

Canonical segmentation

Segmentation is a vital tool in determining the effectiveness of ad spend. Canonical segmentation takes this a step further by relating two (or more) sets of variables to provide more refined analysis. A sample case from the skincare market illustrates how this can be used to advantage in marketing practice

By Michael Lieberman, *Multivariate Solutions*

Segmentation is a term we've written about in the past. A popular example is a technique called latent class segmentation, which is an extension of factor analysis. In tandem segmentation, we join multiple dimensions of statements and arrays in a two-step process that first uses latent class to form first-level segments, and next, *k*-means cluster analysis to bring together the multiple segment solutions into four to five final segments. We then use business rules – comparison variables such as spend and frequency – to separate the final segments. We've also considered customer classification waterfalls and hierarchical clustering.

In this article, we are going back to the future. We'll review a technique that has been around awhile, but has become far more accessible due to the exponential increase in computing power and the ubiquity of open-source statistical software such as R, SPSS, or MaxStat. With such software, we can combine disparate data sets within a survey, thus overcoming differing scales or dominant variables. Since the method is based on canonical correlation, we will call it canonical segmentation.

A well-known problem with cluster analyses is that the segments they create do not always show separation of key attributes – what I like to call business rules. Segmenting on demographics might

be great for looking at demographics, but it might not separate important issues such as how much money people spend on the product, how many times they visit your

restaurant chain, or how readily they join your loyalty programme.

Canonical segmentation can solve this dilemma by including business rules in the actual cluster process. This method combines two or more (though two is recommended) basis sets of variables. It relies not only on the sets of variables, but specifically on the relationship between them. A single variable in either set is significant only to the extent that it relates to the other set of variables. The end result from canonical segmentation is a set of segments that differ on both sets of variables between the segments, yet are relatively consistent on both sets of variables within the segment. Put simply, we segment two sets of variables first, run a canonical correlation to determine which variables are most closely related, then segment on those relationships.

Canonical is a statistical term for analysing latent variables – those which are not directly observed – that represent multiple variables which are directly observed. This use of latent variables is common among other

FIGURE 1: SKINCARE LATENT CLASS ANALYSIS AND CANONICAL CORRELATION: LATENT RELATIONSHIP AMONG TWO SETS

Correlations in brackets for skincare attributes are factor correlations from the first step

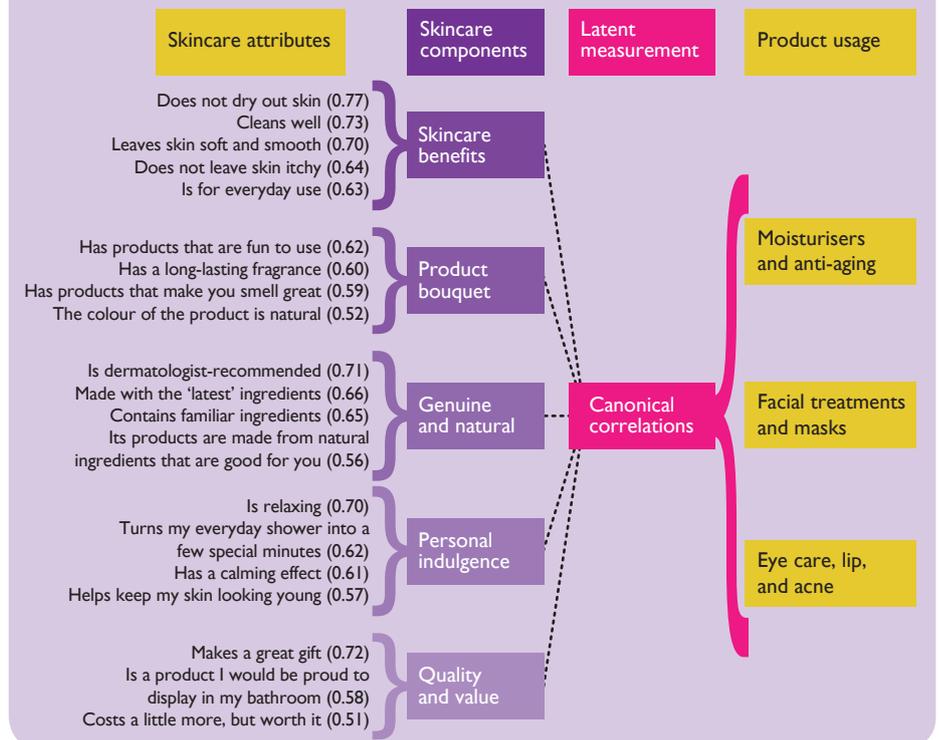
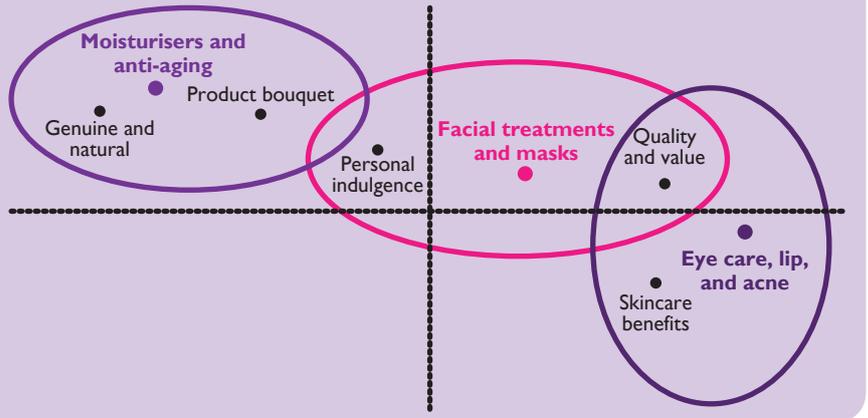


