

Analytics Using SAS® Customer Intelligence: Multivariate Testing Processes for Digital Marketing Campaigns

Malcolm Lightbody, Todd Barlow SAS Institute Inc.

ABSTRACT

Multi variate website testing helps marketers understand which combinations of factors, working together, are most important for driving conversion. A real-world objection to running multi variate tests is the volume of traffic needed to draw reliable conclusions. This paper introduces a novel and practical method for doing multi variate web site testing. It is based on reducing a full factorial test to a smaller number of design points and then fitting a predictive model to estimate outcomes of untested combinations from measured results. The marketer can then choose to follow up the multi variate test with a subsequent A/B test once a small number of highly performing combinations have been identified.

INTRODUCTION

As a mantra, “fail fast, fail often” (Babineaux, Krumboltz 2013) is only meaningful if failure contributes to lessons learned. Failure on its own is not enough. Being able to differentiate between success and failure is a critical part of learning. And being able to monitor the effect of actions on outcomes reinforces the good and provides contrast with the bad.

Some business environments lend themselves more easily to measurement than others. Digital is one area that is intrinsically measurable. Websites, mobile apps and emails can be dynamically built so that different versions, with varying content, can be presented to different users. This can be done in real time for web, mobile and social channels.

Testing different versions and combinations of content is a standard tool of the digital marketer. At its heart, such testing is conceptually simple. Try out N versions of a page or piece of content, count how many visitors see each version, and, of those, how many react in a meaningful way (convert). A whole industry has arisen to help the digital marketer understand how to interpret the results of such tests and how to translate these results into actions that add business value.

A/B testing is the prominent methodology, and allows the digital marketer to compare the effect of changing a single page element on achievement of a business goal. This element could be the content of a spot on the page, different color backgrounds, different font sizes, different versions of a call to action and so on. Each of these elements is known as a ‘variable’ and the different options are levels of that variable. The defining characteristic of A/B testing is that it allows one variable to be tested at a time.

Often, a marketer is interested in changing more than one variable at a time, because they believe that there is an interaction between variables that results in a combined effect that is greater than the sum of the parts. Multi-variate testing is the tool of choice for measuring these interactions. However, as illustrated below, multi-variate testing has practical disadvantages because of traffic volumes required for statistical significance.

To see the practical challenge associated with multi-variate testing, consider a page that has four content spots (variables) and each spot has four different pieces of content that could be displayed in it (the levels). The number of unique versions of that page is $4 \times 4 \times 4 \times 4$, which is 256 and represents the so called ‘full factorial’ design. And the size of this full factorial design increases rapidly with the number of variables and levels.

Suppose that, to provide reliable results, each of the 256 versions of the page must be shown to at least 10,000 visitors. That means that to test every combination would require a total of over 2.5 million visitors. Such a test would take many months if not longer on a typical web site. By which time, the results will no longer be relevant.

Marketers wishing to measure the effects of different combinations of content are then forced to face these limitations and either adopt an alternative experimental design methodology or severely limit the

size and scope (and learning opportunity) of this test, or to abandon it altogether and opt for multiple A/B tests instead. Multiple A/B tests still only capture the effect of one variable at a time and cannot measure interaction effects.

SAS has developed a solution to the problem of optimizing the design of multi variate tests in the face of limitations caused by cost constraints. Originally designed for applications in manufacturing, where each 'variant' is expensive or hard to perform, this solution allows a multi variate test to be optimized by minimizing the cost whilst still requiring a minimum level of reliability (SAS Institute Inc, 2014).

This paper describes the application of this methodology to solving the scalability problem associated with on-line multi variate testing and allow marketers with smaller traffic volumes to benefit from the resulting insights. The "cost" in this case is the traffic (or time) needed to try out each variant. The solution described below uses optimization to minimize the number of variants and automatically select the best sub-set from the full factorial design. Furthermore, the user can easily create follow-up A/B tests using the top performing variants if the multi-variate test results in a close tie.

