

Paper 096-2013

Incremental Response Modeling Using SAS[®] Enterprise Miner[™]

Taiyeong Lee, Ruiwen Zhang, Xiangxiang Meng, and Laura Ryan

SAS Institute Inc.

ABSTRACT

Direct marketing campaigns that use conventional predictive models target all customers who are likely to buy. This approach can lead to wasting money on customers who will buy regardless of the marketing contact. However, incremental response models that use a pair of training data sets (treatment and control) measure the incremental effectiveness of direct marketing. These models look for customers who are likely to buy or respond positively to marketing campaigns when they are targeted but are not likely to buy if they are not targeted. The revenue generated from those customers is called incremental revenue. This paper shows how to find that profitable customer group and how to maximize return on investment by using SAS[®] Enterprise Miner[™].

INTRODUCTION

Traditional response models have long been used to predict who is likely to respond to an action, such as a marketing incentive. In such models, all customers in a group receive the promotion, their responses are recorded, and a predictive model is built to separate likely responders from those unlikely to respond. This is done through a number of predictive modeling methods such as decision trees, neural networks, or regression models.

In essence, customers who receive the offer can be viewed as receiving a treatment. Because the group consists only of treated customers, the traditional response model cannot take into account that some, perhaps many, of those treated customers would have responded regardless of the incentive. The traditional response model yields the likely responders, but they might have responded even in the absence of an incentive. Therefore, relying on this type of modeling can lead to wasted effort and money and cannot capture the true effectiveness of a marketing campaign.

The incremental response model divides customers into four groups: (1) those who respond only when targeted with a marketing action, (2) those who respond regardless of contact, (3) those who do not respond, regardless of contact, and (4) those who are less likely to respond because of contact. The model enables an organization to focus on the customers who respond because of the marketing action, thus decreasing unnecessary marketing costs and increasing ROI. This model also enables an organization to avoid the negative effects of marketing to customers who do not respond. These customers might be annoyed by the marketing action, or they might prefer to be left alone.

An incremental response model uses two randomly selected data sets, which are called control and treatment as in a clinical trial setting. The treatment group receives the promotion offer, but the control group does not. Figure 1 shows where the incremental response (customer group 1) can be identified in the process of modeling.

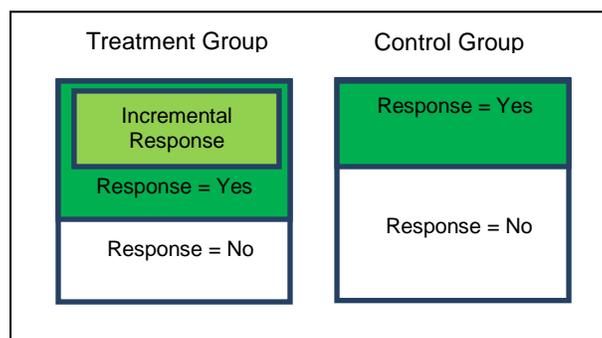


Figure 1. Potential Incremental Response

Although not much research exists in this area, some can be found in Radcliffe and Surry (1999), Lo (2002), and Hansotia and Rukstales (2002). Radcliffe and Surry illustrate a well-defined concept of incremental response with distinct customer groups. Hansotia uses a decision tree algorithm that has incremental response rate as a split criterion. Lo uses a single regression equation for both treatment and control and a holdout sample from the treatment group. Larsen (2010) also introduces a bifurcated logistic regression and generalized naïve Bayes classifier for the incremental response model.

This paper is organized as follows: First, basic descriptive incremental statistics from treatment and control data sets are explained, and then two different modeling techniques are explained for the incremental response model. The next section extends the response model to an incremental sales model by using Heckman's two-step modeling, and then it shows heuristic model diagnostics and how to choose profitable customers to generate the incremental revenue. Finally, a case study with a real data set is presented.

BASIC INCREMENTAL STATISTICS

First, simulated data that show apparent incremental responses are created in a similar way to Lo (2002). The data contain two targets (Response and Sales), four explanatory variables (X1–X4), and one treatment variable (Promotion) that indicates treatment or control group. It also has a cost variable and a customer ID variable. Each treatment or control group has 10,000 observations. These simulated example data are used for the purpose of illustration throughout the paper except for the case study example section. The data partition is not used in this example. Figure 2 shows the diagram and the variable setting for the Incremental Response node.

