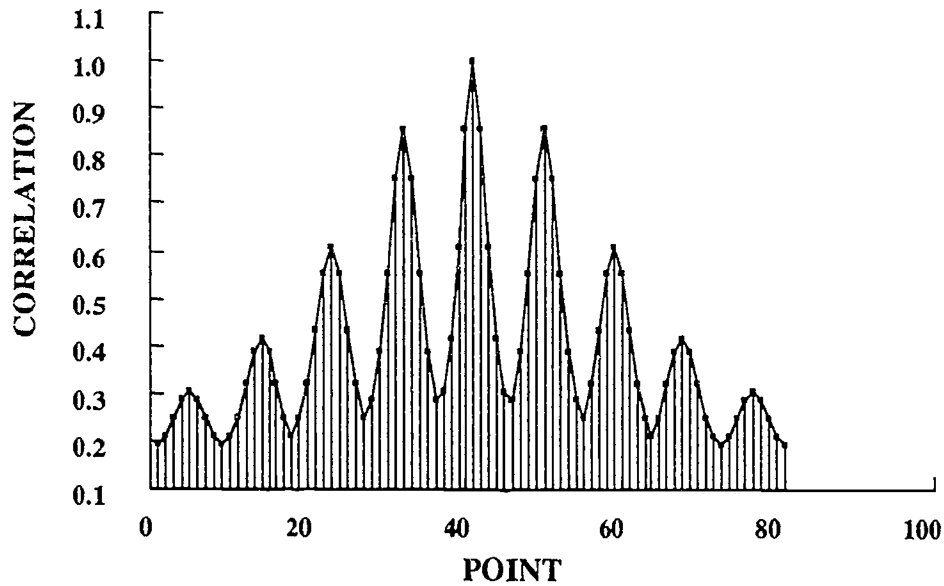

Figure 4 - Detection of Point Peak Pattern

As can be seen from the graph in figure 4 the correlation is the highest for point number 41 which is the one at the center of the grid containing eighty one points. This shows that the correlational approach is able to identify the focal point or the individual at the focal point in the ideal single point peak pattern.

Two other types of patterns are shown in figures 5 and 7 which can be included in the category of point diffusion patterns. The pattern in figure 5 shows a linear planar pattern where individuals at two opposite ends of the planes have the highest and the lowest gain scores respectively. The aim here is to detect the orientation of the plane with respect to the XY surface. In order to detect the orientation, a multiple regression containing Z as the outcome and X and Y as the independent variables is set up for each angle around the origin along XY plane. The proper angle is indicated by the equation for which the t-values of the regression coefficients are maximum. The results are shown in figure 6. The angle of the plane where the highest t-values occur is 45° which is how the simulated data was set up.

