Kamen's quick clustering of participation data — a research note

L. Pitt
Department of Marketing, School of Business, Western Australian College, Perth, Western Australia

D. Nel*
Department of Marketing, University of Pretoria, Hillcrest, Pretoria, 0002 Republic of South Africa

This brief research note revisits a simple, but very useful clustering procedure developed by Kamen (1970), and illustrates its use in the clustering of attitudinal/perceptual data. For the purpose of illustrating the technique, perceptions of peer group participation in potentially corrupt situations in business were used as the data set. The mean responses, standard deviations and medians of 458 managers served as input for a correlation matrix from which the variables were clustered. The clusters formed by the analysis have been interpreted as 'The Insiders', 'Felons', 'Happy Holidays', 'The Fiddlers', 'A Bit on the Side', and 'The Innocents'. From the clusters identified it was evident that some situations were similar in nature. Quick clustering of the pilot study data is regarded as successful and could therefore lead to improved questionnaire design as well as the elimination of similar questions.

Hierdie kort navorsingsnota beskou weer eens 'n eenvoudige, maar baie handige trosanalise-prosedure ontwikkel deur Kamen (1970) en illustreer die toepassing van trosanalise met houdings/persepsie-data. Ten einde die tegniek te illustreer, is persepsies van portuurgroepdeelname in moontlike korrupte besigheidspraktyke gebruik as datastel. Gemiddeldes, standaardafwykings en mediane is vir 458 bestuurders bereken, wat gedien het as insette om 'n korrelasiematriks te ontwikkel waaruit die veranderlikes in trosse gegroepeer is. Die trosse wat na vore gekom het uit die analise is geïnterpreteer as 'Die Binnekringers', 'Oortreders', 'Heerlike Vakansies', 'Die Knoeiers', 'Ietwat eenkant' en 'Die Onskuldiges'. Uit die geïdentifiseerde trosse was dit duidelik dat sekere situasies eenders van aard was. Die onderwerping van die loodsstudiedata aan 'n spoedige trosanalise-prosedure word as suksesvol beskou en kan aanleiding gee tot verbeterde vraelysontwerp, of om onnodige duplisering van soortgelyke vrae uit te skakel.

* To whom correspondence should be addressed

Introduction

In many application settings there is reason to believe that the set of objects under study can be clustered into subgroups that differ in meaningful ways. Researchers and decision makers often need to segment variables such as objects, ideas and attitudes into groups, in such a way that there is maximum with-in group homogeneity. One of the best-known techniques for doing this is cluster analysis. Cluster analysis is a set of statistical techniques for grouping objects into homogenous sets on the basis of observed data (Wentz, 1979: 435). The objective in most cluster applications is to arrive at clusters of objects that display intra-cluster variation relative to the inter-cluster variation (Dillon & Goldstein, 1984: Ch 5). It differs from other methods of classification, such as discriminant analysis, in that the number and characteristics of the groups are to be derived from the data and are not usually known prior to the analysis (Afifi & Clark, 1984: 379).

The most patent advantage of the technique(s) is that it allows the researcher to start with nothing but the attributes of the variable that need to be classified, and the attributes are allowed to determine which variables will be grouped together, or 'clustered'. Several different algorithms are available for conducting cluster analyses — Punj & Stewart (1983) have described 14 different clustering programmes/packages. Sokel & Sneath (1963) provide a detailed examination of clustering approaches, whilst Green & Tull (1978) offer an excellent general discussion of clustering techniques. Depending on the choice of method, nominal, interval, or ratio data may be analysed.

Although cluster analysis is most commonly used to group like objects (such as people, products, cities) in preliminary examination of data, researchers often want to examine the inter-relationships between variables which are not strictly 'objects' — attitudes, perceptions, rankings, etc. Psychographic studies which use banks of statements from Likert-type scales, for example, might be expedited by the facility to group similar statements for the purposes of preliminary data analysis — to get an overall 'picture' of the data. 'In short, it is possible to find applications of cluster analysis in virtually any field of research' (Afifi & Clark, 1984: 381). Frequently however, the cluster analysis procedures available will entail too much additional time and effort, both from the point of view of data processing, and with regard to complexity. This brief research note revisits a simple, but very useful clustering procedure developed by Kamen (1970), and illustrates its use in the clustering of attitudinal/perceptual data.

