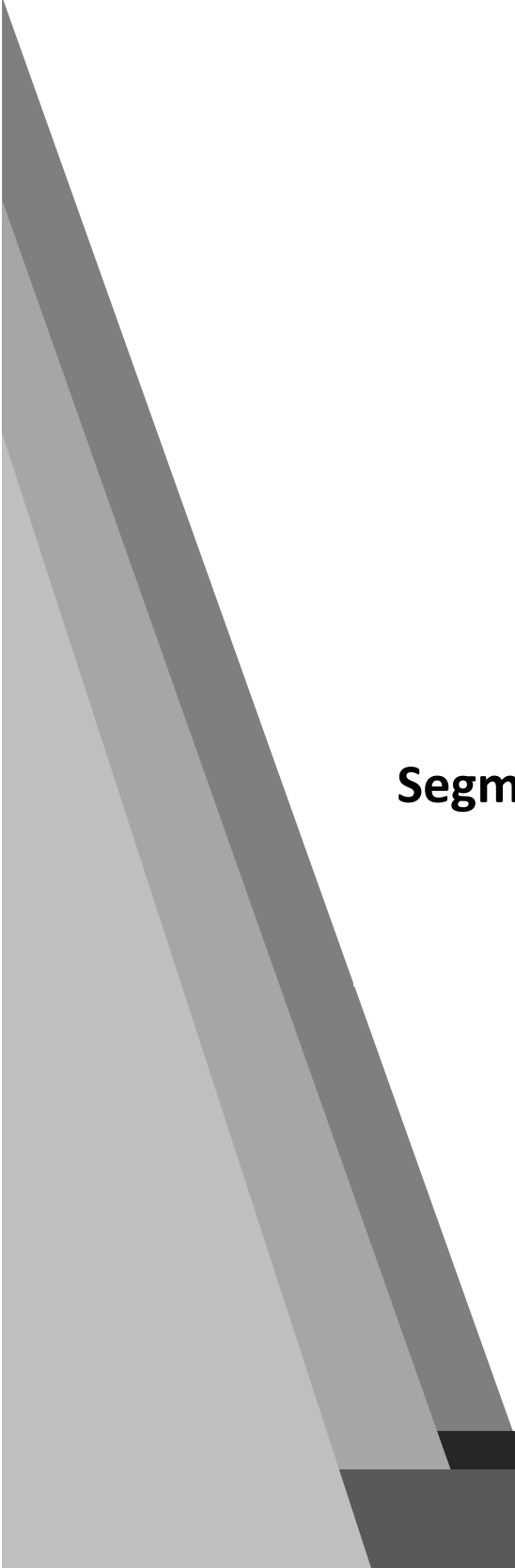




Sawtooth Software

RESEARCH PAPER SERIES

Segmentation: Four Common Types of Segmentation

Keith Chrzan
Sawtooth Software, Inc.

Segmentation: Four Common Types of Segmentation

Keith Chrzan, Sawtooth Software

Introduction

Segmentation helps marketers understand how groups of customers differ with respect to the products, messaging or positioning that appeal to them. Understanding these differences gives marketers more leverage in designing or selling products to their customers.

Many discussions of segmentation delve into a red herring topic about kinds of basis variables: should you segment on demographic variables, on psychographic or attitudinal variables, or maybe on needs or behaviors. Nothing at all, however, prevents us from segmenting our data on any combination of the above and in fact most of the segmentation studies I work on do involve combinations. To think about “types of segmentation,” I think it makes more sense to consider where the segmentation data comes from and how the segmentation variables relate to one another. Just knowing these two things can go a long way in pointing us to the analysis methods that will best meet our study’s objectives.

Are we segmenting data from a conjoint or choice experiment?

Sometimes we want to have “needs-based” segmentations, where the needs are measured with a conjoint or MaxDiff experiment. For segmenting data from choice experiments, use latent class multinomial logit (MNL). Eagle and Magidson (2019) and Lyon (2019) present compelling evidence that clustering methods (including latent class clustering) are inappropriate for segmenting conjoint utilities. Sawtooth Software offers options for this kind of Latent Class MNL analysis in both Lighthouse Studio and in its standalone Latent Class software.

For segmentations based on a mix of utility models from a choice experiment and data from outside the choice experiment, latent class MNL is still the way to go, and the Latent Gold software can handle models like this.

Do we have a dependent variable?

Sometimes our study has a single variable we’d really like to see the segments differ on. A common example comes when segmenting a market to find the target segment(s) for a new product launch and we want to find segments that are most interested in buying the product. We call these supervised segmentations, because the dependent variable (purchase intent) plays a supervisory role in defining segments. When we have no such dependent variable, we call that an unsupervised segmentation.

Supervised segmentation

Supervised segmentations usually involve tree-based models – classification trees, regression trees and conditional inference trees. The first supervised segmentation I worked on was for DIRECTV, back before it was called DIRECTV: the client wanted to identify the target market for satellite-delivered TV programming so that they could spend their limited advertising budget telling the right people about the

new service. In addition to measuring a large number of demographic and psychographic variables, we asked respondents how much they expected to spend on a satellite programming service we described. Our tree-based segmentation identified six segments of respondents, two with high levels of expected spend, one with medium spend expectations and three anticipating low levels of expected spend. A benefit of supervised segmentation is that we need not put a whole lot of thought into what basis variables to use. We can toss in any variables (even if they're a mix of metric or ordinal or categorical variables) and let the decision tree tell us which combination best identifies segments. A side benefit of tree-based segmentation is that the tree also serves as the typing tool (a typical segmentation deliverable) for classifying future respondents into the segments – there is no need to run a separate analysis, like a discriminant analysis for example, to create a typing tool.

Unsupervised segmentation

In the absence of a supervising variable, we have unsupervised segmentation. These typically use some form of cluster analysis or latent class clustering. We select a set of basis variables (the variables that the algorithm will use to create segments) and we let the algorithm identify groups that differ on those variables. Unsupervised segmentation is about the most difficult thing marketing research analysts are called upon to do – no supervision guides the model and the analyst needs to make several decisions, based on limited information, which can greatly affect the results of the segmentation. This large topic appears in Part 2 of this white paper: “Segmentation – How to Do It Badly and Well.”

Lightly Supervised Segmentation

I use a third type of supervision more often than either of the above, however – call it “lightly supervised” (“semi-supervised” sounds better, but that’s already the name of something else). As the name suggests, a lightly supervised segmentation sits in the middle ground between supervised and unsupervised segmentations. We usually choose a lightly supervised segmentation when we do have a variable we want to see differ across segments, but our sample is less than the many hundreds needed to support a rich tree-based segmentation.

In a lightly supervised segmentation, we choose our basis variables in a specific way:

1. First, we identify a variable we would like to see differ by the segments – this is our lightly supervising variable.
2. Next, we examine correlations between that variable and all the plausible basis variables (all the variables that could make the segments of respondents identifiable for marketing efforts).
3. We include as basis variables only those variables that are significantly (or highly) correlated with the lightly supervising variable.

The resulting segments almost always differ significantly with respect to the lightly supervising variable, even though it was itself never included as a basis variable.

What’s Next?

An unsupervised/cluster-based segmentation requires enough modeling decisions as to deserve a separate, lengthier (and geekier) discussion: Segmentation – How to Do It Badly and Well.