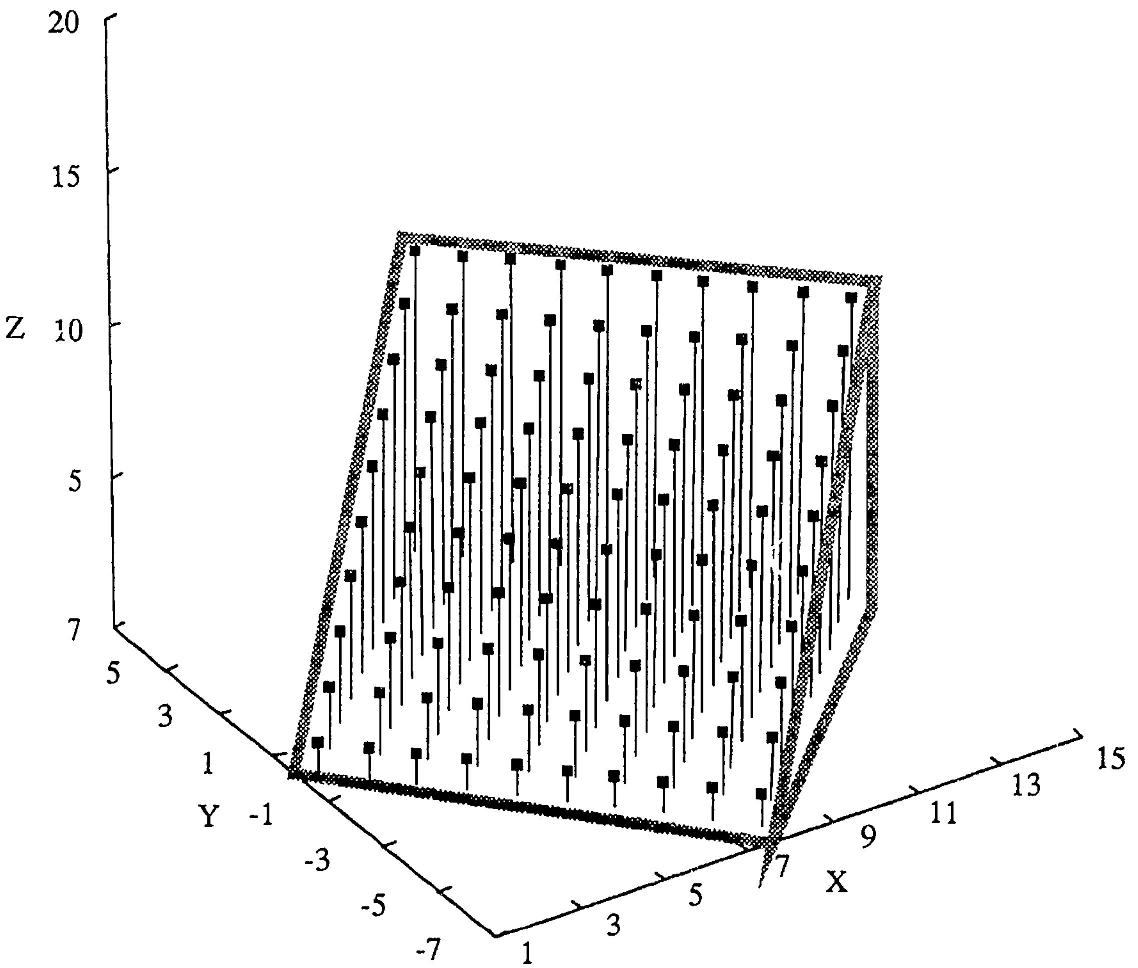

Figure 5 - Simulated Linear Planar Pattern

Figure 7 shows a circular pattern where the gain scores start declining from a group of individual in a circular fashion and reaches the lowest point at almost the same region of the map. The principle of proximal similarity in this pattern is obeyed along a circle than more easily understandable patterns presented earlier. In case of a circular pattern the location of the radial straight line which divides the highest and the lowest gain scores is important. In this case the gain score is regressed on the angle of projection of the corresponding point on either the X or the Y axis. Again, the dividing line is detected at the points for which the t-values of the regression coefficients are maximum. Figure 8 shows that such a line exists at 45° angle which conforms to the simulation parameters.