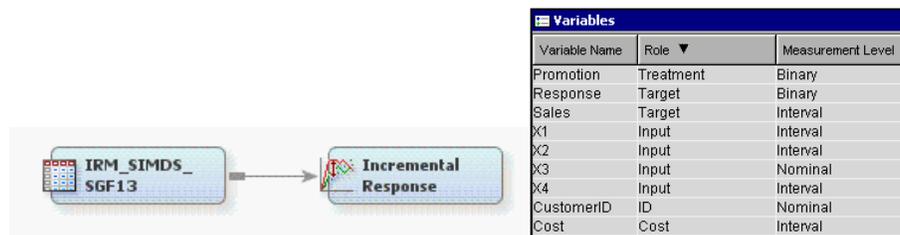


Figure 2. SAS Enterprise Miner Flow and Variable List

To view the basic difference statistics, select **View > Model > Response Outcome Summary Table** after the node is run. Display 1 shows the resulting window.

| Data Role | Variable | Label | Treatment:1 | Control:0 | Difference |
|-----------|-------------|------------------------|-------------|-----------|------------|
| Train | NumCust | Number of observations | 10000 | 10000 | 0 |
| Train | NumResp | Number of Response | 6175 | 3327 | 2848 |
| Train | rate | Rate of Response | 0.6175 | 0.3327 | 0.2848 |
| Train | Ave_outcome | Average Sales | 155.3788 | 129.2173 | 26.16147 |
| Train | Tot_outcome | Total Sales | 959464 | 429906 | 529558 |

Display 1. Summary of Response and Outcome

The treatment group has a 62% response rate, and the control group has a 33% response rate. The promotion has resulted in about 28% incremental responses and average incremental sales of approximately \$26.00. The 28% of customers are called the true responders of the promotion, but you don't know who they are. Finding those customers is the main purpose of incremental response modeling. SAS Enterprise Miner also shows bar charts of the basic statistics across the training and validation data sets. The treatment and control levels can be switched by using treatment level selection. Descending order is the default, so the treatment level is set to 1 in this example.

VARIABLE PRESCREENING

Variable selection is an important step in predictive modeling because it reduces the dimension of the model, avoids overfitting, and improves model stability and accuracy. In the context of incremental response modeling, the target is the incremental effect that is calculated as the difference between treatment and control outcomes. This kind of twofold calculation increases the risk of overfitting and eventually jeopardizes the predictive performance of the model, especially when the incremental effect is relatively small, as it is in many practical cases. This encourages some robust variable selection prior to building the model.

SAS Enterprise Miner implements the net information value (NIV) that is used by Larsen (2010) and uses its variation for variable selection in the incremental response modeling. This method is intuitive, easy to implement, and highly flexible and effective as a data exploratory tool.

Weight of Evidence and Information Value

Denote $Y \in \{0,1\}$, and assume $Y = 1$ is the success or response. In the discrete form, group the predictor variable $X \in \mathcal{X}$ into I mutually exclusive bins. Then the weight of evidence (WOE) for each bin is calculated as

$$WOE_i = \log \frac{P(X = x_i|Y = 1)}{P(X = x_i|Y = 0)} \text{ for } i = 1, 2, \dots, I$$

and the information value (IV) is calculated as

$$IV = \sum_i (P(X = x_i|Y = 1) - P(X = x_i|Y = 0)) \cdot WOE_i$$

The WOE is calculated after the predictors are put into meaningful bins. However, in practice, you usually select 20 equally sized bins based on the quantiles and the data distribution. Then, by analyzing WOE among those consecutive bins, you can explore the predictor variables over different ranges to understand where the predictive strength of this particular variable comes from.

The IV measures the strength of correlation between the explanatory variable and the response for the binary outcome. The empirical rule of thumb (Siddiqi 2005) is used for assessing the IV: a value less than 0.02 indicates that the variable is not predictive, whereas a value greater than 0.3 shows that the variable has strong predictive power.

Net Weight of Evidence and Net Information Value

Given the Treatment (T) and Control (C), the WOE and IV concepts can be applied to the incremental response model with a modification. Larsen (2010) suggests net weight of evidence (NWOE) as

$$NWOE = \log \frac{P(X = x_i|Y = 1)_T / P(X = x_i|Y = 0)_T}{P(X = x_i|Y = 1)_C / P(X = x_i|Y = 0)_C}$$

where $P(X|Y)_T$ and $P(X|Y)_C$ denote the conditional probabilities for treatment group and control group, respectively. NWOE is the log-odds ratio that compares the response odds for the treatment group and the response odds for the control group.