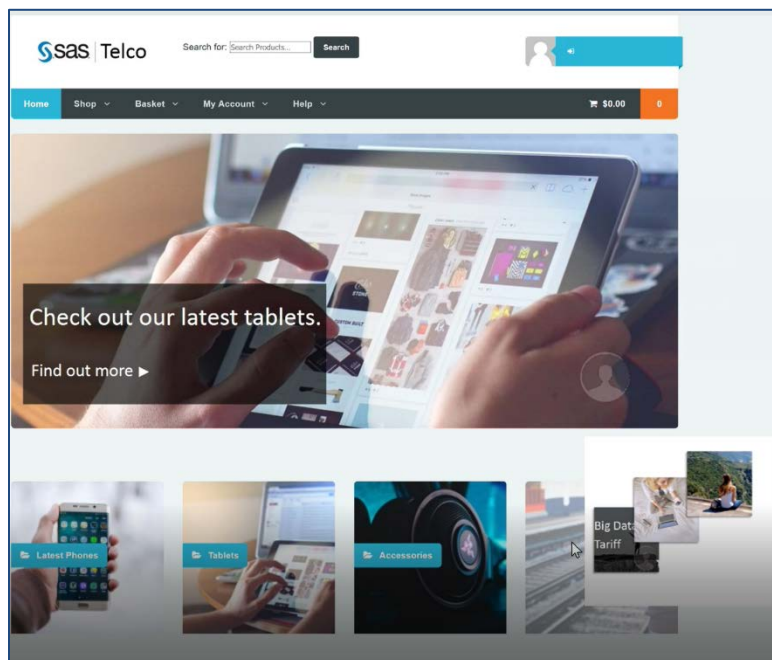
MULTI-VARIATE TESTING IN SAS CUSTOMER INTELLIGENCE 360

SAS Customer Intelligence 360 introduces the concept of analytic Blue Boxes. Each Blue Box is a self-contained embedded applied analytic solution to a specific business problem. Multi-variate testing is one of the analytics Blue Boxes in Customer Intelligence 360. It has well defined inputs and outputs and its internals are designed so that the marketer gets value from day one.

Making advanced analytics accessible and actionable is a key tenet of Customer Intelligence 360. The marketer can simply focus on what they want to achieve from a business perspective. The next few sections describe the marketer workflow as he or she defines, sets up, optimizes and analyses a multi-variate test.

SETTING UP THE TEST

As with all testing exercises, the starting point is to determine what the purpose of the test is. In this example Kelly, the web marketer, wants to figure out what combination of content leads to greatest conversion from the home page. Kelly works for a telco company, and they offer the typical range of products and services. The home page contains four content spots. Kelly is interested in understanding how different combinations of images across the spots interact to influence conversion.

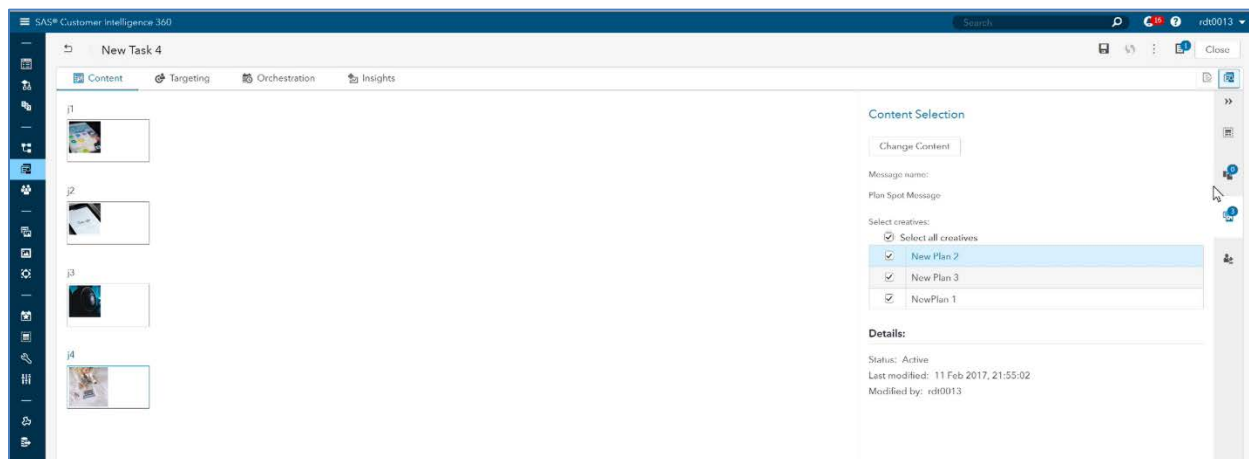


Display 1. Home page of the Telco website, with four content spots in the lower half of the page.

