ABSTRACT

It is more important than ever to accurately select receptive consumers for direct solicitation. Increasingly saturated markets, the growing costs of direct mail and telemarketing, decreasing consumer responsiveness and increasing dissatisfaction with direct marketing methods combine to require a more finely tuned approach.

With a database of customer information and the LOGISTIC procedure, a predictive model can be produced and put into operation very quickly. This paper will address aspects building a marketing response model using the SAS® system in a financial services setting. Topics to be discussed include experiment design, data screening, preliminary data analysis and characteristic selection, model selection, as well as validation and tracking issues.

INTRODUCTION

Today it is vital for marketers to make the most of every customer contact opportunity. In addition to the direct costs associated with each contact are the less tangible harm to the company name for “junk mail” or “harassing phone calls”. The long-term goal of any marketing modeler should be a world in which no consumer receives “junk” mail. That is, every contact the have with a company is just the sort of thing they would be interested in. There would seem to be only two ways of accomplishing this: get to know each customer as individuals or develop your predictive modeling skills to the point where you can, with near certainty, target only customers interested in an unsolicited promotion.

This paper will give an overview of how the SAS system can be used to help build response models in the context of a financial services organization. It will not discuss in detail the statistical algorithms behind the SAS procedures. PROC LOGISTIC, when coupled with careful planning and a knowledge of both your customer and product, can help you create marketing response models that will reduce your costs and increase consumer response rates.

METHOD

For demonstration purposes, this paper will use examples from the solicitation by telemarketing of retail credit card credit holders for an extended warranty product. This offer doubles the manufacturers warranty on any product purchased with the card. Any cardholder who agrees to sign up for the product and who pays for the first month’s premium on their card will be considered a positive response; everyone else is a negative.

Experiment Design

Experiment design is arguably the most important step in the development of a marketing response model. If the design is flawed, no amount of data manipulation or clever statistical technique will be able to compensate for it. Experiment design seeks to obtain results for a sample of accounts, such that a logistic regression model can be built that is applicable to the entire target market. In other words, to build a proper response model, you need to test the marketing offer on a random sample (or pseudo-random sample) of accounts, build the model based on that, and then apply the model to the entire population. To that end, two questions must be answered: Who should be solicited? and How many should be included in the sample?

When selecting a sample for model building purposes, it’s important to make sure that the sample is homogeneous and from the same population that will be targeted with the promotion. Cardholders who will never be selected for logical reasons (e.g. closed accounts, accounts not in good standing) should be excluded from the sample, as should those who don’t fit the marketing profile (e.g. in a region that’s not targeted, outside a specific demographic). If cardholders will not be solicited more than once, this should be reflected in the sample as well.

Because the responses will be binary (i.e. signed up or didn’t), the expected response rate is the only major factor to consider, outside of cost, in selecting a sample size. The total size of the sample is important, but it is just as important to have a sufficient number of responders to build a model. A good rule of thumb is that at least 150 responders are necessary to build a basic model. When selecting a sample size, consider other analyses in addition to modeling that might be conducted (e.g. financial, telemarketing script) and adjust the size accordingly. A typical response rate to telemarketing is about 3% so in our example of an extended warranty product we selected 10,000 cardholders for our sample.

Sometimes it may be more sensible to not build a model, and these instances should be identified during the experiment design phase. Examples include when the population is expected to change significantly, when the
majority of customers have already been solicited or when a mass mailing strategy is already in place.

Data Collection & Screening

While waiting for the results of the test mailing is an ideal time to collect the data on the independent variables. These values will either be purchased or be found in your data warehouse. Experience and knowledge of your customer base and product is invaluable here (e.g. don’t sell refrigerators to people North of 60°!). Don’t ignore experience with similar products and the input of others.

Table 1 gives a listing of the data we collected for our example. Normally many more pieces of data will be collected and analyzed. A good rule of thumb is to start with ~ 20 to 30 variables and narrow it down to 6 to 12.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
<th>Valid Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg_Bal</td>
<td>Average card balance over past 12 months</td>
<td>-$250 to $4,500</td>
</tr>
<tr>
<td>Card_tenure</td>
<td>Age of Account in months</td>
<td>0 to 505</td>
</tr>
<tr>
<td>CtcMemb</td>
<td>Member of roadside assistance program</td>
<td>1 = ‘Member’; 0 = ‘Non-Member’</td>
</tr>
<tr>
<td>Risk_Scr</td>
<td>Current Credit Risk Score</td>
<td>500 to 750</td>
</tr>
<tr>
<td>WPromo</td>
<td>Number of times promoted with similar products in the past</td>
<td>0 to 2</td>
</tr>
</tbody>
</table>

Once the data has been collected, be on the look out for data that doesn’t make sense (e.g. year of birth 1850) and try to find the source of the problem. At this stage it is important to consult with the parties that have the most detailed knowledge of the data in question, whether it be the data analysis group, marketing, finance or credit risk.