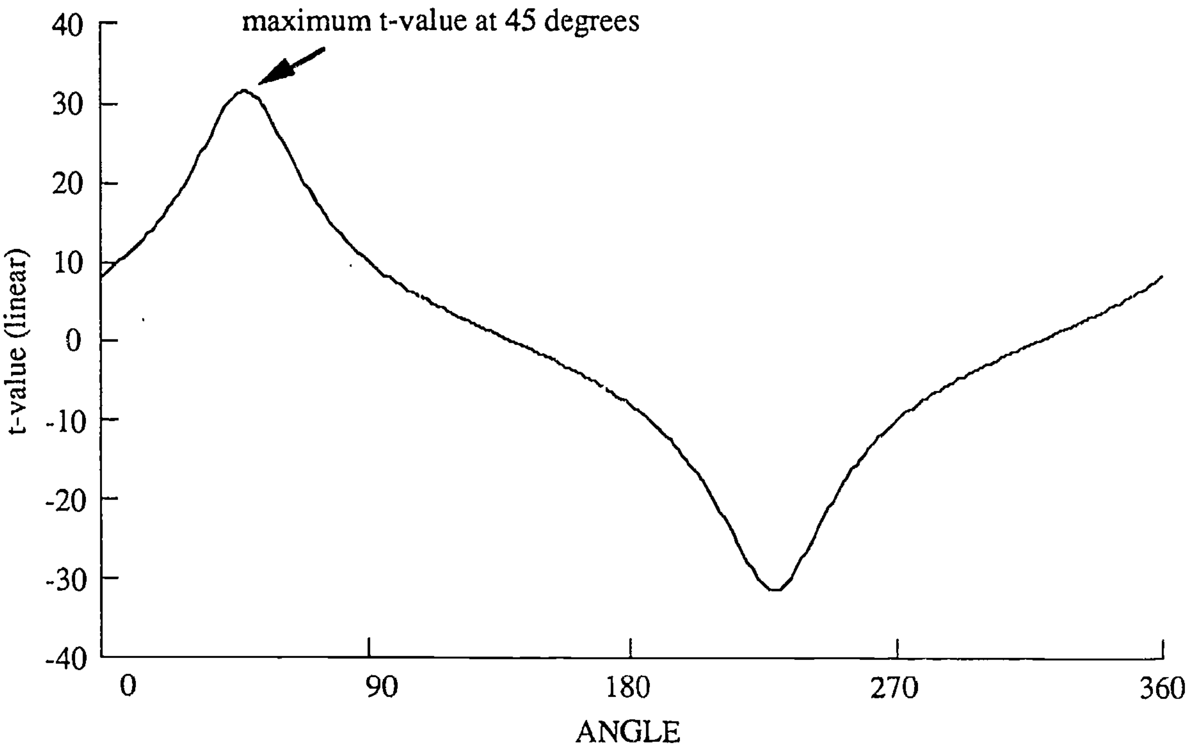

Figure 6 - Detection of Linear Planar Pattern

Conclusion

Ability to detect non-random patterns in either a person map or a concept map helps the researcher both methodologically and in terms of predictability. It may be noted here that the simulated data set and the methods of pattern detection can be used in concept maps in a similar manner. In the case of concept maps the points will represent different concepts, constructs, or issues as the case may be instead of representing individuals.

Methodologically, the existence of any point diffusion pattern implies the presence of proximal similarity. Since Campbell (1986) has relabeled external validity as principle of proximal similarity, the method of detecting point diffusion patterns may strengthen the validity of research findings provided such maps as in figure 2 can be produced and patterns are detected in such maps.

In terms of predictability, the detection of point diffusion patterns enable the researcher to be more specific about the kind of individuals who are more likely to benefit from similar programs in future or about the areas (concepts) in which similar programs are likely to produce greater impact. While more conventional statistical methods such as factorial designs suit this purpose for few variables, for a large number of variables such methods may become computationally more difficult accompanied by rapid loss of statistical power with increasing number of cells.

A lot more work needs to be done with real life data sets for establishing this method as a

