

# Highlight Case Study 2:

## *Customer Response Model development*

### **Executive Summary**

The client is a major retailer specializing in mail order and Internet sales but also offering its products through multiple company-owned national and international outlets.

### **Business objective:**

Predict consumer response to catalog mailings for a variety of the client's products. In all, this project involved the development and testing of 130 models.

### **Methodology:**

Build a model, using the customer's internal data merged with sourced data, to predict consumer purchase interests and response behavior.

### **Data receiving mechanism**

Client sends the dataset in a transport file form and places it on the server. Once the data is saved into the project folder, we use SAS to convert the flat file into a SAS dataset. The file consists of all variables of continuous and categorical records.

### **Data Audit and cleansing**

Overall data health and hygiene assessment was done to provide baseline metrics on the following:

- We first cross checked with the number of records in the flat file
- We then cross-checked with the specifications list, send by the client, for whether all the variables mentioned as available in the layout file are indeed present in the data or not.
- We then assessed the number of records for deceased individuals
- Next, we compute the frequency for each variable to check the percentage of missing values and typographical errors.

### **Data and the Statistics (for each model)**

95,862 records were received from Client (400 MB)

31,188 is the sample size for the Response model

64,672 is the sample size for Expected demand model

## **Model Building:**

### ***Modeling:***

1. Run the regression model and select the significant variables
  - a. For all the selected variables:
    - i. Create a bi-variate graph for each selected variable to identify the linearity and the relationship with the dependent variable.
    - ii. For all the binary variables, ensure that the coverage is “adequate”.
    - iii. Check for multicollinearity and remove if it exists.
  - b. Run the Regression model with the selected variables
2. Check decile report for all the variables.
3. Retain variables which will satisfy the decile report and max R-Square in the regression model.
4. For the variables which were included in the model, re-code to apply a “cap”, if necessary (we usually cap variables at the 99<sup>th</sup> percentile).
5. Re-run the model with the capped variables to get the final quality output
6. Develop the lift chart for the model sample

### **Model Validation:**

Using the predicted probabilities we develop the lift chart for the validation sample. Additionally, we assessed the relative values and direction of the parameters in each of the 130 models for their consistency with the client’s prior intuition.

### **Software used:**

SAS 9.1, SAS/Stat, SAS Analytics PRO on Windows-XP server