

Eleven Multivariate Analysis Techniques: Key Tools In Your Marketing Research Survival Kit

By Michael Richarme, Ph.D.

Situation 1: A harried executive walks into your office with a stack of printouts. She says, “You’re the marketing research whiz—tell me how many of this new red widget we are going to sell next year. Oh, yeah, we don’t know what price we can get for it either.”

Situation 2: Another harried executive (they all seem to be that way) calls you into his office and shows you three proposed advertising campaigns for next year. He asks, “Which one should I use? They all look pretty good to me.”

Situation 3: During the annual budget meeting, the sales manager wants to know why two of his main competitors are gaining share. Do they have better widgets? Do their products appeal to different types of customers? What is going on in the market?

All of these situations are real, and they happen every day across corporate America. Fortunately, all of these questions are ones to which solid, quantifiable answers can be provided.

An astute marketing researcher quickly develops a plan of action to address the situation. The researcher realizes that each question requires a specific type of analysis, and reaches into the analysis tool bag for...

Over the past 20 years, the dramatic increase in desktop computing power has resulted in a corresponding increase in the availability of computation intensive statistical software. Programs like SAS and SPSS, once restricted to mainframe utilization, are now readily available in Windows-based, menu-driven

packages. The marketing research analyst now has access to a much broader array of sophisticated techniques with which to explore the data. The challenge becomes knowing which technique to select, and clearly understanding its strengths and weaknesses. As my father once said to me, “If you only have a hammer, then every problem starts to look like a nail.”

Overview

The purpose of this white paper is to provide an executive understanding of 11 multivariate analysis techniques, resulting in an understanding of the appropriate uses for each of the techniques. This is not a discussion of the underlying statistics of each technique; it is a field guide to understanding the types of research questions that can be formulated and the capabilities and limitations of each technique in answering those questions.

In order to understand multivariate analysis, it is important to understand some of the terminology. A variate is a weighted combination of variables. The purpose of the analysis is to find the best combination of weights. Nonmetric data refers to data that are either qualitative or categorical in nature. Metric data refers to data that are quantitative, and interval or ratio in nature.

Initial Step—Data Quality

Before launching into an analysis technique, it is important to have a clear understanding of the form and quality of the data. The form of

The purpose of this white paper is to provide an executive understanding of 11 multivariate analysis techniques, resulting in an understanding of the appropriate uses for each of the techniques.

the data refers to whether the data are nonmetric or metric. The quality of the data refers to how normally distributed the data are. The first few techniques discussed are sensitive to the linearity, normality, and equal variance assumptions of the data. Examinations of distribution, skewness, and kurtosis are helpful in examining distribution. Also, it is important to understand the magnitude of missing values in observations and to determine whether to ignore them or impute values to the missing observations. Another data quality measure is outliers, and it is important to determine whether the outliers should be removed. If they are kept, they may cause a distortion to the data; if they are eliminated, they may help with the assumptions of normality. The key is to attempt to understand what the outliers represent.

Multiple Regression Analysis

Multiple regression is the most commonly utilized multivariate technique. It examines the relationship between a single metric dependent variable and two or more metric independent variables. The technique relies upon determining the linear relationship with the lowest sum of squared variances; therefore, assumptions of normality, linearity, and equal variance are carefully observed. The beta coefficients (weights) are the marginal impacts of each variable, and the size of the weight can be interpreted directly. Multiple regression is often used as a forecasting tool.

Logistic Regression Analysis

Sometimes referred to as “choice models,” this technique is a variation of multiple regression that allows for the prediction of an event. It is allowable to utilize nonmetric (typically binary) dependent variables, as the objective is to arrive at a probabilistic assessment of a binary choice. The independent variables can be either discrete or continuous. A contingency table is produced, which shows the classification of observations as to whether the observed and predicted events match. The sum of events that were predicted to occur which actually did occur and the events that were predicted not to occur which actually did not occur, divided by the total number of events, is a measure of the effectiveness of the model. This tool helps predict

the choices consumers might make when presented with alternatives.

Discriminant Analysis

The purpose of discriminant analysis is to correctly classify observations or people into homogeneous groups. The independent variables must be metric and must have a high degree of normality. Discriminant analysis builds a linear discriminant function, which can then be used to classify the observations. The overall fit is assessed by looking at the degree to which the group means differ (Wilkes Lambda or D2) and how well the model classifies. To determine which variables have the most impact on the discriminant function, it is possible to look at partial F values. The higher the partial F, the more impact that variable has on the discriminant function. This tool helps categorize people, like buyers and nonbuyers.

Multivariate Analysis of Variance (MANOVA)

This technique examines the relationship between several categorical independent variables and two or more metric dependent variables. Whereas analysis of variance (ANOVA) assesses the differences between groups (by using T tests for 2 means and F tests between 3 or more means), MANOVA examines the dependence relationship between a set of dependent measures across a set of groups. Typically this analysis is used in experimental design, and usually a hypothesized relationship between dependent measures is used. This technique is slightly different in that the independent variables are categorical and the dependent variable is metric. Sample size is an issue, with 15-20 observations needed per cell. However, too many observations per cell (over 30) and the technique loses its practical significance. Cell sizes should be roughly equal, with the largest cell having less than 1.5 times the observations of the smallest cell. That is because, in this technique, normality of the dependent variables is important. The model fit is determined by examining mean vector equivalents across groups. If there is a significant difference in the means, the null hypothesis can be rejected and treatment differences can be determined.

