Getting your money's worth: virtual targeting

It is all about 'bang for the buck'. **Michael Lieberman**, Multivariate Solutions, explains a model that helps you to reach the right audience more efficiently

'IT IS ALL ABOUT 'bang for the buck'. You have a database of, say, ten million voters. Or ten million consumers. Or one hundred thousand association members. The database is full of 'goodies'. Not only the normal stuff, such as demographics (gender, age, income, and so on), and political information (party affiliation, donations given, primary and general elections last voted in), but a wealth of other personal information. For example, whether the list holder rents or owns his home, the number of private schools in the area, whether there is a working woman in the household, the number of children in the home, whether they have a DVD or have contributed to a health or environmental organisation, and whether they own a sports utility vehicle or subscribe to magazines.

TABLE 1

Swing voters: sample

- Age under 40 = 3, 41–65 = 2, 65+ = 1
- Attend religious services regularly
- Bank credit card
- Contribution to environmental group
- General election 2000
- General election 2002
- High private school attendance district
- Household party 2 = mixed, 1 = republican,
  3= democrat, 0 = no affiliation
- Married 1 = married 2 = other
- Party registration
  (1 = Rep, 2 = Dem, 3= Ind)
- Primary 2002
- Primary election 2000
- Subscribed to one or more magazines or publications, responded to a survey or entered a sweepstakes within the past two years.
- Unemployment rate in Utopia
- Working women in household

An election is looming, and in effort to reach swing voters or energise those who potentially support your core issue, your organisation would like to hit those potential targets with a direct mail piece or a telephone call. Something that will energise or sway them. Or a credit card company would like to mail out a sampler to a million or so homes, but wants fewer people to discard the piece without so much as a glance.

Let us say you are trying to reach swing voters in the state of Utopia, a swing state where things are not perfect. We know from those existing records that we expect about 20% of the list to be swing voters. You could mail out a flyer to all ten million, knowing that the hit rate is about one-in-five. Or you could target only swing voters and dramatically raise your efficiency and lower your costs. Only you don't know who they are.

So you can't actually target them. But you can build a model and make a very educated guess: you can 'virtually' target them.

In addition to your primary database, you have, say, 10,000 records in which you are able to determine your target group. These records could be drawn from other lists, primary research where ID numbers allow you to identify back to the main list, or company databases. These records give you the ability to build a link between 'swing' voters and characteristics to help define them. What virtual targeting will do is to build and test a profile of who your target is.

Basics of virtual targeting

What distinguishes our target group, swing voters, from non-swing voters? Are there characteristics that could be used to find out their likely perspective? How can we make, on the basis of several individual attributes, one assessment on the likelihood that a given person is 'swing'?

Virtual targeting answers these questions using a blend of statistical techniques that:

1. identifies distinguishing characteristics of the target group, and then

2. builds a linear equation that can be applied to each of our ten million records to calculate a 'score'.

When sorted, the hope is that the group with the highest scores will be the more likely swing voters.

TABLE 2

Discriminating factors

Variables present in model equation	Standardised discriminant coefficients
Party registration (1 = Rep, 2 = Dem, 3= Ind)	0.31
Contribution to environmental group	0.10
Working women in household	0.10
Unemployment rate in Utopia	0.04
Primary 2002	0.00
Attend religious services regularly	–0.05
Married 1 = married 2 = other	–0.05
Subscribed to one or more magazines or publications, responded to a survey or entered a sweepstakes within the past two years.	–0.06
Bank credit card	–0.08
High private school attendance district	–0.08
Age under 40 = 3, 41–65 = 2, 65+ = 1	–0.8
Household party 2 = mixed, 1 = republican, 3= democrat, 0 = no affiliation	–0.15
General election 2002	–0.28
General election 2000	–0.29
Primary election 2000	–3.6

The first step

The first step is to take the myriad of variables available to us and, using our known swings, discover which variables distinguish our target group from our non-target. There are two techniques that can be applied: regression analysis

Michael Lieberman is founder and president of Multivariate Solutions, a statistical and market research consulting firm that works with major advertising, public relations, and political strategy firms.

and CHAID (Chi-square automatic interaction detector) – a chi-square technique that creates a tree. The variables at the top of the tree are the most useful to distinguish between swing/non-swing, and as the variables run down the branches their importance diminishes. Still, a variable that emerges in the top five to six branches is a good candidate for the final model.

A detailed explanation of regression and CHAID analysis is beyond the scope of this article. Basically, what each does in this case is to create a baseline variable – a measure of association – between our target group and characteristics represented by literally hundreds of variables available in the entire database. A weeding out process.

In our example, the swing voters of Utopia, we have run through the first step in the virtual targeting. We have run a regression and CHAID, and the attributes listed in Table 1 have come up as significant.

Not surprisingly, many political variables, such as election frequency and party affiliation, have made it into the model. After all, what we are looking for is a potential politically neutral block of voters. Naturally, those who are not politically neutral (for example, primary voters), would be an evident distinguishing variable.

However, other not-so-obvious demographic and social attributes made it into the model. For example, if the person is married, lives in a district with high private school attendance and has a bank credit card, chances are that his or her 'swingness' can be more easily identified.

Step two

Now we know what should be placed into the model. Either the initial regressions or CHAID trees have told us. So, the next step is to run the model.

There are a number of multivariate techniques that can be used for this. They can have fancy names, such as logistic regression or forecasting membership by way of using an exponential probability model. These work in the proper situations, and sound pleasingly fancy to satisfy the clients that we are adding enough 'sauce' to the equation, so that it will be sophisticated. In truth, many credit card companies run these techniques on their enormous databases with terrific results.