Each spot has a different purpose. One promotes handsets, one tablets, one accessories and the last one different plans. Two of the content spots have three candidate images and two of the spots have four. Meaning that there are one hundred and forty four unique combinations of content across the spots.

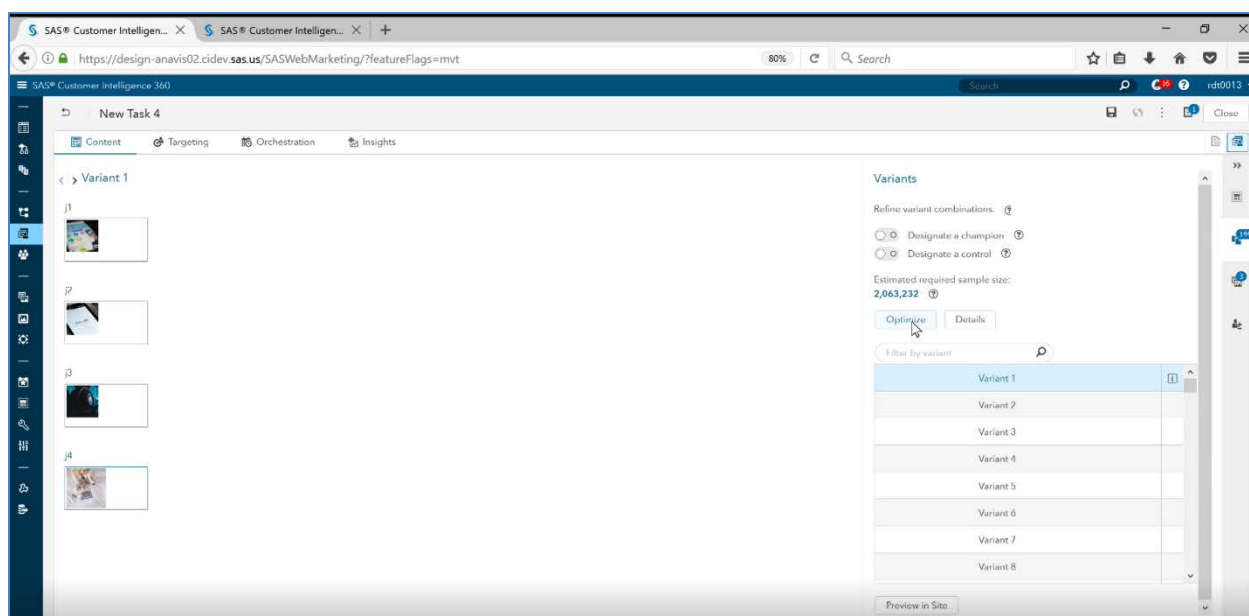
The basic unit of interaction within SAS Customer Intelligence 360 is a task. Tasks target one or more spots and are used to place content within. Decisions on what content to show can be based on rules and also on analytics. Since a multi-variate test consists of multiple variables, a multi-variate test task targets a spot set rather than a single spot.

Before setting up a multi-variate test, Kelly needs to ensure that a spot-set including the desired target content areas has been set up. Once that is done, then defining the test is straightforward. Kelly creates a new multi-variate test task and selects the message for each spot.



Display 2. Setting up the content for the spots. Each of the four spots are shown on the left hand side and the list of candidate creatives for each spot on the right hand side.