Then, the net information value (NIV) is calculated as

$$NIV = \sum_i (P(X = x_i|Y = 1)_T P(X = x_i|Y = 0)_C - P(X = x_i|Y = 0)_T P(X = x_i|Y = 1)_C) \cdot NWOE_i$$

Using the NIV measure, the candidate variables are ranked and the best ones are identified, usually choosing either a fixed number (or percentage) or all those above a certain threshold. There is no empirical rule for the threshold.

Penalized Net Information Value

When the predictive power of the variable drops more markedly in validation data than in training data, concern is raised about the predictive robustness of the variable. A variable that lacks predictive robustness results in mismatched patterns of NWOE being calculated from the validation data set and the training data set. In such a case, the NIV is adjusted with a penalty term, which describes the difference between the NWOE that is calculated from training data and the NWOE that is calculated from validation data.

For each value or bin of predictor variable X , calculate the NWOE from the training data as $NWOE_{train}$, and calculate the NWOE from the validation data as $NWOE_{valid}$. Take the difference as

$$\omega = |NWOE_{train} - NWOE_{valid}|$$

Then, the penalty is calculated as

$$Penalty = \sum_i |P(X = x_i|Y = 1)_T P(X = x_i|Y = 0)_C - P(X = x_i|Y = 0)_T P(X = x_i|Y = 1)_C| \cdot \omega_i$$

The penalized NIV is

$$PNIV = NIV - penalty$$

The penalty function is especially crucial in evaluating NIV (Larsen 2010). Because the variance of NWOE aggregates the variance of WOE from the treatment group and from the control group, NWOE tends to vary more compared to the regular WOE values. Therefore, it is more common to see different NWOE patterns from training and validation data. The penalized NIV provides more robustness for variable selection in incremental response modeling. SAS Enterprise Miner uses the penalized NIV automatically when a validation data set is used.

INCREMENTAL RESPONSE SCORE MODELING

In incremental response modeling, one of the basic models is the difference score model, which measures the differences between predicted values from the model of the treatment group and from the model of the control group. These difference values are called difference scores. The ranked difference scores are binned in descending order for model assessment, and the final top subset is selected as the true responders to the marketing promotion.

The predictive model can be built in two different ways: one consists of two separate models from the treatment group and control group, and the other uses a single, combined equation.

Suppose there are two data sets, D_T and D_C , which are the treatment group and the control group, respectively. Denote a dependent variable y and explanatory variables x , and denote the number of observations in each group, n_T and n_C . Then,

$$D_T = \{x_i, y_i\}_{i=1}^{n_T}$$

$$D_C = \{x_i, y_i\}_{i=1}^{n_C}$$

$$D = D_T \cup D_C = \{x_i, y_i\}_{i=1}^n, \quad n = n_T + n_C$$

Without loss of generality, a linear model is considered as follows:

$$Y = X\beta + \epsilon$$

Difference Score from Two Separate Models

Two models are built separately on D_T and D_C :

$$\hat{Y}_T = X_T \hat{\beta}_T$$

$$\hat{Y}_C = X_C \hat{\beta}_C$$

Then, both models are used to calculate predicted values from the entire data sets ($D = D_T \cup D_C$). The difference scores can be obtained from the predicted values as

$$\widehat{DS}_i = (\hat{Y}_T - \hat{Y}_C)_i \text{ for } i = 1, 2, \dots, n$$

Customers who have a positive value of \widehat{DS}_i are initially considered as the incremental responders as a result of the promotion campaign. However, to determine the final set of incremental responders, further analyses must be performed with the ranked difference scores in decreasing order:

$$\widehat{DS}_{(i)} = (\hat{Y}_T - \hat{Y}_C)_{(i)} \text{ for } i = 1, 2, \dots, n$$

The section “Model Diagnostics” shows how to use these ranked difference scores.

SAS Enterprise Miner uses logistic regression with variable selection separately for each model so that each model has more flexibility in regard to its own data set (treatment data and control data).