However, Utopia is a state that likes its meat and potatoes. And, to be fair, our interest is to produce clear results that can easily be back-coded to the main list. So here I choose to use discriminant analysis, a multivariate technique that measures our input variables and produces coefficients that give us a measure of how much each attribute discriminates between swing voters and those who are committed.

Discriminant analysis produces a 'discriminant function'. That is, a linear equation where coefficients are multiplied against the respondent attributes to produce a score. Derived from the discriminant score, a likelihood of each group membership (ie swingness) is calculated, based on people who we know are swing voters from the smaller sample. To put it simply, the respondent fills out the form and gets a score, which is then compared to a chart to see if s/he has a good chance of being a swing voter.

TABLE 3

Calculating the discriminant score

Variables present in model equation	Raw discriminant function coefficients	Coded Answer to Question	Sum of raw discriminant
Party registration (1 = Rep, 2 = Dem, 3= Ind)	0.75	1	0.75
Contribution to environmental group	0.73	0	0.00
Working women in household	0.63	1	0.63
Unemployment rate in Utopia	0.19	1	0.19
Primary 2002	0.16	3	0.47
Attend religious services regularly	0.13	0	0.00
Married 1 = married 2 = other	0.13	1	0.13
Subscribed to one or more magazines or publications, responded to a survey or entered a sweepstakes within the past two years.	0.12	7	0.83
Bank credit card	0.11	1	0.11
High private school attendance district	0.00	1	0.00
Age under 40 = 3, 41–65 = 2, 65+ = 1	0.00	1	0.00
Household party 2 = mixed, 1 = republican, 3= democrat, 0 = no affiliation	–0.02	4	–0.6
General election 2002	–0.04	3	–0.11
General election 2000	–0.23	1	–0.23
Primary election 2000	–0.36	2	–0.73
Discriminant Score		1.9950	

As in all sophisticated statistical analyses, a blizzard output accompanies the procedure. There are three outputs that we need to examine: the beta scores of the discriminant function (known as the raw coefficients); the standardised coefficients (which tell us which are the best variables); and the discriminant score, coupled with the percentage likelihood that that score describes a member of our target group – swing.

The raw and standardised coefficients are used for descriptive and classification purposes, which I will cover below. The discriminant score, when calculated afterwards, is the instrument used for future classification.

In the virtual targeting model there is one more measure which is not necessarily used with discriminant analysis. We are looking for the strength of the model as it specifically applies to identifying swing voters. The method is straightforward. The software, after it runs the analysis, gives each respondent in the analysis a score. We sort the list from highest to lowest score, then look at, say, the top 10%. The idea is to see how much better the sorted list is than a random sample. For example, with our Utopian list, we expect 20% of voters to be swing. If we take the top 10% of our sorted list, and 30% or them are identified swing voters, we can see that our list is 50% more efficient than a random sample.

TABLE 4

Discriminant scores table

Score	Probability (%)
2.65	82
2.64	80
2.61	80
2.59	79
2.56	79
2.37	76
2.36	76
2.00	61
1.99	61
1.42	50
1.41	79
0.00	33
–0.01	33

TABLE 5

How many voters are swing voters?

% of total sample	% of sample that is swing voters	Index (base = 20%)	% better
0	0	0	0
10	41	205	105
20	35	175	75
30	31	155	55
40	29	145	45
50	27	135	35
60	24	118	18

What happens next?

Table 2 shows, not surprisingly, that the most discriminating factor is that the person is an independent. That is, not a member of either political parties. The two other telling factors to determine swing are that the person contributes to a religious group and that there is a working woman in the household.

Unsurprisingly, if a person has voted in the primary election in 2000 or in a recent general election s/he has a high negative coefficient. It is unlikely that s/he is a swing voter.

Virtual targeting is both descriptive and predictive. The descriptive side, illustrated in Table 3, explains which factors rise to the top (or bottom) when running the model. This can be very interesting information. However, the real power of the technique lies in the simple ability to predict a person's group. This is where the real 'bang for the buck' comes in.

Table 3 also illustrates how a given person receives a discriminant score. The raw coefficients (not standardised, as above) are multiplied by a respondent's answer, then tallied to create one 'score'.

At the bottom of the table this example's score has been calculated. It is 1.9950. So, is that good?

The final useful output in our example is a list of all discriminant scores and the probability of that score's respondent being a swing voter. This output can be sorted and displayed in a table which is partially shown in Table 4.

This functions as a 'look-up' table. When one person goes through all the survey and has their score totalled, that score can be compared to scores on this look-up table to see what is the percentage chance that person is a swing voter. In our case, 1.9950 has a 60% chance of being swing.

The last, most important question: how good is the model? Would it make sense to score all ten million?

Reading Table 5 from left to right goes like this. When the known swing voters are scored, and the scores are sorted highest to lowest, what percentage of the top 10% are swing voters? The answer, according to Table 5 (second column), is 41%. We would expect one-in-five (20%) to be swing if people were just randomly selected. So, if you divide 41/20, you get 2.05. Or, in other words, the model has more than doubled the efficiency of finding swing voters. The index, which multiplies this number by 100, is 205. That is high.

If you then look at the top 20% of the sorted sample, 35% of those are swing. Or, the model is 1.75 times more efficient, with an index of 175.

As we work our way down the sample in order of score, the efficiency lessens. This is to be expected, since lower scores indicate less likelihood of being a swing voter.

Think about it. The organisation is sending out one million pieces. If it does not run the virtual targeting, it can expect to reach about 200,000 swing voters.

If it does run the virtual targeting, and applies the scores to the general database, it can expect to reach 410,000 swing voters spending the same amount of money.

That's 'bang for the buck'. ■

michael@mvsolution.com