Measurement and Tracking - The Tactical Perspective

We recently discussed the importance of measurement and tracking as a key component within the data mining process. Besides the importance of being able to analyze results arising from these tracking reports, we also mentioned the importance of understanding the overall business objectives of a given marketing campaign. It is these objectives that ultimately provide the framework in which to design and create these reports. These reports could be designed to provide both strategic as well as tactical information. From the strategic viewpoint, we demonstrated, albeit in a simplistic manner, how one could create sensitivity reports which demonstrate the return on investment (ROI) impact of introducing CRM programs. We also viewed these reports in a number of different ways by introducing different assumptions such as the introduction of a targeting model.

The intention of this article is to discuss measurement and tracking from a tactical perspective. How do we set up the appropriate testing matrix for a given business initiative or marketing campaign? What kind of learning do we want to obtain and how do we evaluate the performance of a given initiative or campaign? For instance, if we are deploying some kind of targeting tool within a campaign, we need to obtain learning about its performance. However, it is the amount and complexity of the required learning which will dictate the design of the measurement and tracking system. For instance, the simplest kind of tracking is to simply determine the performance of a particular initiative. The initiative could be anything such as an event, campaign, type of offer, type of communication, targeting tool, etc. In this case, the tracking requirements are quite simple in the sense that two cell codes/groups are created. The first cell code or group is the control cell where there is no occurrence of the initiative while the second group or test cell is impacted by the initiative. This is rather easy if we are interested in only one piece of learning from a campaign. Yet, in practice, marketers are always trying to obtain as much learning as possible. In setting up a measurement matrix, complexity is not the issue; rather, the real issue is costs. For each given initiative, test cells need to be set up. Suppose that we want to evaluate four different offers along with a control offer. Five cells in this case would need to be created.

Offer Type	
	Control
1	
2	
3	
4	
Control	

Now suppose that we wanted to test four different offers along with four different communication pieces. More importantly, suppose we wanted to test the interaction of offer and communication. Consequently, we would need to create 17 separate cells (4 offers x 4 communication types + 1 control group). Listed below is a schematic of this tracking matrix with live promotion quantities.

Offer Type		Communication Piece			
	Control	A	B	C	D
1		5,000	5,000	5,000	5,000
2		5,000	5,000	5,000	5,000
3		5,000	5,000	5,000	5,000
4		5,000	5,000	5,000	**200,000**
Control	5,000				

In any tracking scenario, there is always one cell which the marketing team has to select as its winning cell. Marketers will place as many of their names here since this is the cell which is expected to generate the largest ROI. In our example here, the winning cell is D4 while the other 16 cells essentially represent the investment costs of promoting 80,000 names or 5,000 names per cell. This can be considered a significant investment since 20% (80,000/280,000) of the promotion quantity is devoted to learning. At a $1.00 per piece, this can translate to $80,000 in investment devoted to learning.

Suppose, we wanted to introduce other factors into the mix such as region (four groups) and targeting (four groups), the tracking matrix in our previous example would then explode into a 257-cell matrix (4 offers x 4 communications x 4 regions x 4 target groups + 1 control group). Obviously, this becomes unwieldy.

This is where the use of statistics such as conjoint analysis or Analysis of Variance (ANOVA) can be used to reduce the number of test cells to a more manageable number. We are not going to delve into the mathematics behind these techniques other than to state that these techniques are useful in providing some science regarding the interaction between factors. Suppose that conjoint analysis/ANOVA demonstrates that there is no interaction of targeting and region with the other factors. Then, the following matrix could be produced:

Offer Type		Communication			
	Control	A	B	C	D
1		X	X	X	X
2		X	X	X	X
3		X	X	X	X
4		X	X	X	X
Control	X				
Region					
Maritimes	X				
Quebec	X				
Ontario	X				
Rest of Canada	X				
Targeting					
Segment 1	X				
Segment 2	X				
Segment 3	X				
Segment 4	X				

As you can see here, through the use of statistical analysis, we have reduced the number of cells from 257 to 25 cells. This is another illustration of the role of statistics within database marketing programs. Yet, in this scenario, its objective is not the creation of targeting tools but rather determining what cells need to be tested within a given marketing campaign.