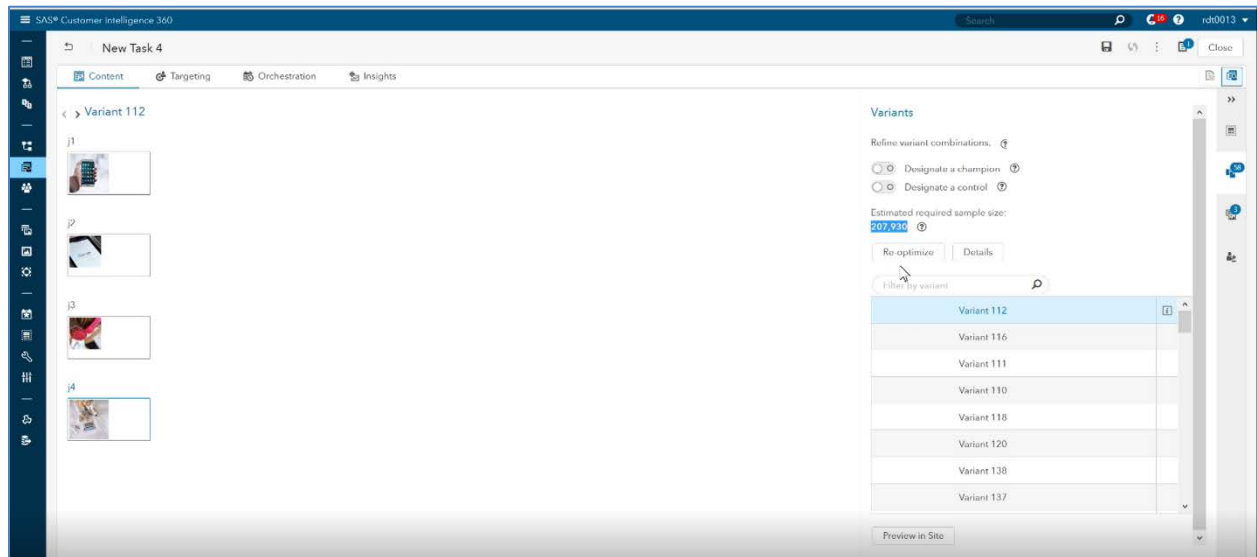
A message contains one or more creatives (in this case, each creative is a different image). Kelly defines the candidate images for each spot. Once the creatives and images have been selected, Kelly can generate the base full factorial design.



Display 3. Generate the initial full factorial design. This also includes a calculation of the required sample size for the full factorial design, in this case over two million impressions.

EXPLORING THE OPTIONS

Kelly now has several options available. The first is to automatically optimize the design of the test. On pressing the button, the system automatically analyses the problem and works behind the scenes to determine the minimum number of variants (combinations of content) that can be used to draw reliable conclusions from the test. Kelly can then view the design and simply start the test.



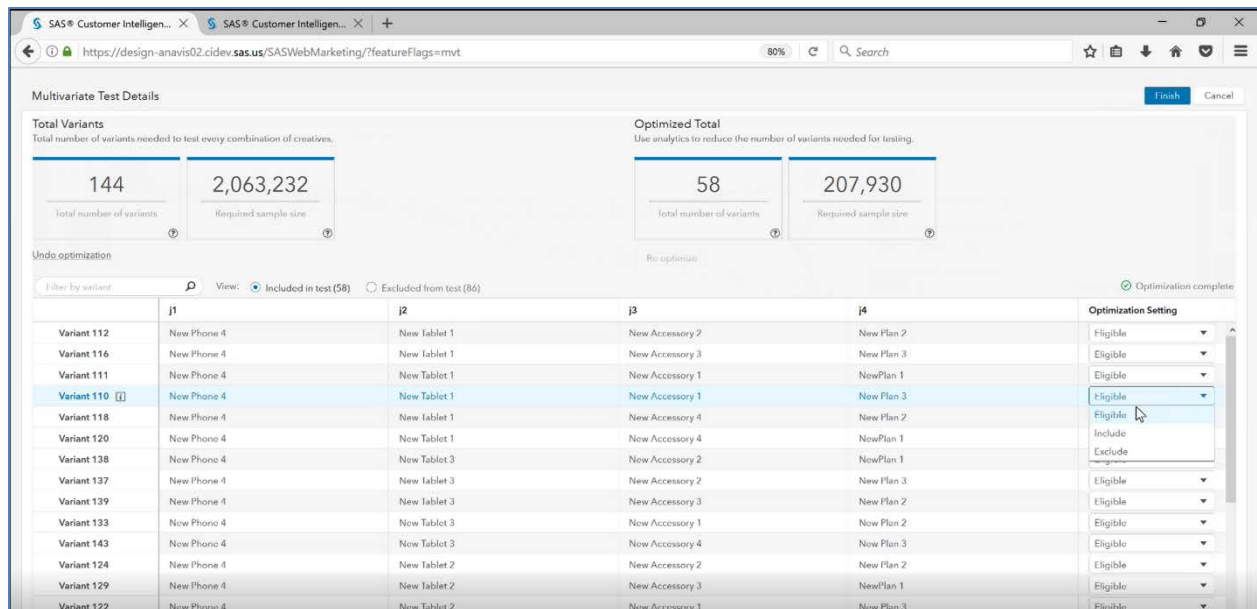
Display 4. The automatic optimized design.

