TARGET MARKETING WITH LOGIT REGRESSION

By Jeffrey S. Morrison

Target marketing requires selecting customers on the basis of socio-demographic characteristics ... logit regression can help to identify the right characteristics ... demonstrates with the help of an example how this method can be used.

As competition increases and budgets grow tighter, companies are looking for cheaper and more effective ways to sell their products and services. Greater emphasis is placed on identifying those customers who are most likely to purchase. In the past, targeting those customers may have been done in an ad-hoc manner depending on how historical sales data look in tables and cross-tabs. Now the movement is toward a statistical modeling framework which can be implemented at product introduction or any time during the product's life cycle.

AD-HOC ANALYSIS

Although the ad-hoc analysis of tables and cross-tabs for targeting customers is very informative, there are some drawbacks. First, it is difficult to perform more than a two or three dimensional look at the data. What you would like to do is to account for as many customer characteristics (factors) as possible while trying to answer a particular question. For example, do family size, household income, credit card debt, and consumer age have an impact on the decision to purchase a product? These types of questions can be answered only with a statistical based model.

Second, conclusions drawn from cross-tab analysis may not take into account sampling error. Constructing a statistically based model allows the analyst to quantify behavior for a number of factors simultaneously and draw conclusions that are statistically valid. A statistically based framework also has the added flexibility of testing for numerous specifications. As a result, a *best* solution can be selected from a set of plausible solutions. Benchmarks can then be established for comparisons with future analyses as more of the product is sold.

LOGIT REGRESSION

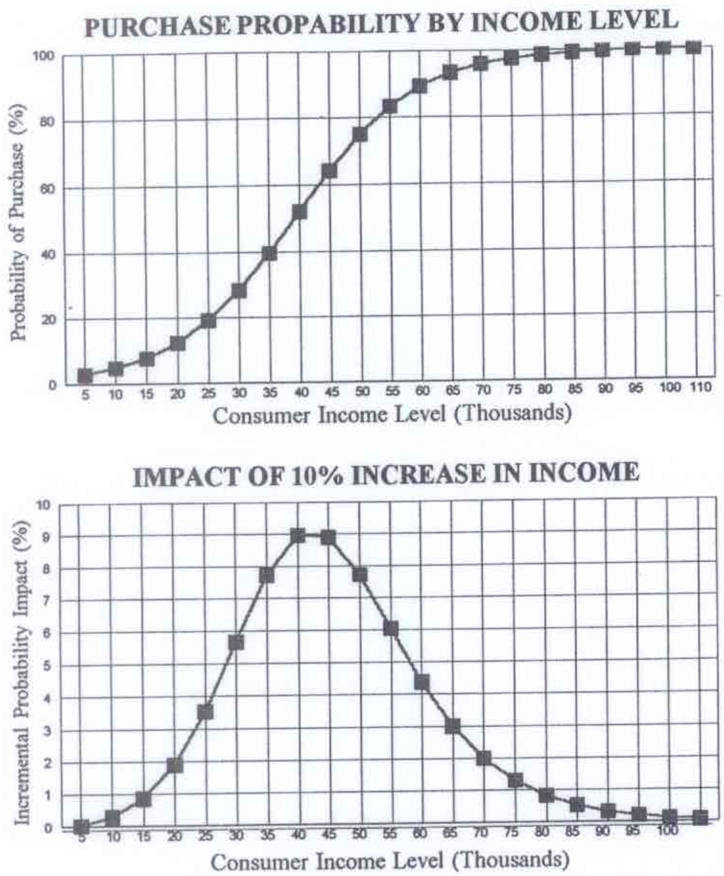
Standard econometric methods like ordinary least squares were designed for evaluating variables that can assume any value within a range, i.e. continuous variables. These methods are usually appropriate for examining data which have been accumulated over time into totals representing aggregate market response. When the outcome variable is not continuous, for example, to buy or not to buy a product, other techniques will be needed to properly measure and evaluate the decision making process. These techniques are called Discrete Choice Models.

JEFFREY S. MORRISON

Mr. Morrison is Director of Modeling at Equifax, based in Atlanta, Ga. His work experience includes the development of new product diffusion models, quantitatively based target marketing systems, and customer satisfaction simulation models for the telephone industry. In 1992, he won the Outstanding Speaker Award at the National Telecommunication Forecasting Conference.

The discrete choice technique discussed here is Logit Regression. Numerous statistical packages are available to handle these types of models (SAS, SPSS, SHAZAM, LIMDEP). In the simplest of cases, the consumer is faced with only two choices: (1) to purchase a product or (2) not to purchase. The consumer is assumed, in general, to make the decision in such a way as to maximize his or her utility. One of the advantages of discrete choice methods is that it treats the decision making process in a probabilistic manner. Once the equation is estimated, we can project the probability that a consumer will purchase the product based upon a set of explanatory criteria (education, income, age, etc.). This probability (ranging from 0 to 100%) can be interpreted as a *score* and used to rank each customer from those who are most likely to purchase to those least likely to purchase the product. Setting up the data for estimation looks very similar to ordinary least squares. The main difference is that the dependent variable (whether or not the product was purchased) is coded as a zero or one. Moreover, the logit procedure is non-linear and actually estimates the log of the odds of purchase.