Difference Score from a Combined Model

This model was used in Lo (2002). He used a new indicator variable, T_i , $T_i = 1$ for D_T and $T_i = 0$ for D_C , and he fit the model to the entire, combined data ($D = D_T \cup D_C$):

$$Y = X\beta + T\gamma + (XT)\phi + \epsilon$$

Then, the difference scores are obtained from the estimated equation, $\hat{Y} = X\hat{\beta} + T\hat{\gamma} + (XT)\hat{\phi}$, as follows:

$$\hat{Y}_T = X\hat{\beta} + \hat{\gamma} + X\hat{\phi}$$

$$\hat{Y}_C = X\hat{\beta}$$

$$\widehat{DS}_i = \hat{Y}_T - \hat{Y}_C = \hat{\gamma} + X\hat{\phi} \text{ for } i = 1, 2, \dots, n$$

This is a simplified expression of the difference scores in linear model, but in general they are calculated from the form of conditional expectation of Y :

$$\widehat{DS}_i = \hat{Y}_T - \hat{Y}_C = E(Y|X, T = 1) - E(Y|X, T = 0) \text{ for } i = 1, 2, \dots, n$$

One advantage of this model is that it shows some influential variables directly to the incremental response through the significant parameter estimates (for example, $\hat{\phi}$ in the linear model).

In SAS Enterprise Miner, the model can be chosen by selecting Yes for the Combined Model property, which indicates whether the treatment variable is included as a predictor in the model.

In addition to the variable prescreening technique, any well-known model selection method can be applied to each modeling process. Because SAS Enterprise Miner uses regression-based modeling techniques, forward, backward, and stepwise selection methods are provided.

INCREMENTAL SALES MODEL

An incremental response model often handles two targets: a binary response and an interval variable such as sales, revenue, and so on. The incremental response model that handles both targets is also known as an incremental sales model. The goal of the incremental sales model is to find customers who are likely to spend incrementally when they receive a promotion. An issue of unobserved target values occurs here. The interval target is observed only when the response is received, which introduces selection bias to the model. To overcome the problem, the Heckman selection model (Heckman 1979) is used when both interval and response targets exist in incremental modeling. Therefore, the interval target model (outcome model) is fitted by including the inverse Mills ratio in the model.

Suppose there are two target variables: Y is the sales amount per customer, and Z is a response variable for the promotion. Y is observed only when the customer responds to the promotion. So the predicted model can be formulated as

$$Y = X\beta + \epsilon \text{ if } Z = 1$$

$$Y = \text{not observed if } Z = 0$$

$$Z^* = X^*\gamma + \omega$$

$$Z = \begin{cases} 1 & \text{if } Z^* > 0 \\ 0 & \text{if } Z^* \leq 0 \end{cases}$$

where Z^* is a latent variable. X^* and X are not independently distributed, and they usually have the same or many common variables in incremental models. So the error ϵ is not independent of the error ω . The conditional

expectation of the observed Y given X is

$$E[Y|X, Z = 1] = X\beta + E[\epsilon|\omega > -X^*\gamma]$$

The estimated β is no longer unbiased if $\text{Corr}(\epsilon, \omega) \neq 0$. To correct the selection bias, Heckman (1979) proposes the two-step method under the assumption of the bivariate normal joint distribution of ϵ and ω , $\epsilon \sim N(0, \sigma)$ and $\omega \sim N(0, 1)$ $\text{Corr}(\epsilon, \omega) = \rho$. He also proposes the probit model for the selection equation. From the truncated bivariate normal distribution, the conditional expectation of error term becomes

$$E[\epsilon|\omega > -X^*\gamma] = \rho\sigma \frac{\phi(-X^*\gamma)}{1 - \Phi(-X^*\gamma)} = \rho\sigma \frac{\phi(X^*\gamma)}{\Phi(X^*\gamma)}$$

where ϕ is the standard normal density function and Φ is the cumulative distribution function of ϕ . The inverse Mills ratio in this model is $\frac{\phi(X^*\gamma)}{\Phi(X^*\gamma)}$ ($= \lambda$). Heckman shows that the selection bias can also be considered as omitted variable bias, and his two-step method includes the inverse Mills ratio as an independent variable in the outcome regression equation. In summary, (1) the probit model is used to estimate γ in the selection equation, (2) the inverse Mills ratio, λ , is computed, and (3) Y is regressed on X and λ .

After two predicted outcome (interval target) models are built from the treatment and control group data sets, the incremental sales model follows the same process as the incremental response model does with the difference scores in order to identify the customers who are likely to spend incrementally when they are targeted by the marketing campaign.