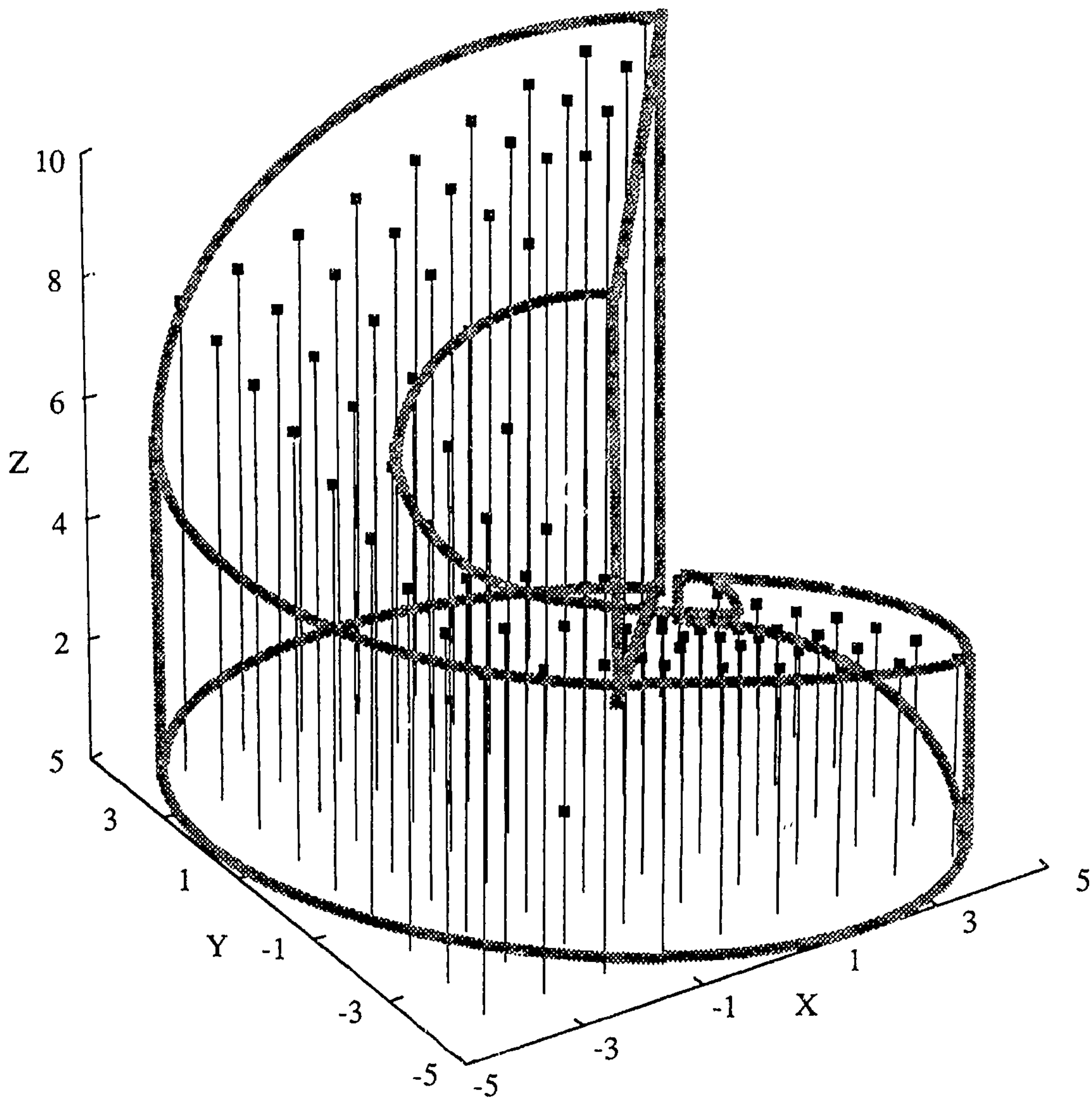

Figure 7 - Simulated Linear Circular Pattern

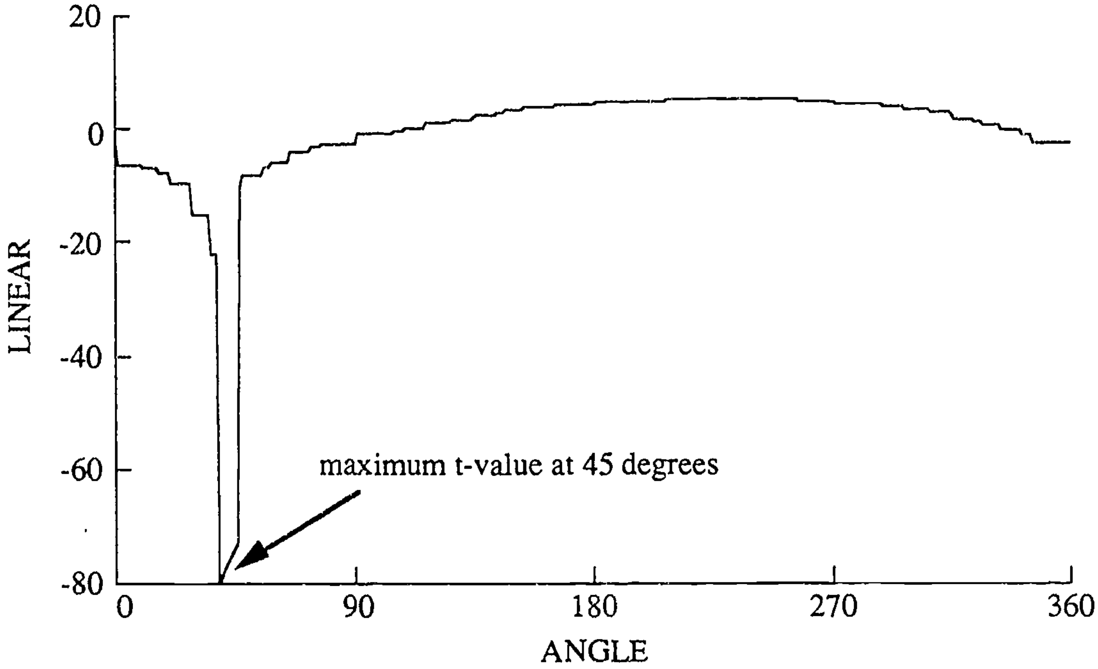

Figure 8 - Detection of Linear Circular Pattern

viable alternative to conventional statistical methods. However, this is a modest beginning to convince and encourage researchers in evaluation focused studies to experiment with such methods which, in this author's opinion, produce a practicable blend of qualitative thinking and quantitative methods.

Reference

Campbell, D. T. (1986) Relabeling internal and external validity for applied social scientists. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 67-77). San Francisco: Jossey-Bass Inc.

Campbell, D. T. (1991) Coherentist empiricism, hermeneutics, and the commensurability of paradigms. *International Journal of Educational Research, 15*(1), 587 - 597

SenGupta, S (1992) *Using multidimensional scaling to detect patterns across individuals in program evaluation.* Unpublished Master's thesis, Cornell University

Shepard, R. N. (1972) A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.). *Multidimensional scaling* (vol.1, pp. 21-47). New York: Seminar Press.

Trochim, W. M. K. (1989) Outcome pattern matching and program theory. *Evaluation and Program Planning, 12*, 355-366.