Check also for outliers to determine reasonableness. A good rule of thumb for normally distributed data is to look closely at any data points more than 1.5 times the interquartile range from either the 25th or 75th percentile and more than 3 standard deviations from the mean.

In our example, we ran the UNIVARIATE procedure on each variable. The code below runs a summary data analysis for the Risk Score. The output follows.

```
PROC UNIVARIATE DATA=VARIABLE_SELECTION;
VAR RISK_SCR;
TITLE 'Analysis of Risk Score';
RUN;
```

Analysis of Risk Score

The UNIVARIATE Procedure

Variable: RISK_SCR (RISK_SCR)

Moments

N                                  10000 Sum Weights        10000
Mean                                624.6158 Sum Observations  6246158
Std Deviation                      60.47539 Variance            3657.27292
Skewness                           -0.1341477 Kurtosis          1.14597041
Uncorrected SS                     3938018048 Corrected SS     36569071.9
Coeff Variation                    9.6820143 Std Error Mean     0.6047539

Basic Statistical Measures

Location                   Variability
Mean                        624.6158 Std Deviation  60.47539
Median                      625.0000 Variance         3657
Mode                        619.0000 Range            956.00000
Interquartile Range         80.00000

Tests for Location: Mu0=0

Student’s t t  1032.843  Pr > |t|    <.0001
Sign        M    4999.5  Pr >= |M|   <.0001
Signed Rank S  24997500  Pr >= |S|   <.0001

Quantiles (Definition 5)

Quantile    Estimate
100% Max            956
99%               761
95%               724
90%               703
75% Q3             665
50% Median         625
25% Q1             585
10%                548
5%                 526
1%                 482
0% Min             0

Extreme Observations

----Lowest----  ----Highest---

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5621</td>
<td>832</td>
<td>5206</td>
</tr>
<tr>
<td>384</td>
<td>4200</td>
<td>841</td>
<td>7410</td>
</tr>
<tr>
<td>384</td>
<td>559</td>
<td>845</td>
<td>7019</td>
</tr>
<tr>
<td>415</td>
<td>4180</td>
<td>855</td>
<td>8572</td>
</tr>
<tr>
<td>415</td>
<td>1813</td>
<td>956</td>
<td>1233</td>
</tr>
</tbody>
</table>

From the interquartile range (80) and the 25% and 75% quartiles (585 and 665 respectively), we calculate an acceptable minimum and maximum of 465 and 785. Calculating three standard deviations from the mean we compute a minimum and maximum of 443 and 806. There are data points that fall outside of our two
acceptable ranges, but after speaking with the credit risk department, we realize that all but the risk score of 0 and 956 are acceptable. Looking more closely at those two points we determine that the 0 was miscoded and should be 600 and the 956 should be 656.

Similar analyses of the other potential variables were conducted and we determined the data was acceptable.

**Preliminary Data Analysis & Characteristic Selection**

Once the response file is back from the telemarketing firm it’s a simple matter to merge the responses with the data file collected during the previous phase. It’s good practice to merge such data on both an account identifier and a tracking code:

```sas
DATA TOTAL_RESPONSE_FILE;
  MERGE VARIABLE_SELECTION (IN=A)
       RESPONSE_FILE (IN=B);
  BY ACCOUNT_ID TRACK_CODE;
  IF A AND B;
RUN:
```

It is standard practice to hold back a portion of the completed file to validate the model once it’s built. Typically, this will be 10 to 20% of the total base. For the purposes of our example, however, all the observations were included.

Selection of characteristics for the model is informed not just by numerical analysis but also by knowledge of the product and customer base. It is important to select characteristics that are predictive, independent, stable with respect to the others and to the responses, and that make logical sense.

Predictive variables can be identified using the FREQ procedure for discrete data and the TTEST procedure for continuous variables. These are not the most exhaustive tests, nor even the most statistically appropriate, but at this stage, we are interested only in a rough idea of the data. We will refine these variables further once we get to the modeling stage.