Kelly can cycle through the different combinations of content by clicking on the variant number in the top left hand corner. These are the variants that test participants will see during the test. It is not the full set of variants from the full factorial. However, we will be able to predict conversion rate for all variants in the full factorial. Note the new sample size of just over two hundred thousand, which is about ten percent of the original. This means that the test, which was initially not practical, is now feasible to conduct.

OPTIMIZING THE TEST

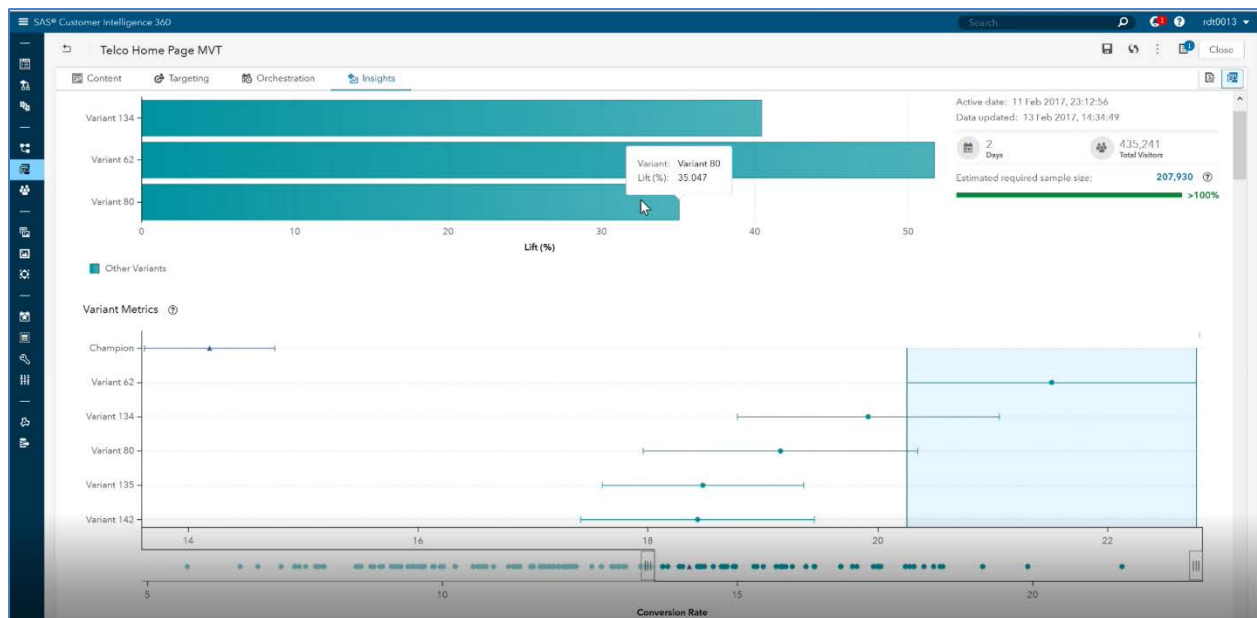
Kelly can also work with the multi-variate test Blue Box interactively. By selecting the “Details” button above, Kelly gets to add constraints to the optimization. For example, Kelly can decide that specific variants (combinations of content) must always be shown, and that other combinations must not be shown together. Prior to every optimization run, Kelly can explicitly include or exclude any variant. Re-optimizing then forces these variants in (or out) of the solution.

Once Kelly is satisfied with the setup of the experiment, the task can be saved and put into operation. As the task is running, Customer Intelligence 360 automatically shares out traffic amongst the variants to be tested, per the automatically calculated sample sizes and weightings. Progress of the test can be monitored by opening and viewing the insights tab of the corresponding task.



Display 5. Scenario building – add constraints that explicitly include or exclude specific variants

INTERPRETING THE RESULTS



Display 6. Insights – the result of the test and follow up action

At any point, Kelly can open the task and monitor the results so far. Best practice is that a test should not be ended until the pre-determined minimum sample size is met. Progress towards the total is shown by the green bar in the top right quadrant of the screen. In this case, the test is complete, and the progress bar shows one hundred percent.