FIGURE 2: TYPES OF SKINCARE USAGE

Moisturisers. Moisturisers and night creams
Face wash and cleansers. Exfoliators. Face wipes
Anti-aging. Teeth whitening
Treatments. Face serums. Blemish and acne treatments
Face masks. Sheet masks
Eye care. Eye creams and treatments. Eye masks
Lip treatments. Lip balms and treatments. Lip sunscreen
Acne blemishes

FIGURE 3: CORRESPONDENCE MAP



techniques such as structural equations modelling. The ability to analyse groups of variables into one latent variable lends canonical correlation a great deal of agility.

To illustrate how this technique works in practice, let's consider a sample case from the skincare market, using canonical correlation to link two cluster groups.

Brand choice or purchase behaviour is a common basis for segmentation. It can easily be paired with demographics or

combination of variables in the attitude set; and find a linear combination of variables in the usage set.

First, we conduct a factor analysis on each set of basis variables separately. Seeking the two sides to be as related as possible, we choose the most intuitive solution. Next, all the variables in one set are regressed on all those in the other – not individually, but 'as a set'. This is the core of canonical correlation, namely the creation of 'latent variables'.

prone to care about genuine and natural qualities and product bouquet.

The primarily facial treatments and masks segment combines that particular product usage with personal gratification together with quality and value.

Finally, the eye care, lip and acne segment is mostly purchasing based on benefits to combat acne and imperfections. These consumers, too, are quality and value conscious.

Our last step would be to recreate the segments for future surveys. We can use discriminant analysis, logistic regression or CART/CHAID analyses to find simpler ways to identify them rapidly. Although 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 by employing predictive analytic regression to weed out insignificant variables, then deploying our reclassification scheme with up to 300% increase in efficiency.

“Put simply, we segment two sets of variables first, run a canonical correlation to determine which variables are most closely related, then segment on those relationships”

attitudes. My mentor Dave Lyon writes: “And while the general concept of attitudes is usually taken to mean ‘needs’ or ‘wants’, it generally refers to benefits sought in the marketplace.” Thus, brand descriptions. In addition, attitudes are commonly used for lifestyle and psychographics.

Often it is difficult to specify in advance which two sets of variables are most ideal for our canonical segmentation. Sometimes we need to run several versions of the analysis before we find the best solution. For this study, our skincare client has chosen to segment on skincare attitudes and what would be a strong business rule – category product usage. Our steps are to: find a linear

Our first step is to conduct a latent class segmentation on usage needs for skincare products (factor correlations in Figure 1). Next, we segment types of skincare product usage (Figure 2). Finally, we derive the canonical functions (also Figure 1). The value of canonical correlations and the resulting segmentation is the mix of the two segmentations. The output of the latent variables can be shown in a correspondence map, which clearly outlines our segments (Figure 3).

Relationships are defined by proximity in our map. Our maintenance segment (moisturisers and anti-aging) is a combination of that usage where consumers are more

CONCLUSION

Not every dollar spent on advertising and sales is created equal – some of those dollars generate far more revenue than others. Segmentation is a vital tool in the effort to determine which is which. Canonical segmentation holds an important position in the quiver of multivariate clustering methods. As bandwidth progressively expands, it will become even easier to deploy.

This article was first published in Admap magazine February 2017 ©Warc www.warc.com/admap