In our example, the variable CTACMEMB is clearly predictive, based on the results generated by the following SAS code, particularly the Chi Square Stat for the table.

```sas
PROC FREQ DATA=TOTAL_RESPONSE_FILE;
  TABLES RESPONSE*CTACMEMB/NOPERCENT
  NOROW CHISQ;
  TITLE 'Testing Predictiveness of CTAC';
RUN;
```

PROC TTEST applied to a continuous variable such as AVG_BAL yields an indication of its quality as a predictor. From the example below, we can see that there is strong evidence to suggest AVG_BAL is also a good indicator of responsiveness.

```sas
PROC TTEST;
  CLASS RESPONSE;
  VAR AVG_BAL;
  TITLE 'Testing Variance of Avg_Bal wrt response';
RUN;
```

**Testing Predictiveness of CTAC**

The FREQ Procedure

<table>
<thead>
<tr>
<th>Table of RESPONSE by CTACMEMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESPONSE (RESPONSE)</td>
</tr>
<tr>
<td>CTACMEMB (CTACMEMB)</td>
</tr>
<tr>
<td>Frequency Col Pct Total</td>
</tr>
<tr>
<td>0   8034  1516  9550</td>
</tr>
<tr>
<td>95.84  93.75</td>
</tr>
<tr>
<td>1   349  101  450</td>
</tr>
<tr>
<td>4.16  6.25</td>
</tr>
<tr>
<td>Total 8383  1617 10000</td>
</tr>
</tbody>
</table>

Statistics for Table of RESPONSE by CTACMEMB

- **Chi-Square**: 13.6852
- **Likelihood Ratio Chi-Square**: 12.5499
- **Continuity Adj. Chi-Square**: 13.2048
- **Mantel-Haenszel Chi-Square**: 13.6838
- **Phi Coefficient**: 0.0370
- **Contingency Coefficient**: 0.0370
- **Cramer's V**: 0.0370

**Fisher's Exact Test**

- **Cell (1,1) Frequency (F)**: 8034
- **Left-sided Pr <= F**: 0.9998
- **Right-sided Pr >= F**: 2.375E-04
- **Table Probability (P)**: 8.742E-05
- **Two-sided Pr <= P**: 3.855E-04

Sample Size = 10000

| Variable Class N Mean Std Dev Std Dev Std Dev Std Dev Std Dev Std Err T-Tests |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Variable Class  | Lower CL        | Upper CL        | Lower CL        | Upper CL        | Lower CL        | Upper CL        | Lower CL        | Upper CL        |
| AVG_BAL 0       | 311.93          | 316.36          | 320.91          | 3.2372          | 313.0          | 318.25          | 322.72          | 15.352          |
| AVG_BAL 1       | 334.31          | 356.16          | 381.09          | 16.79           | 313.9          | 318.25          | 322.72          | 15.352          |
After conducting similar analysis for the six variables examined and through discussions with Finance, Marketing and Credit Risk, we arrived at the following groupings of the variables (Table 2). Each of the new variables in a group are mutually exclusive so each observation will have one and only one variable in each group set to 1 with all the others set to 0.

**Table 2.**

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Source Variable</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABAL01</td>
<td>Avg_bal</td>
<td>≤ $250</td>
</tr>
<tr>
<td>ABAL02</td>
<td>Avg_bal</td>
<td>($250,$750]</td>
</tr>
<tr>
<td>ABAL03</td>
<td>Avg_bal</td>
<td>&gt; $750</td>
</tr>
<tr>
<td>CTEN01</td>
<td>Card_ten</td>
<td>≤ 12</td>
</tr>
<tr>
<td>CTEN02</td>
<td>Card_ten</td>
<td>(12,24]</td>
</tr>
<tr>
<td>CTEN03</td>
<td>Card_ten</td>
<td>&gt; 24</td>
</tr>
<tr>
<td>CTMEM01</td>
<td>Ctac_memb</td>
<td>0</td>
</tr>
<tr>
<td>CTMEM02</td>
<td>Ctac_memb</td>
<td>1</td>
</tr>
<tr>
<td>GEO01</td>
<td>Geo_Code</td>
<td>‘R’, ‘S’, ‘N’</td>
</tr>
<tr>
<td>GEO02</td>
<td>Geo_Code</td>
<td>‘T’, ‘U’</td>
</tr>
<tr>
<td>RS01</td>
<td>Risk_scr</td>
<td>≤ 600</td>
</tr>
<tr>
<td>RS02</td>
<td>Risk_scr</td>
<td>(600,650]</td>
</tr>
<tr>
<td>RS03</td>
<td>Risk_scr</td>
<td>&gt; 650</td>
</tr>
<tr>
<td>WP01</td>
<td>WPPromo</td>
<td>0</td>
</tr>
<tr>
<td>WP02</td>
<td>WPPromo</td>
<td>&gt; 0</td>
</tr>
</tbody>
</table>

The final step in characteristic selection involves screening to make sure collinearity will not overly influence the model. In the final model output this may appear as

1. Variables that have a negative coefficient when the preliminary analysis suggestions it should be positive
2. Coefficients that change dramatically with the addition of an x variable
3. Variables that come in as non-significant despite preliminary analysis that suggests otherwise
4. A number of variables that do not pass significance tests but yield a model with excellent fit statistics

The quickest way to screen for this is by running the CORR procedure on the variables and looking more closely at values higher than 0.5.

