ABSTRACT

In the area of Marketing Research, measurement of attitudes, opinions, preferences and perceptions often involve ranking judgements. For market segmentation, product positioning, test marketing etc. one needs to cluster these rank ordered data. While clustering literature abounds in methods relevant to Interval/Ratio and Nominal scaled variables, cluster analysis of rank ordered data is mostly unavailable. This thesis dwells on this aspect of clustering rank ordered data.

The notion of cluster is derived from the concepts of distance and closeness, similarity and dissimilarity measures among the data. At the very beginning we introduce to rank correlation coefficients - Kendall's "TAU" and Spearman's "RHO", which have been traditionally discussed in literature. It is shown that while "RHO" has more discriminating power, "TAU" is on a sound theoretical footing. We then proceed construct new rank correlation coefficients which have at least as much discriminating power as "RHO" and as much theoretical and intuitive basis as "TAU".

The first rank correlation coefficient we come up with is Gamma ab. A given n element rank vector is converted into vector C of length Nc2 and Gamma ab is computed as inner product of C and the C-vector corresponding to the ideal preference ranking. It turns out that "Gamma ab" is nothing but a reinterpretation of "RHO" and has the same discriminating power as that of "RHO". It also possesses the theoretical soundness of "TAU".

The Lexicographic Rank correlation coefficient of a ranking R1 is also defined with respect to ideal preference ranking Ri. It exploits the features of radix base notation of positional number system. Using a normalization procedure we show how to select an appropriate base. As an interesting by product, it is also shown as how rankings become indistinguishable when the cardinality of the rank vector exceeds 9. It is shown how the theory behind the definition of the coefficient is in consonance with what Miller discovered regarding limit of span of absolute judgement of human memory.

The pure weights coefficient - another heuristic of similarity between rankings takes into account two features of rank ordered data - the preference of one brand or another and the placing of this preference in the rank list. In order to arrive at weights, use is made of Weber's law in psychophysics.

In defining the vector coefficient of rankings, ranks attached, for example to brands etc., are viewed as vectors (with magnitude and direction) and we make use of vector products. This coefficient reveals and elegant structure relationship that exists among various rank listings.

In order to study the performance of various coefficients defined earlier in the context of clustering, a clustering experiments was conducted based on a sample survey among I.I.M. (B) student population. We give details of the methodology adopted for obtaining real world data. Of the four new coefficients defined, the first three resulted in better grouping than that obtained using "TAU" and "RHO". The performance of vector coefficient was inferior to the rest. The cluster generated by it do not conform to the conventional hierarchical structure although its use in detecting random rankings data is suggested.

Finally a thorough survey of clustering techniques and certain issues related to market research are highlighted. A novel interpretation for hierarchical clustering of rank ordered data is also presented.