Factor Analysis

When there are many variables in a research design, it is often helpful to reduce the variables to a smaller set of factors. This is an independence technique, in which there is no dependent variable. Rather, the researcher is looking for the underlying structure of the data matrix. Ideally, the independent variables are normal and continuous, with at least 3 to 5 variables loading onto a factor. The sample size should be over 50 observations, with over 5 observations per variable. Multicollinearity is generally preferred between the variables, as the correlations are key to data reduction. Kaiser's Measure of Statistical Adequacy (MSA) is a measure of the degree to which every variable can be predicted by all other variables. An overall MSA of .80 or higher is very good, with a measure of under .50 deemed poor.

There are two main factor analysis methods: common factor analysis, which extracts factors based on the variance shared by the factors, and principal component analysis, which extracts factors based on the total variance of the factors. Common factor analysis is used to look for the latent (underlying) factors, whereas principal components analysis is used to find the fewest number of variables that explain the most variance. The first factor extracted explains the most variance. Typically, factors are extracted as long as the eigenvalues are greater than 1.0 or the scree test visually indicates how many factors to extract. The factor loadings are the correlations between the factor and the variables. Typically a factor loading of .4 or higher is required to attribute a specific variable to a factor. An orthogonal rotation assumes no correlation between the factors, whereas an oblique rotation is used when some relationship is believed to exist.

Cluster Analysis

The purpose of cluster analysis is to reduce a large data set to meaningful subgroups of individuals or objects. The division is accomplished on the basis of similarity of the objects across a set of specified characteristics. Outliers are a problem with this technique, often caused by too many irrelevant variables. The sample should be representative of the population, and it is desirable to have uncorrelated factors. There are three main clustering

methods: hierarchical, which is a treelike process appropriate for smaller data sets; nonhierarchical, which requires specification of the number of clusters a priori, and a combination of both. There are 4 main rules for developing clusters: the clusters should be different, they should be reachable, they should be measurable, and the clusters should be profitable (big enough to matter). This is a great tool for market segmentation.

Multidimensional Scaling (MDS)

The purpose of MDS is to transform consumer judgments of similarity into distances represented in multidimensional space. This is a decompositional approach that uses perceptual mapping to present the dimensions. As an exploratory technique, it is useful in examining unrecognized dimensions about products and in uncovering comparative evaluations of products when the basis for comparison is unknown. Typically there must be at least 4 times as many objects being evaluated as dimensions. It is possible to evaluate the objects with nonmetric preference rankings or metric similarities (paired comparison) ratings. Kruskal's Stress measure is a "badness of fit" measure; a stress percentage of 0 indicates a perfect fit, and over 20% is a poor fit. The dimensions can be interpreted either subjectively by letting the respondents identify the dimensions or objectively by the researcher.

Correspondence Analysis

This technique provides for dimensional reduction of object ratings on a set of attributes, resulting in a perceptual map of the ratings. However, unlike MDS, both independent variables and dependent variables are examined at the same time. This technique is more similar in nature to factor analysis. It is a compositional technique, and is useful when there are many attributes and many companies. It is most often used in assessing the effectiveness of advertising campaigns. It is also used when the attributes are too similar for factor analysis to be meaningful. The main structural approach is the development of a contingency (crosstab) table. This means that the form of the variables should be nonmetric. The model can be assessed by examining the Chi-square value for the model. Correspondence analysis is difficult to interpret,

as the dimensions are a combination of independent and dependent variables.

Conjoint Analysis

Conjoint analysis is often referred to as “trade-off analysis,” in that it allows for the evaluation of objects and the various levels of the attributes to be examined. It is both a compositional technique and a dependence technique, in that a level of preference for a combination of attributes and levels is developed. A part-worth, or utility, is calculated for each level of each attribute, and combinations of attributes at specific levels are summed to develop the overall preference for the attribute at each level. Models can be built which identify the ideal levels and combinations of attributes for products and services.

Canonical Correlation

The most flexible of the multivariate techniques, canonical correlation simultaneously correlates several independent variables and several dependent variables. This powerful technique utilizes metric independent variables, unlike MANOVA, such as sales, satisfaction levels, and usage levels. It can also utilize nonmetric categorical variables. This technique has the fewest restrictions of any of the multivariate techniques, so the results should be interpreted with caution due to the relaxed assumptions. Often, the dependent variables are related, and the independent

variables are related, so finding a relationship is difficult without a technique like canonical correlation.

Structural Equation Modeling (SEM)

Unlike the other multivariate techniques discussed, structural equation modeling (SEM) examines multiple relationships between sets of variables simultaneously. This represents a family of techniques, including LISREL, latent variable analysis, and confirmatory factor analysis. SEM can incorporate latent variables, which either are not or cannot be measured directly into the analysis. For example, intelligence levels can only be inferred, with direct measurement of variables like test scores, level of education, grade point average, and other related measures. These tools are often used to evaluate many scaled attributes or build summated scales.

Conclusions

Each of the multivariate techniques described above has a specific type of research question for which it is best suited. Each technique also has certain strengths and weaknesses that should be clearly understood by the analyst before attempting to interpret the results of the technique. Current statistical packages (SAS, SPSS, S-Plus, and others) make it increasingly easy to run a procedure, but the results can be disastrously misinterpreted without adequate care.

About Decision Analyst

Contact Decision Analyst via email at jthomas@decisionanalyst.com or by phone at **1-800-262-5974** or **1-817-640-6166**.

Decision Analyst is a global marketing research and analytical consulting firm. The company specializes in advertising testing, strategy research, new products research, and advanced modeling for marketing-decision optimization.



604 Avenue H East ■ Arlington, TX 76011-3100, USA
1.817.640.6166 or 1.800. ANALYSIS ■ www.decisionanalyst.com