**Model Selection**

Once the variables to input to the model have been selected, the process moves quite quickly. To begin with, we use forward, backward, stepwise and no selection on all the variables to get a general sense of how the variables are working together. PROC LOGISTIC is the sensible choice for building this sort of model because binary data like these violate several of the assumptions of linear regression, most significantly the assumption of constant variance.

**Summary of Stepwise Selection**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Number</th>
<th>Score</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP01</td>
<td>1</td>
<td>1</td>
<td>178.6441</td>
</tr>
<tr>
<td>RS01</td>
<td>1</td>
<td>2</td>
<td>39.7511</td>
</tr>
<tr>
<td>CTEN01</td>
<td>1</td>
<td>3</td>
<td>35.2942</td>
</tr>
<tr>
<td>ABAL02</td>
<td>1</td>
<td>4</td>
<td>21.7852</td>
</tr>
<tr>
<td>CTMEM02</td>
<td>1</td>
<td>5</td>
<td>14.8424</td>
</tr>
<tr>
<td>ABAL01</td>
<td>1</td>
<td>6</td>
<td>14.2742</td>
</tr>
<tr>
<td>CTEN02</td>
<td>1</td>
<td>7</td>
<td>12.0620</td>
</tr>
<tr>
<td>RS02</td>
<td>1</td>
<td>8</td>
<td>10.0220</td>
</tr>
<tr>
<td>GEO01</td>
<td>1</td>
<td>9</td>
<td>10.0667</td>
</tr>
</tbody>
</table>
The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter DF Estimate Error Chi-Square Pr > ChiSq
Intercept 1 -3.5819 0.1482 583.7802 <.0001
ABAL01 1 -0.6569 0.1758 13.9570 0.0002
ABAL02 1 -0.6918 0.1105 39.1748 <.0001
RS01 1 0.8445 0.1371 44.3401 <.0001
RS02 1 0.4413 0.1371 10.3668 0.0013
CTMEM02 1 0.4544 0.1193 14.5180 0.0001
WP01 1 0.7481 0.1268 40.9149 <.0001
CTEN01 1 0.7481 0.1169 40.9149 <.0001
CTEN02 1 0.4464 0.1309 11.6271 0.0006
GEO01 1 -0.3144 0.0994 10.0012 0.0016

Another thing to consider when looking at this output is the order in which the variables entered the model and the effect they had on it. Dramatic changes in the score Chi-Square with the addition of a single variable could be a sign of instability in the model.

ABAL02 1 -0.6918 0.1105 39.1748 <.0001
RS01 1 0.8445 0.1371 44.3401 <.0001
RS02 1 0.4413 0.1371 10.3668 0.0013
CTMEM02 1 0.4544 0.1193 14.5180 0.0001
WP01 1 0.7481 0.1268 40.9149 <.0001
CTEN01 1 0.7481 0.1169 40.9149 <.0001
CTEN02 1 0.4464 0.1309 11.6271 0.0006
GEO01 1 -0.3144 0.0994 10.0012 0.0016

In the case of our example, we are looking at an optimal solution to the problem.

Odds Ratio Estimates

Point Estimate 95% Wald Confidence Limits
ABAL01 0.518 0.367 0.732
ABAL02 0.501 0.403 0.622
RS01 2.327 1.815 2.983
RS02 1.555 1.188 2.034
CTMEM02 1.575 1.247 1.990
WP01 3.567 2.931 4.342
CTEN01 2.113 1.680 2.657
CTEN02 1.563 1.209 2.020
GEO01 0.730 0.601 0.887

Often the automated runs will yield identical results. In our example above, for instance, variables were only added to the model. None were removed despite it being a stepwise procedure, so the results of the forward selection are identical.

When comparing the models, the AIC (Akaike’s Information Criteria) and SC (Schwartz’s Criteria) consider fit and characteristics: the lower the number the better.

The section “Testing the Global Null Hypothesis” considers whether the model has any predictive value at all. That is, it tests the hypothesis that all the coefficients are zero. The Chi-Square values are all significantly high so that model has at least some predictive value.

The listings for individual characteristics give the coefficient and the related Chi-Square stat. The Pr > ChiSq column tests the null hypothesis and all of these variables pass well. The coefficients should all agree with the preliminary analysis.