CHART 1



Once estimated, the equation can be used to show how the probability of purchase varies across different values of the explanatory variable. Since logit regression is a non-linear technique, it is able to capture certain curvilinear relationships that may exist. The top of Chart 1 shows how the probability of purchase estimated by a logit regression may vary across different income levels. A consumer with an income of $25,000 has a 20% probability of buying the service — all other factors remaining equal. At very low levels of income you might expect a small increase in that variable to have minimum impact on the decision to purchase. This is because certain basic needs such as utilities and fixed costs have to be met. At very high income levels, different factors may contribute to some insensitivity. However, at levels in-between the consumer's decision is border line. Sensitivity to factors like price or income may "tip the scale" in favor of the alternative.

The logit equation can also be used for what-if analysis. For example, how much would the probability of purchase increase if we shifted selling efforts from prospects with an income of $20,000 to those with incomes of $22,000? The bottom part of Chart 1 shows how a 10% increase in consumer income ($20,000 to $22,000) could impact the purchase probability. It indicates that an income increase at that level would only raise the probability of purchase by 2%. However, a 10% variation at middle income ranges ($40,000 to $44,000) could have up to a 9% impact on the propensity to purchase the product.

IMPLEMENTING THE PROCESS

Let's start from the beginning. You have a great idea for a new product and decide to send out a "intent to purchase" survey. The survey consists of questions revealing the prospect's income, marital status, age, geographic location, and predisposition for a variety of products. The respondent is also asked how likely they are to purchase your product at a particular price. These types of surveys indicate stated preferences of the consumer. Based on the information available, a logit regression model could be used to determine what customer or product related characteristics are important in the decision to purchase the product. What we get is a statistical picture of the optimal customer. Maybe he is a single 25-30 year old male who recently purchased a house and has an income greater than $50,000.

The next step is to develop a prospect list for the sales force or direct mail campaigns. In our example, this is done by collecting income, age, and other profile information on potential customers. Some data may have to be purchased from outside sources like Equifax. For each prospect, the values for the key characteristics are applied to the logistic model that was estimated from the survey data. The model is then able to calculate the probability of purchase for each candidate. When completed, a list can be generated which ranks all prospects from highest to lowest in terms of their potential to purchase the product. The lists are then made available to telemarketers or to individuals responsible for promotional campaigns.

The next phase of the process might be to see how well the characteristics of those who said they would buy match those who actually purchased the product. In other words, what people say they will do and what they end up doing may be two different things. Results obtained from historical data are called *revealed preference* studies. Evaluating revealed preference behavior is done the same way as in the stated preference case. Simply design another logit model to test for significant profile characteristics in the new historical data.

AN EXAMPLE

One firm recently implemented a trial using this approach in their small business marketing division. The purpose of the trial was to determine if the development of a logit model using revealed preference data could improve target marketing efforts of the department. A logit model was developed from historical data comprising over 10,000 observations — some who actually bought the product and others who decided against purchase. A number of explanatory variables were identified to contribute to the historic purchase decision. In particular, the customer's two digit S.I.C. classification was found to play an important role in purchase decisions. The logit model was then used to produce a list by ranking each observation from highest to lowest based on the expected purchase probability.

The model's value was tested through a variety of simulations using the historical sample data. A method often used as a starting point simply compares the effectiveness of two lists in identifying those individuals who purchased the product. The first list was developed by randomly selecting customers from the sample data — some who bought the service and some who did not. Because it was randomly generated, no insight into purchase behavior was used. The second list was generated by taking the same sample and computing the probability of purchase for each observation. The list was then sorted from highest to lowest based on the probability score. By comparing these lists side by side, we can evaluate which list is more effective in identifying customers who have purchased the product. Chart 2 shows that after examining 20% of the random list, about 1,000 customers were found to have purchased the product. The list generated from the logit model did better, identified about 1,500 customers who bought the product. Of course, when both lists were fully examined, all purchasers (5,500) were identified.

Instead of using a randomized list for benchmark comparisons, perhaps a more meaningful approach might be to compare lists derived from intuition or sales experience. In our example, maybe intuition tells us that selling efforts directed at S.I.C. 59 (Miscellaneous Retail) have proved worthwhile. A performance comparison could be made by generating a list from the sample data that contained all businesses with industrial classification 59. Maybe that results in 1000 prospective businesses. This list could then be compared to the first 1000 prospects generated by the logit model to determine which method would identify the greatest number of purchasers.

CONCLUSION

The modeling framework discussed provides an effective mechanism for target marketing. By scoring and ranking each prospect through the methods described, a list can be produced which should aid sales personnel in understanding who their customers really are. Since the process can only be as good as the data collected, efforts should be made to obtain quality profile information and a tracking procedure which identifies the marketing channel through which sales are made. Periodic updates can then be made to benchmark and revise profile characteristics and check the effectiveness of market channels.■

CHART 2

**BENCHMARK COMPARISONS
RANDOM LIST VS. LOGIT MODEL**