The data set

The data set used for this illustration of Kamen's Quick Clustering procedure is that of Pitt & Nel (1988), suggested by a pilot study of Pitt (1985). In order to measure the attitudes of a group of managers towards situations in business which could be construed as

potentially corrupt, respondents are presented with a series of scenarios which describe certain situations. They are then asked to indicate how wrong they feel each to be on a scale ranging from 1 (definitely wrong), through 2 (wrong), and 3 (understandable) to 4 (not wrong at all). Using the principle of measuring perceived participation (Wilkes, 1978), respondents are required to indicate the extent to which they are of the opinion that their friends/colleagues would behave in the manner described, ranging from 1 (never), through 2 (very seldom), and 3 (often) to 4 (most of the time). Finally, respondents are asked to indicate what measure(s) the organization, having become aware of the situation described, should adopt. The scenarios used in Pitt (1985) and Pitt & Nel (1988) are presented in Appendix 1.

For the purpose of illustrating Kamen's Quick Clustering technique, this research note analyses the respondents' perceptions of peer group participation in the activities presented in Appendix 1. In Table 1 (Pitt & Nel, 1988) the degree of perceived peer group participation in the scenarios in Appendix 1 is presented, as well as standard deviations and medians. (Pitt, 1985 and Pitt & Nel, 1988 also present analyses of the attitudes towards how wrong the various situations are perceived to be, and the appropriate action of the company.) Obvious questions which arise are: Are there perhaps groups, or clusters of situations, which are similar in nature, and with regard to the extent of participation? Would it be possible to do similar research without having to rely on the opportunity for engaging in them across various organizational position spectra (Ferrell & Gresham, 1985), without having to include a number of scenarios which score low on participation simply because for most respondents, the opportunity is seldom there?

Kamen's quick clustering procedure

Before starting Kamen's (1970) clustering procedure it is necessary to develop a correlation matrix for all the variables (in this case the scenarios and the degree of perceived participation) to be clustered. This has been done in Table 2.

In quick clustering, clusters are formed by the following steps:

1. Underline the highest entry in each column — as done in Table 2.

Table 1 Extent of perceived peer group participation. (Question to managers: How often would your friends/colleagues behave as described. 1 = never; 4 = most of the time; N = 458.)

Scenario	Mean	SD	Median
1. Bribe	1,509	0,848	1
2. Hunting trip	2,382	0,924	2
3. Free lunch	3,264	0,830	3
4. Competitor information	1,456	0,774	1
5. Christmas gift	3,552	0,721	4
6. Sporting event	3,157	0,798	3
7. Large gifts	1,716	0,861	1
8. Geologist knowledge	2,448	1,011	2
9. Free weekend	2,153	0,925	2
10. Fuel siphoning	1,755	0,861	2
11. False expense claims	2,314	0,927	2
12. Free overseas trip	2,236	0,952	2
13. Insider trading	2,245	1,017	2
14. Free groceries	1,814	0,909	2
15. Secretary	2,155	0,959	2

SD — Standard Deviation

(Source: Pitt & Nel, 1988)

Table 2 Correlations among scenarios in terms of perceived peer group participation

Scenario numbers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.	—	0,4524	0,1568	0,6226	-0,0348	0,1371	0,5487	0,2622	0,3247	0,4737	0,4225	0,2901	0,3046	0,4155	0,3335
2.	0,4525	—	0,4015	0,3891	0,2246	0,4409	0,4582	0,4279	0,5355	0,2608	0,3116	0,4221	0,3778	0,4157	0,3554
3.	0,1568	0,4015	—	0,1423	0,4285	0,4196	0,2305	0,2889	0,3176	0,1518	0,1960	0,3334	0,2265	0,2276	0,2591
4.	0,6226	0,3891	0,1423	—	0,0842	0,1175	0,6280	0,2781	0,4311	0,4767	0,3975	0,3109	0,3084	0,4756	0,3764
5.	-0,0348	0,2246	0,4285	-0,0842	—	0,4198	0,0205	0,1975	0,1915	-0,0604	0,0932	0,1350-	0,0989	0,0801	0,0563
6.	0,1371	0,4409	0,4196	0,1175	0,4197	—	0,2307	0,2572	0,2787	0,0784	0,2231	0,2795	0,1953	0,2336	0,2199
7.	0,5487	0,4582	0,2305	0,6280	0,0205	0,2307	—	0,4250	0,4965	0,4580	0,3641	0,3992	0,4017	0,5924	0,4270
8.	0,2622	0,4279	0,2889	0,2781	0,1975	0,2572	0,4250	—	0,4436	0,3322	0,4191	0,4310	0,7019	0,4717	0,4610
9.	0,3247	0,5355	0,3176	0,4311	0,1915	0,2787	0,4965	0,4436	—	0,3355	0,3010	0,4531	0,3671	0,5023	0,4518
10.	0,4737	0,2608	0,1518	0,4767	-0,0604	0,0784	0,4580	0,3322	0,3355	—	0,5188	0,3000	0,3634	0,4091	0,3404
11.	0,4225	0,3116	0,1960	0,3975	0,0932	0,2231	0,3641	0,4191	0,3010	0,5188	—	0,3446	0,4359	0,4669	0,4129
12.	0,2901	0,4221	0,3334	0,3109	0,1350	0,2795	0,3992	0,4310	0,4531	0,3000	0,3446	—	0,4374	0,4098	0,4009
13.	0,3046	0,3778	0,2265	0,3084	0,0989	0,1953	0,4017	0,7019	0,3671	0,3634	0,4359	0,4374	—	0,4755	0,4189
14.	0,4155	0,4157	0,2276	0,4756	0,0801	0,2336	0,5924	0,4717	0,5023	0,4091	0,4669	0,4098	0,4755	—	0,4094
15.	0,3335	0,3554	0,2591	0,3764	0,0563	0,2199	0,4270	0,4610	0,4518	0,3404	0,4129	0,4009	0,4189	0,4904	—