Another question that often gets asked when designing a testing matrix is the sample size that should be allotted to each cell. The answer to this question, like to most other statistical questions, is "it depends." By this answer, we mean that there are a variety of factors that impact sample size. Factors such as the overall performance rate, error range and confidence level all influence sample size. The actual formula can be written as:

$$\textbf{Sample Size} = \frac{\textbf{Confidence Level}^2 * \textbf{Performance Rate} * (1 - \textbf{Performance Rate})}{\textbf{Error Range} * \textbf{Error Rate}}$$

Let's take a look at some examples to obtain some better insight on how this formula actually works. Assume we have a response rate of 1% (performance rate), a confidence level of 95% (1.96 z-score), and the fact that we want the results to be statistically significant if they are 0.2% different from the mean (error range), then our minimum required sample size is simply

$$\textbf{Sample Size} = \frac{1.96^2 * 0.01 * (0.99)}{0.002 * 0.002} = 9{,}508$$

If we increase the allowable range of error to 0.4%, then intuitively we would expect the minimum required sample size to decrease. In fact, it does decrease to 2,377. Increases to the confidence level will always increase our sample size. For instance, increasing the confidence level from 95% to 99%, the sample size increases from 9,508 to 16,475. For those of you who are quick with numbers, a performance rate of 50% always yields the largest sample size.

This formula can be plugged into any spreadsheet application. It has tremendous utility to the marketer in its ability to provide a range of required sample sizes based on the differing assumptions of performance rate, range of error and confidence level.

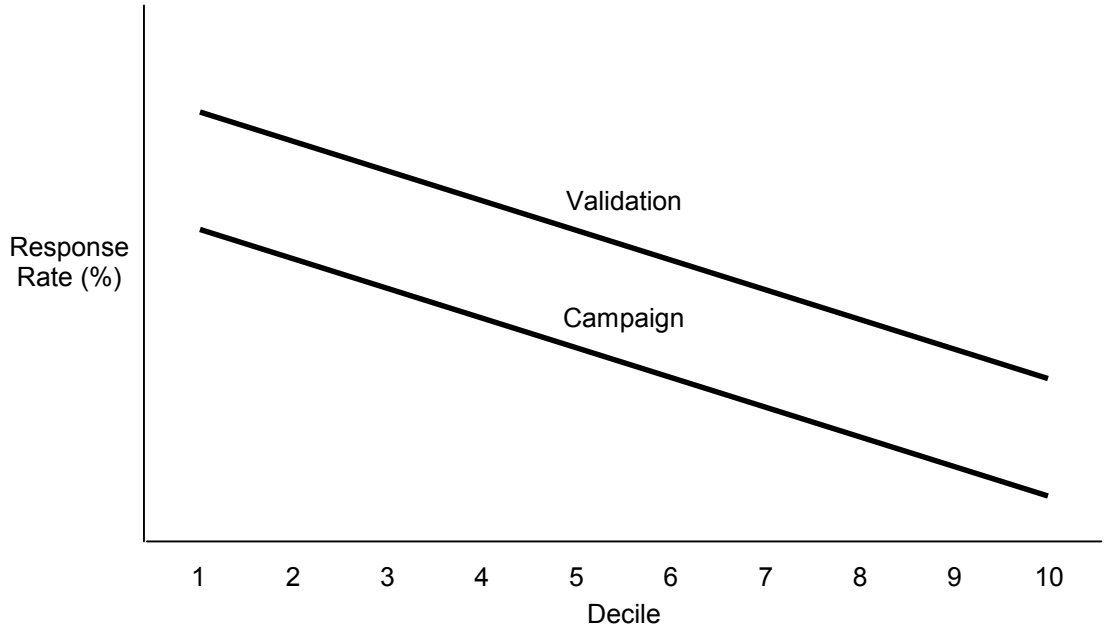
We have discussed at length the tracking and measuring of different business objectives. But suppose we want to be more granular in our results performance by determining the optimization of a given business objective. This scenario is especially applicable to targeting tools that are used within a marketing campaign. With a control or random cell set aside and then used in comparison to the targeted group, we can quickly ascertain if the model is effective by observing if the results are superior to the control group. But is it optimal? Is the targeting performing as what was observed when we built the tool? This is where we need to create some kind of comparison between when the model was built and its current application. By checking the performance results of the model when it was initially developed, we then have a benchmark in which to compare subsequent targeting application results.

For instance, suppose we built a response model whereby customers are ranked into 10 deciles with decile 1 being the highest scored and decile 10 being the worst scored. The observed response rate from the validation sample for model development is then reported for each decile and plotted as a line or curve. This line or curve is often referred to as a Lorenz curve. The

objective of any model or targeting is to maximize the slope of the Lorenz curve. A flat line is a complete failure while a vertical line represents perfection.

