Concurrent Segmentation

Background

Standard cluster-based segmentation schemes often encounter a standard problem: the segments they create fail to show worthwhile differences on anything except the "basis variables"—the variables that were used to create them in the first place. The earliest attempts at segmentation were usually based on demographics; however, such demographically based segments were often found to show only weak differences in other measures such as attitudes or brands used – the types of measures where differences would be important and useful.

In response to this problem, early forms of needs- or benefits-based segmentation were developed as a replacement for demographics-based segmentation. Those early benefits segmentations used attitudinal batteries or product attribute importance ratings to produce segments with very different attitudes or needs. But, as one might expect, a simple reversal of the earlier problem resulted: the resulting attitudinal segments were often very difficult to identify or reach due to a lack of sharp demographic differences.

A seemingly obvious way to overcome these problems is to cluster on two sets of basis variables simultaneously, such as on demographics and attitudes together. This simple approach, however, encounters technical problems of its own. First, scale differences in the variables can create mathematical problems and necessitate arbitrary decisions about rescaling—decisions that do, unfortunately, affect the outcome. Second, even after those issues are addressed, or even if scaling is compatible between two different types of variables, it is all too common to find that the end result is a mix of incoherent segments. Some segments, for example, may be internally very similar on demographics but not on attitudes, while others may be attitudinally coherent but not show any demographic consistency. In effect, the jointly defined segments often look as if they had come from two different segmentations that were pasted together incorrectly.

Solution

Concurrent segmentation is a way of finding usable segments on the basis of two sets of basis variables. It relies not just on the two sets of variables by themselves, but specifically on the *relationships* between them. A variable in either set will influence the segments only to the extent that it can predict one or more variables in the other set. Variables that are unrelated to those in the other set of basis variables are effectively ignored. The net result from concurrent segmentation is a set of segments that differ substantially on *both* sets of variables between segments, and are relatively consistent on *both* sets of variables within segments.

It is true that, just as one would have to expect, concurrent segmentation produces segments that are not quite as sharply different on either set of basis variables as traditional single-focus segmentation would be. However, the differences are typically 70% to 80% as large as those obtained in single-basis segmentations. This small loss is more than made up for by the fact that the segments differ usefully on both sets of basis variables. In the case of attitudes and demographics as the basis variables, this means that we can not only understand what each segment wants, but have reasonably precise ways to target them in terms of demographics. This is clearly far better than having slightly better differentiated needs but no way of targeting, or having very clear demographic cells with no real differences in attitudes.

<u>Practical Implementation</u>

Of course, the example of a demographics vs. attitudes pairing is just one possible pair of basis variables. Others can be used depending on the detailed objectives of the study. Brand choice or purchase behavior (or other non-brand behavioral measures) is a common basis for segmentation, and can be paired with either demographics or attitudes. And while the general concept of "attitudes" is usually taken to mean "needs" or "wants" or "benefits sought" in the marketplace, it can also be defined as perceptions of brands, or as more general attitudes such as psychographics or lifestyle measures.

Often, it is impossible to specify in advance exactly which two sets of variables are crucial. In those cases, we can produce concurrent segmentations on several different pairs of bases, and examine each to see which segments make the most sense or would be the most useful. In other cases, we may feel there are a number of different sets of variables that should be included. In that situation, we can group the possible variables into two "supersets" and proceed with concurrent segmentation on the supersets. We might try grouping them together in different ways, or including different subsets of them and, as in the first case, examine the results to see which make the most sense.

In practice, we often use 6 to 12 different "pairings" of sets of variables and use concurrent segmentation to produce "candidate segmentations" from each. For each pairing, we would typically produce solutions having from three to ten segments, so we might have a total of 50 to 100 (eight numbers of segments times 6 to 12 pairings) candidate segmentations to look at. We screen these candidates statistically to see how many significant differences are generated by each candidate segmentation. We summarize the number of significant differences by questionnaire topic for each candidate segmentation, and inspect that summary to find about three candidates that perform well and are based on different pairings of variables. For those final candidates, we produce detailed profiles, and examine them from the standpoint of managerial usefulness and actionability. In practice, we almost always find that one of the final candidates is a good final solution worthy of recommendation to the client. However, we can also go back to repeat all or part of the process with different candidates or with different pairings of variables, based on what deficiencies we seen in the candidates. Often, we examine the final candidates in conjunction with our client, in a preliminary work-session.

Once segments are developed through concurrent segmentation, we can use discriminant analysis or CART/CHAID analyses to find simpler ways to rapidly identify them in future surveys, in prospect screening situations or in database classification situations. Even though numerous variables may have been used in the original definition of the segments, we can usually find a much smaller set of questions or variables (often, 5 to 15) that can reproduce the segments with reasonable accuracy. When the ability to operationally classify the segments is crucial, we can conduct our analyses and evaluations of potential segments not in terms of their original "pure" definitions but in terms of the segments "as classified." In other words, we redefine the segments as the being the ones produced by the discriminant or CART analyses. Thus, they become ones we could operationally recognize with perfect accuracy. We then conduct the evaluation of candidate segmentations based on their operational definitions rather than on theoretical definitions that cannot be operationalized accurately.