In the top left had quadrant Kelly sees a bar graph showing details of the top three performing variants. The chart in the lower half of the screen provides more details on the performance of each variant. The

horizontal axis represents conversion rate and the vertical axis shows variant number. The variants are ordered from best to worst performing.

The confidence interval round each measured conversion rate is shown and, for the best performing variant, this is also highlighted by the blue shading. Sometimes there is a clear winner and the action is clear – place this winning variant into production. Other times, as in the above example, Kelly sees that there is no clear winner. Variant 62 has the best conversion rate, but there is overlap between its confidence interval and those from variants 134 and 80. In this case, Kelly may decide to run a further “tie breaker” test. Customer Intelligence 360 allows the marketer to automatically deploy an A/B test including just the top N variants. This will allow the marketer to make a better decision as to what variant to place into production.

A feature of the approach taken for multi-variate testing in Customer Intelligence 360 is that the results from the tested variants are used to fit a model that predicts the likely outcome for the untested variants. The second part of the insights tab shows the combined (measured and predicted) results.

The screenshot shows the 'Insights' tab in the SAS Customer Intelligence 360 interface. The primary metric is 'Web Click Through'. A status bar indicates 'In Progress' with a note: 'Either the required sample size, lift, or confidence threshold have not been met yet'. The table below displays performance data for various variants, ordered by conversion rate.

Variant Name	Visitors	Conversions	Conversion Rate (%)	Champion Lift (%)
Variant 113		7,559	1,072	14.18 (±0.57)
Variant 62*				21.51 (±1.26)
Variant 134*				19.92 (±1.14)
Variant 80*				19.15 (±1.20)
Variant 135*				18.48 (±0.88)
Variant 142*				18.43 (±1.02)
Variant 90		7,472	1,369	18.32 (±0.60) ↑
Variant 126*				18.16 (±1.01)
Variant 22*				17.99 (±1.23)
Variant 18*				17.91 (±0.94)
Variant 143*				17.84 (±1.09)
Variant 133		7,483	1,304	17.43 (±0.62) ↑
Variant 114*				17.39 (±0.98)
Variant 98		7,560	1,313	17.37 (±0.62) ↑

Display 7. Insights – measured conversion rates have actual counts, estimates conversion rates have none

Kelly can select to see only the measured results or only the predicted values. When viewing both together, the actual results are, of course, those with actual views and conversion counts.

OPTIMIZATION DETAILS

SAS' PROC OPTEX is at the heart of multi-variate testing in Customer Intelligence 360. It was initially develop to free the scientist from much of the statistical burden associated with designing an experiment and focus instead on the science. This very much fits into the philosophy behind embedding analytics into Customer Intelligence 360 and allowing the marketer to focus on the business problem at hand.

OPTEX from SAS Q/C User Guide

The OPTEX procedure searches for optimal experimental designs (SAS Institute Inc, 2014). You specify a set of candidate design points and a linear model, and the procedure chooses points so that the terms in the model can be estimated as efficiently as possible. Most experimental situations call for standard designs, such as fractional factorials, orthogonal arrays, central composite designs, or Box-Behnken designs. Standard designs have assured degrees of precision and orthogonality that are important for the exploratory nature of experimentation. In some situations, however, standard designs are not available, such as when:

- not all combinations of the factor levels are feasible
- the region of experimentation is irregularly shaped
- resource limitations restrict the number of experiments that can be performed
- there is a nonstandard linear or a nonlinear model

The OPTEX procedure can generate an efficient experimental design for any of these situations. Website multi-variate testing is a suitable candidate for OPTEX because there are:

- Constraints on the number of variants that are practical to test
- Constraints on required or forbidden combinations of content
- Desired accuracy of the model fitting algorithm

Proc OPTEX is highly flexible and has many input parameters and options. This means that it can cover many different business scenarios and its use can be tuned as circumstances demand. The ones that are most relevant to its use for website multi-variate testing are the desired number of output rows (each output row represents one variant to be tested) and the model specification. In the future, some of the additional parameters, such as model type may be exposed to more expert users within Customer Intelligence 360.

