

## **How Not to Do Predictive Modeling**

*By Jim Wheaton  
Principal, Wheaton Group*

*Original version of an article that appeared in the August 1994 issue of  
“The Cowles Report on Database Marketing”*

Articles and speeches on predictive modeling invariably focus on successful case studies. However, sometimes even more can be learned from mistakes. In my travels as Vice President of Research and Consulting Services for Neodata, I certainly have run into my fair share of mistakes. For example:

### **No Net/Net List Deals**

A prospecting model was built for a fundraiser using individual/household overlay demographics. Back tests of the model on live mailings showed impressive segmentation power. Unfortunately, the fundraiser did not have access to net/net list rental arrangements, which meant that the names eliminated by the model would have to be paid for (as would not be the case with a ZIP model).

Consider, for example, a hypothetical list with a published cost of \$100/M, for which the model eliminated the bottom 8 deciles. The actual, in-the-mail cost would be \$500/M, which clearly is not cost effective.

Fortunately, our research group performed financial analysis on the model before it was used in live mailings, and proved that under no realistic circumstances would the model ever be cost effective without net/net rental arrangements. (In fact, the analysis suggested that there exists no realistic circumstance in which any individual/household model will ever work for any direct marketer without the existence of net/nets.)

### **Modeled Out of Business**

A predictive model was built to rank-order existing customers in terms of their probability of repurchasing in the future. Unfortunately, this model drove every single-buyer into deciles that were below the mail/no mail cutoff recommended by the research company. Fortunately, the client quickly realized that this strategy – although effective for maximizing short-term profits – would in the long-term drive the business into bankruptcy. (After all, the only way for single-buyers to become multi-buyers is to mail them!)

**Decile vs. Decile**

A predictive model was built by a “10 to 1 Decile shop”; that is, by a research company that labeled its best buyers “Decile 10” and its worst buyers “Decile 1.” Unfortunately, the rest of the direct marketing world rank orders from Decile 1 to Decile 10. The model was forwarded to a service bureau that had no previous relationship to the research company, with written instructions to “pull off the top four Deciles.” The service bureau, mindful as it was of industry standards, proceeded to select Deciles 1 to 4, which resulted in the worst 40% of the file being mailed!

**Where’s the News?**

A research company built a predictive model for a cataloger in which everyone was eligible to be scored: multi-buyers, single-buyers, inactives, inquiries, and cross-sell candidates from other catalog titles within the overall corporate umbrella. Unfortunately, regression models – as is true with all other predictive statistical techniques – take the path of least resistance when attempting to segment by the probability of future response. Therefore, the result was a “sediment model,” in which multi-buyers were the primary residents of the top couple of deciles, single-buyers the residents of the next two, followed – sequentially – by inquiries, inactives, and cross-sell candidates.

Because the direct marketer already knew that multi-buyers generally perform better than single-buyers, who in turn generally perform better than inactives – and so on – the model essentially was worthless.

**ZIP-Less Lift**

A research company built a ZIP Code model to segment outside list rental prospects for a very targeted cataloger. Unfortunately, ZIP Code prospecting models generally do not display “lift,” top 10% to average, of more than 140 (i.e., with an overall response rate of 1%, Decile 1 will not pull more than 1.40%). Because of the very circumscribed audience for this cataloger's product, only a handful of affordable rental lists are available, and all of those had response rates that were several times higher than average. As a result:

- For the handful of affordable rental lists, even Decile 10 (the worst) names performed above the mail/no mail cutoff.
- For all other rental lists, even Decile 1 (the best) names performed below the mail/no mail cutoff.

Therefore, the ZIP model, although statistically successful at differentiating responders from non-responders, was worthless from a business point of view.

## **Skeletons From My Closet**

### First Skeleton

Anyone who has built a large number of predictive models will make some mistakes. This is inevitable, given the complex processes that are involved in a successful model build and implementation. Because I am no exception, I'll conclude by confessing to some "past skeletons in my own closet":

As part of a large database deal with a cataloger, we agreed to put predictive models into production immediately after the completion of the database. Therefore, "time 0" analysis files were available only for mailings that were done off the previous service bureau's database structure. This meant that the models had to be constructed off the previous structure and then converted to the existing database structure.

Unfortunately, the previous database had many significant data anomalies. This resulted in a large percentage of the records in the new database not having a one-to-one correspondence in the values for many of the key fields (i.e., the previous database showed a net dollar amount of \$300 for a given order, but warehoused raw transaction information showed a gross dollar amount of \$300 and a returns dollar amount of \$250).

As a result, the models were compromised and did not provide the expected segmentation power.

During this same project, it was noticed that there was an unusually large percentage of customers who had only one order. The client's answer was that the catalog had been growing rapidly over the past year. The real reason was discovered only later: about \$80 million dollars of transactions, representing about 2.5 years of history, was missing from the database.

Needless to say, the models were compromised even further when what appeared to be the single-buyer inhabitants of the bottom deciles – who were in fact multi-buyers – ordered merchandise with a vengeance!

### Second Skeleton

A model was built for a retail client to predict future purchase behavior, which did not perform well in the mail. A post mortem turned up nothing unusual until a serendipitous conversation with a programmer who had participated in bringing up this client's database several years earlier:

Unfortunately, although transaction information had been warehoused for several years, large scale gaps had existed in the data (e.g., many transactions had no dollar amount). The client contact, who had subsequently left the company, had a solution for this: simply plug artificial information, which would correspond to the average for all records which contained the data in question.

My response was that this was impossible because the Exploratory Data Analysis that was performed as part of the modeling project would have picked this up. The programmer's reply was that the client was particularly clever, and had written a program to generate a random “plus or minus” factor around the average that would correspond to the distribution of values for all records which contained the data in question!

*Jim Wheaton is a Principal at Wheaton Group, and can be reached at 919-969-8859 or [jim.wheaton@wheatongroup.com](mailto:jim.wheaton@wheatongroup.com). The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts. Jim is also a Co-Founder of Data University [www.datauniversity.org](http://www.datauniversity.org).*