Finally, the “Association of Predicted Probabilities and Observed Responses” gives more summary statistics. The Percent Discordant, Somer’s D and Gamma tests all suggest that the model is strong.

In determining which variables to add, delete or merge, consider the ChiSq tests. Very close coefficients may need to be merged. The degree of differentiation between the two can be tested with the CONTRAST statement.

Another thing to consider when looking at this output is the order in which the variables entered the model and the effect they had on it. Dramatic changes in the score Chi-Sq with the addition of a single variable could be a sign of instability in the model.

In the case of our example, we are looking at an optimal solution to the problem.

Validation

The simplest way to validate the model results is to create a lift table by dividing the population into deciles based on their predicted value (from logout), then compare these to the actual observed. If a validation sample has been held back, it can also be scored and a similar table created for comparison purposes. The following SAS code create our deciles and lift chart data:

```
PROC SORT;
  BY DESCENDING PRED;
RUN;
```

```
DATA LOGOUT;
  PERCENT TIED 2.5 TAU-A 0.039 PAIRS 4297500 C 0.729;
```

Often the automated runs will yield identical results. In our example above, for instance, variables were only added to the model. None were removed despite it being a stepwise procedure, so the results of the forward selection are identical.

When comparing the models, the AIC (Akaike’s Information Criteria) and SC (Schwartz’s Criteria) consider fit and characteristics: the lower the number the better.

The section “Testing the Global Null Hypothesis” considers whether the model has any predictive value at all. That is, it tests the hypothesis that all the coefficients are zero. The Chi-Square values are all significantly high so that model has at least some predictive value.

The listings for individual characteristics give the coefficient and the related Chi-Square stat. The Pr > ChiSq column tests the null hypothesis and all of these variables pass well. The coefficients should all agree with the preliminary analysis.

Finally, the “Association of Predicted Probabilities and Observed Responses” gives more summary statistics. The Percent Discordant, Somer’s D and Gamma tests all suggest that the model is strong.

In determining which variables to add, delete or merge, consider the ChiSq tests. Very close coefficients may need to be merged. The degree of differentiation between the two can be tested with the CONTRAST statement.

Another thing to consider when looking at this output is the order in which the variables entered the model and the effect they had on it. Dramatic changes in the score Chi-Sq with the addition of a single variable could be a sign of instability in the model.

In the case of our example, we are looking at an optimal solution to the problem.

```
PROC SORT;
  BY DESCENDING PRED;
RUN;
```

```
DATA LOGOUT;
  SET LOGOUT;
  DEC=INT(_N_/((10000-10+1)/10+1)+1);
RUN;
```

Often the automated runs will yield identical results. In our example above, for instance, variables were only added to the model. None were removed despite it being a stepwise procedure, so the results of the forward selection are identical.

When comparing the models, the AIC (Akaike’s Information Criteria) and SC (Schwartz’s Criteria) consider fit and characteristics: the lower the number the better.

The section “Testing the Global Null Hypothesis” considers whether the model has any predictive value at all. That is, it tests the hypothesis that all the coefficients are zero. The Chi-Square values are all significantly high so that model has at least some predictive value.

The listings for individual characteristics give the coefficient and the related Chi-Square stat. The Pr > ChiSq column tests the null hypothesis and all of these variables pass well. The coefficients should all agree with the preliminary analysis.

Finally, the “Association of Predicted Probabilities and Observed Responses” gives more summary statistics. The Percent Discordant, Somer’s D and Gamma tests all suggest that the model is strong.
Implementation and Tracking

The greatest challenges in implementing a model are ensuring the SAS programs are coded properly and that the population against which the model is used is analogous to the population upon which it was built. The only guaranteed method of ensuring that the code is accurate is to score a sample of observations by hand. It is also sensible to use the data retrieval programs in development and production, if possible.

Tracking is most easily accomplished by using the lift table which, when combined with a graph, clearly shows the lift a model provides over a random selection. That is, it clearly demonstrates the gain in efficiency of selecting marketing leads using a model over a random selection.

CONCLUSION

This paper has demonstrated that it is possible to build and implement a response model using basic SAS procedures in a very short period of time. Further refinements can be made to the resulting model, such as response net of cancellations or with profit factored in.

AUTHOR CONTACT

David Marsh
Canadian Tire Financial Services
555 Prince Charles Dr.
Welland, Ontario
L3C 6B5
CANADA

Voice: (905) 735-3131 ext. 3860
Fax: (905) 714-2701
Internet: david.marsh@ctal.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 registered trademarks or trademarks of their respective companies.