ANALYTIC PROCESS FLOW

Several analytic steps take place behind the scenes during the setup and subsequent learning phase of the multi-variate test. These range from designing the test all the way through to building the estimator model that predicts what the conversion rate would have been for unshown variants. The following eight step flow describes this process at a high level:

1. **Create starting point full factorial design** - PROC PLAN to create Full Factorial Design, which are the inputs to the optimization step
2. **Calculate required sample sizes** - PROC POWER is used to calculate initial population size to conduct test. Also, used in cost savings comparison versus running multiple A/B tests
3. **Define constraints** – from full factorial design apply constraints on variants which would not be considered for optimal design matrix (ex. Black font on a black background)
4. **Additional setup** - apply any augmented rules to be applied (reference cells or champion cells)
5. **Optimize the design** – use PROC OPTEX to generate an optimal design matrix (a subset of the full factorial design). Sample size is recalculated based on this design.
6. **Run experiment** - distribute content based on optimal design matrix and collect results
7. **Build analytic estimation model** - use collected results and fit a predictive model. Conversion is the binary target of the model and the variant parameters are the inputs

8. **Score the estimation model** - apply the model to produce estimates for untested variants in the full factorial design

SAMPLE SIZE ESTIMATION

Sample size estimation in the MVT complex can be quite complex. To balance complexity, ease of use, and effectiveness in an operational setting, Customer Intelligence 360 uses an approach to calculating sample sizes based on SAS PROC POWER (SAS Institute Inc, 2008).

The starting point is the full Factorial sample size required to satisfy confidence and power constraints subject to the marketer provided minimum detectable difference. We are not looking for a specific single winning variant, rather that the experiment is designed with the goal as identifying a small number of variants as candidates for a subsequent A/B test. This means that we are, in effect, pooling variances of a small subset of variants. The effect of this can be estimated by making an adjustment to the minimum detectable difference.

This has a two-fold effect. First, the sample size for each variant is reduced by this pooled variance effect. Second, the optimization step reduces the full factorial design to a smaller number of variants. So the total sample size required for the test is significantly reduced. Together, these effects provide the marketer with significant business benefit.

MODEL FITTING

PROC OPTEX allows the user to specify a model to extrapolate measured results and allow untested variant performance to be estimated. The model form must be specified and the resulting optimize design is dependent on this. As a rule of thumb, OPTEX requires a minimum of ten design points (individual variants) plus one design point for a model intercept term and one design point for each interaction type in the model.

As an example, choosing a main effects model with the telco example from the previous section with four variables (content spots) leads to a minimum of fifteen design points. OPTEX chooses the best fifteen from the total one hundred and forty-four design points in the full factorial design. Obviously, the accuracy of the estimation model is dependent on the number of design points. A trade off needs to be found between accuracy and achieving the objective of reducing the size of the multi-variate test.

A stepwise logistic regression is used, with conversion as the target and the independent variables are the spots/levels. For example, in the telco case with four spots, each spot has several candidate images that can be shown. The combination used in each presented variant is recorded, along with the outcome. This input data set will have as many rows as the pre-determined sample size.

Variant ID	Spot1	Spot2	Spot3	Spot4	Convert
10001	Image1a	Image2d	Image3b	Image4d	1
10002	Image1a	Image2b	Image3e	Image4b	0
10003	Image1c	Image2c	Image3f	Image4a	1

Table 1. Example input data set for multi-variant testing model fitting

The output from the model is a scoring algorithm that then takes the image combinations for untested variants and predicts the conversion likelihood. These results are then presented back to the marketer.

CONCLUSION

Up to now, multi-variate testing has been limited to small experiments for all but the busiest web sites. Embedding SAS Proc OPTEX plus model estimation capabilities within Customer Intelligence 360 brings the power of multi-variate testing to more users, without requiring them to have intimate knowledge of design of experiment theory. Whilst multi-variate testing will always require larger sample sizes than simple A/B testing, the capabilities presented here show how many more practical use cases can be addressed.

REFERENCES

Babineau, R, Krumboltz, J. 2013. *"Fail Fast, Fail Often: How Losing Can Help You Win"*. Tarcher Perigee (December 26, 2013).

SAS Institute Inc 2014: "SAS/QC® 13.2 User's Guide: The OPTEX Procedure": SAS Institute Inc, 2014

SAS Institute Inc 2008: "SAS/STAT® 9.2 User's Guide The POWER Procedure" SAS Institute Inc 2008

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Malcolm Lightbody
SAS Institute Inc
malcolm.lightbody@sas.com
www.sas.com

Todd Barlow
SAS Institute Inc
todd.barlow@sas.com
www.sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.