2. From the underlined entries select the highest in the entire table. In this case it is 0,70197, which corresponds to scenarios 8 (geologist knowledge) and 13 (insider trading). These scenarios constitute the first cluster.

3. Then read across rows 8 and 13. If another scenario other than the scenario in each of the rows is underlined, this brand is added to the first cluster. As no other scenarios are underlined, cluster 1 remains scenarios 8 and 13.

4. The procedure repeats itself by selecting the highest entry in the table after 8 and 13 have been omitted. The highest remaining value is 0,6280, which corresponds to scenarios 4 and 7. These two scenarios represent the second cluster.

5. Again read across the row in which the coefficient is found, that is rows 4 and 7, to see if another scenario other than scenarios 4 and 7 is underlined. In this case scenarios 1 and 14 are added to the cluster.

6. This procedure is continued until all the remaining scenarios have been assigned to a particular cluster. In this instance cluster 3 contains scenarios 2, 9, 6 and 12; cluster 4 contains scenarios 10 and 11; cluster 5 contains scenario 15; and, cluster 6 contains scenarios 3 and 5.

Cluster analysis, for the most part, requires judgement on the part of the researcher as to the number and composition of the clusters to be formed. Obviously this is not always an easy task to undertake.

The clusters formed by the preceding analysis have been interpreted as follows, hopefully with appropriate names given to them.

Cluster 1 'The Insiders'
(Scenarios 8 and 13)

Consists of situations pertaining to the use of information privy to the organization.

Cluster 2 'Felons'
(Scenarios 1, 4, 7 and 14)

Consists of situations where the action is either criminal (and would be punishable in a court of law), or the gift so large as to constitute enrichment above and beyond the nature of the performance which it is supposed to warrant. The material enrichment comes from *outside* the organization however, as opposed to situations 10 and 11 (cluster 4), where the organization itself is the direct victim.

Cluster 3 'Happy Holidays'
(Scenarios 2, 9 , 6 and 12)

These situations all resolve around the receipt of free time spent away from home/the office, and would conceivably not involve the recipient in any untoward activity. They do differ in magnitude however — from tickets to a sports event to a free overseas trip.

Cluster 4 'The Fiddlers'
(Scenarios 10 and 11)

Situations 10 and 11 both entail a direct defrauding of the company — in this case the victim is the organization, at the mercy of people who have been put in a position of trust.

Cluster 5 'A bit on the side'
(Scenario 15)

While it was at no time intimated what the 'night out' entailed, respondents obviously read rather a lot into the scenario, and it forms a cluster by itself.

Cluster 6 'The Innocents'
(Scenarios 3 and 5)

As would be expected, situations 3 and 5, the lunch and the small Christmas gift, which are intrinsically harmless under most circumstances, are clustered together.

Conclusions

From the clusters identified in the aforegoing analysis, it is evident that there are situations similar in nature and with regard to degree of perceived participation. Quick clustering of pilot study data should therefore lead to improved questionnaire design, both in terms of question semantics and also with regard to the length of questionnaires — a number of 'almost duplicated' scenarios, such as one of scenarios 7 or 14, or one of 2 or 9 could either be eliminated, or replaced with a scenario which addresses another dimension of the ethical spectrum, in the case of this study. Another benefit of the quick clustering procedure could be that it identifies an excess of similar scenarios which may report low perceived participation simply because there is not a great deal of opportunity for participation — scenarios 8 and 13, which both entail insider trading, are a case in point. In many organizations there may be low participation in insider trading not because employees are so honest, but simply because the opportunity does not often present itself. The researcher finds himself on the fine line of trying to measure 'would they?', rather than 'do they?'