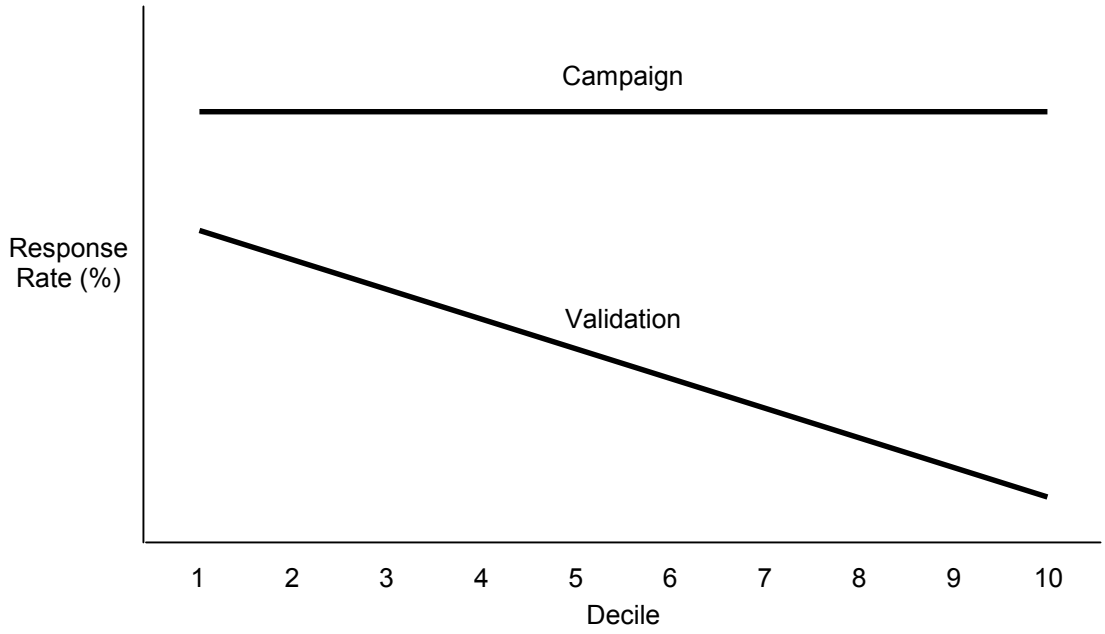
Once this model is applied to a current campaign, we can conduct the same above exercise and create a Lorenz curve for the current campaign. We then have a means of determining optimality by comparing the slopes of the two Lorenz curves. If the slopes are identical, then the model is performing optimally. If the slope of the campaign is less than the slope for model development, then the model is not performing optimally. The key business decision in this type of situation, though, is to determine whether or not we can live with this sub-optimal performance or invest in the redevelopment of a new model.

Let's take a look at a couple of examples to better illustrate this concept. Listed below is an example of two Lorenz curves, the first one reporting the results during model development (Validation) and the other one reporting the results after the model was applied for a campaign (Campaign).



Despite the fact that the overall campaign's response rate is performing much worse than what was observed during model development, the targeting tool is performing very effectively as its ability to rank-order response rate has more or less remain unchanged between model development (validation) and its current application (campaign). Since the slope of the two lines is identical, we can conclude that the model's current performance is optimal. In this particular case, it was critical to evaluate the model's performance in a granular manner since the first inclination of the business team might be to blame the model for the overall campaign's poor performance.

Listed below is another example of reviewing a model's performance during its current application.



Here we observe that the overall campaign is performing much better than what was observed during model development. In this particular case, the first inclination might be to credit some of the campaign's success to the model when in fact the model provided no contribution towards its success. This is due to the fact that there is no rank-ordering of response rate by decile as demonstrated by the flat line Campaign Lorenz curve. The slope of the Campaign Lorenz curve in this case is almost non-existent. The business decision in this case would be rather simple in that we definitely need to redevelop a new model.

These are just some examples of what one might consider in designing a measurement and tracking system. The key towards the creation of any measurement and tracking system from a tactical standpoint is to establish your priorities and objectives and plan accordingly. Yet, this planning can be irrelevant if we do not properly understand how to use the data within the existing information environment and more importantly how to use it in a cost effective manner.