

Art, Science, and the Modern-day Modeler

*Bradley Rukstales
President*

Customer Asset Consulting Group, Inc.

Well, I was working at home late one night on my laptop. My wife was in the other room. I have a habit of talking to myself as I work, clarifying my thoughts. After some time had passed, I exclaimed “look at the curves on this model!” to no-one in particular. I was overjoyed with the success I was seeing in the regression. My wife, however, ran over to see what *exactly* I was up to. We laughed, and I went back to work, looking at all sorts of models with all sorts of curves.

So, if you asked three statisticians to build a predictive model from the same set of data, you’ll get at least three different models. All of them may “perform” similarly, but may vary significantly in the types of variables included, or in the approach taken in building the model. Why would that be? It’s been a long-acknowledged understanding that modeling is part art, part science.

The science of modeling comes from the statistical toolkit that an analyst brings to the problem. This includes knowledge of different statistical procedures themselves (types of regression models, data reduction techniques, observation clustering techniques, etc.), the programming skills to make the statistical procedures hum in a business environment, and the ability to differentiate the world of academia from that

of the real world of messy data. A thorough knowledge of sample size requirements and statistical significance measures is also a key requirement of any modeler.

The art of modeling arises in many forms. One is in the training and background of the analyst. For instance, in a traditional regression model, econometricians and psychologists pay much attention to the interpretation of *betas (the coefficients)*, pure statisticians keep their eye on the *error*, and business analysts are most concerned with the prediction (the left hand side of the equation). These are broad generalizations, and a good analyst is thinking of all of these things to some degree, but it is a good indicator of the types of mindsets that modelers bring to the table.

The art of modeling is also a function of experience. There’s a level of intellectual curiosity that is required to become a good modeler. Every project starts with a blank screen, a dataset, and a goal. Years of intellectual curiosity, forming an individualized point of view through trial and error, leads to a unique method and process for each modeler. Yes, there are moments where time dictates that one plugs and chugs through a model using nothing but the tried and true. However, few modelers can handle a steady diet of these

projects before demanding an opportunity to explore new and challenging frontiers.

This also leads to a process of continuous improvement. Analysis methods that worked in one project are often duplicated in other projects, usually through the building of a tool kit of programs (macros) that improve the efficiency of analysis. One such example could be a program that reviews variable transformations and identifies the best transformations for inclusion in the model. Another would be methods of identifying important interaction terms for potential inclusion in a model. Through this process, the quality of the models should improve, if the methods are based on both statistical and empirical results.

All modelers and analysts face the issue of messy data. In fact, the task of building a model is relatively minor compared to the demanding undertaking of putting the data in the proper form for analysis. Even within companies where the data is stored in easy-to-access tables, well-householded, and updated and cleaned regularly, the issues of sample selection, time-dependencies, and level of summarization of detailed transaction information will require time and energy to define and program.

Yet, in this, too, there is art. Do we try to include seasonality in a model? Geography? What level? Do we include prior contact information in the model, and how should we summarize this? Is there information from the call center, from real-estate, or about competitive marketing density that can be included? Again, experience, and

intellectual curiosity will lead to new and innovative approaches in the area of data summarization. A decade ago, the kitchen sink was much smaller, and could be included. In today's data environment, the level of detail at which information is stored opens up wide opportunities for exploration and experimentation.

One thing that entry-level modelers and analysts straight of school presume is that data is normally distributed (bell-curve), and that models can achieve a high level of significance in prediction. They soon realize that if they can reduce the non-response rate to a mailing from 98% to 95%, they've found gold. They're also still wrong 95% of the time. There are very few normally-distributed variables in a typical customer database. However, this is a core assumption in much of statistical theory. Analysts must be taught to forget many of the rules that are standard in the classroom.

Once upon a time, there was the belief among some that any marketer could build a model with a desk-top database interface. A non-technical user could frame a problem, run a query, push a few buttons, and review the results of their model. The idea was that a model is a series of rules, and a science so routine that it can be simplified for a non-technical audience. Companies have invested millions of dollars on developing this type of software. Yet, I've not had one client in our practice say that they've even *tried* this type of product.

By and large, these sales pitches have been replaced by the new pitch: We'll make the modeler's time more productive by



automating the routine tasks. Here, they may have a point. These software tools can be helpful, as long as they have the functionality required for a modeler to explore a wide range of statistical options and methodologies, and are integrated into the database. The challenge is still in the area of data summarization and variable creation, but these tools can have a role to play.

So, if you've made it to this point, you're wondering what it actually is that I'm trying to say. Our industry is simple: target an audience that wants your product and is willing and able to pay what you ask, so you make a profit. It is complex: goals, methodologies, data complexities, aggressive project timelines and production schedules, and accurate business forecasting of marketplace impact. Dedicated professionals abound in all of the disciplines required for marketing success, full of experience, expertise, and passion for excellence. For this day, this one day, a nod is given to the artist, the scientist, the modern-day modeler.