There are a number of techniques, such as factor analyses and the more sophisticated clustering techniques which would achieve at least the same, and probably better results than the quick clustering approach. While it is not our intention to examine the results of more advanced cluster analyses, we have used the SAS FASTCLUS procedure, as well as the hierarchical procedures under STATGRAPHICS in the analysis of the perceived participation data to achieve similar results. Furthermore, the utilization of focus groups and more exhaustive pre-testing of questionnaires should result in more effective questionnaires. However, it is apparent that Kamen's Quick Clustering Technique does offer the important advantages of convenience and simplicity.

References

Afifi, A.A. & Clark, V. 1984. *Computer-Aided Multivariate Analysis*. Lifetime Learning Publications, Belmont, California.

Dillon, W.R. & Goldstein, M. 1984. *Multivariate Analysis: Methods and Applications*. New York, John Wiley and Sons.

Ferrell, O.C. & Gresham, L.G. 1985. A Contingency Framework for Understanding Ethical Decision-making in Marketing. *J. Market*. 49, 87–96.

Green, P.E. & Tull, D.S. 1978. *Research for Marketing Decisions*, 4th Ed. Englewood Cliffs, New Jersey, Prentice-Hall Inc.

Kamen, J.M. 1970. Quick Clustering. *J. Market. Res.*, May.

Pitt, L.F. 1985. Managerial Attitudes towards Corruption — A Pilot Study. *S. Afr. J. Bus. Mgmt.* vol. 16, 27–30.

Pitt, L.F. & Nel, D. 1988. The Wearer's Merit — A Comparison of the Attitudes of Suppliers and Buyers towards Corruption in Business. *Industr. Market. Purchas.* vol. 3.1.

Punj, G. & Stewart, G.W. 1983. Cluster Analysis in Marketing Research: Review and Suggestions for Application. *J. Market. Res.* May.

Sokel, R.R. & Sneath, P.H.A. 1963. *Principles of Numerical Taxonomy*. San Francisco, W.H. Freeman.

Wentz, W.B. 1979. *Marketing Research — Management, Methods and Cases*, 2nd Edition. New York, Harper and Row.

Wilkes, R.E. 1978. Fraudulent behaviour by consumers. *J. Market*. October.

Appendix 1 Scenarios

1. A purchasing manager is offered the sum of R5 000 on condition that he awards a large contract to a certain company. He accepts the offer.

2. A project engineer awards a sizable tender. A week later the company invites him on an all expenses-paid shooting trip to their private game reserve. He accepts their offer.

3. Members of a company negotiating team accept a potential supplier's invitation to lunch.

4. On condition that he let them know of rival tender prices, a tender official is offered the opportunity of having his home carpeted throughout. He supplies every price tendered.

5. A company official receives a bottle of whiskey from a supplier as a Christmas gift. He phones to say thanks.

6. Shortly after awarding a large contract, a company official and his son are invited to join the suppliers at their private box at a major rugby match. They accept the offer.

7. Shortly before the announcement of a large new tender, one of the tender officials has a sheep and a case of whiskey delivered to his home by one of the parties. He accepts the gifts.

8. A geologist working for a major mining group gains important information regarding the development of a new reef. He immediately buys 1000 company shares.

9. A branch manager receives a voucher for a weekend's stay at a holiday resort from a supplier, as a birthday present. His secretary telephones his thanks.

10. A senior manager siphons petrol from his company car and transfers it to his wife's car for her shopping.

11. The marketing manager and his wife entertain friends at a top restaurant. He books the bill to his expense account...'dined with potential customers'.

12. A company is known to be in the market for twenty heavy vehicles. One of the potential suppliers invites the specification engineer on an overseas trip so that he can 'visit their highly sophisticated facilities'. He accepts the offer.

13. The financial manager knows that the next company report will be the best yet. He instructs his broker to purchase R10 000's worth of shares on his behalf.

14. On returning from a visit to a supplier a buyer finds his car boot filled with groceries. He does nothing about this.

15. On arrival at his hotel in another city, a company purchasing manager receives a phone call from the secretary of a supplier's sales manager. Her boss, she says, has instructed her to take him out for a 'night on the town'. He accepts her invitation.