MODEL DIAGNOSTICS

Model diagnostics can be obtained through a heuristic method by ranking the control and treatment data sets in descending order by the difference score ($\bar{D}\bar{S}_i$). The ranked observations are divided into bins. If the number of bins is 10, the bins are deciles. For each bin, the average predicted values are calculated from both the treatment model and the control model. The predicted incremental response is the difference between the two average predicted values for each bin. Similarly, for each bin, the average observed responses are calculated from both models and the observed incremental response accordingly. Display 2 shows the model diagnostics from the estimated response model, which uses the example data set without prescreening variables.

| Table: Incremental Response Model Diagnostics | | | | | | |
|---|---------------------|-------------------|---------------------|--------------------|------------------|--------------------|
| Percentile | Predicted Treatment | Predicted Control | Predicted Increment | Observed Treatment | Observed Control | Observed Increment |
| 10 | 0.731977 | 0.354541 | 0.377435 | 0.735209 | 0.367389 | 0.367819 |
| 20 | 0.711739 | 0.36625 | 0.345489 | 0.722112 | 0.365462 | 0.35665 |
| 30 | 0.703177 | 0.375665 | 0.327512 | 0.712551 | 0.37253 | 0.340021 |
| 40 | 0.693125 | 0.380453 | 0.312672 | 0.683049 | 0.36989 | 0.313159 |
| 50 | 0.669399 | 0.371129 | 0.29827 | 0.668675 | 0.360558 | 0.308117 |
| 60 | 0.650663 | 0.367496 | 0.283167 | 0.636816 | 0.350754 | 0.286062 |
| 70 | 0.625271 | 0.358865 | 0.266407 | 0.640404 | 0.370297 | 0.270107 |
| 80 | 0.575347 | 0.330758 | 0.24459 | 0.57996 | 0.323123 | 0.256837 |
| 90 | 0.486066 | 0.27287 | 0.213196 | 0.487232 | 0.270323 | 0.216909 |
| 100 | 0.311545 | 0.165989 | 0.145556 | 0.310176 | 0.174847 | 0.135329 |

Display 2. The Model Diagnostics Table for Incremental Response

To constitute a good incremental response model, the top percentiles of actual data should have higher incremental rates, and the bottom percentiles should have lower incremental rates. In addition, the rates should decrease monotonically from the top to the bottom percentiles. According to this criterion, the table in Display 2 shows that the estimated model is reasonably good, as expected from the data generation. No percentiles violate the monotonicity. The prediction power of the model is also good enough, which means that the differences between the estimated and actual incremental response rates are small at all the percentiles.

The same diagnostic method can be applied for the incremental sales model. The data are sorted in decreasing order by the predicted incremental sales (difference scores in sales prediction), and the ranked observations are binned by using the predefined number of bins. Average values for each bin are calculated from both predicted and observed values. The binned incremental sales should decrease monotonically in a well-fitted incremental model.

| Table: Incremental Outcome Model Diagnostics | | | | | | |
|--|---------------------|-------------------|---------------------|--------------------|------------------|--------------------|
| Percentile | Predicted Treatment | Predicted Control | Predicted Increment | Observed Treatment | Observed Control | Observed Increment |
| 10 | 105.8938 | 25.40846 | 80.48539 | 140.1854 | 68.11443 | 72.071 |
| 20 | 89.02516 | 17.64444 | 71.38072 | 122.6114 | 55.53287 | 67.07858 |
| 30 | 84.4372 | 16.92435 | 67.51285 | 108.3317 | 56.77204 | 51.55965 |
| 40 | 80.69064 | 16.13369 | 64.55695 | 106.8882 | 51.20288 | 55.68536 |
| 50 | 77.68305 | 15.73781 | 61.94524 | 104.7295 | 44.70559 | 60.02387 |
| 60 | 74.23168 | 14.99389 | 59.23779 | 100.6642 | 43.12776 | 57.53647 |
| 70 | 70.06349 | 13.76138 | 56.30211 | 92.09026 | 38.6213 | 53.46896 |
| 80 | 63.82105 | 11.30586 | 52.51519 | 83.00709 | 34.65548 | 48.35161 |
| 90 | 52.45254 | 6.100948 | 46.35159 | 60.20342 | 23.12413 | 37.07929 |
| 100 | 30.08354 | 1.701214 | 28.38232 | 39.30228 | 15.21998 | 24.0823 |

Display 3. The Model Diagnostics Table for Incremental Outcome

In this example, the observed increment does not decrease monotonically; the values between the 30th and 50th percentiles are in increasing order. So this model is not good enough. A better model might be considered by tweaking some of model parameters. However, the model is not too bad, because the incremental values at the top 10th and 20th percentiles are higher than those at the bottom percentiles. The predicted values at all bins slightly miss the actual values.

In summary, for the model diagnostics, use the observed incremental value; then (1) compare top percentiles with bottom percentiles, (2) check the decreasing monotonicity, and (3) observe how close the predicted value is to the actual value.

SAS Enterprise Miner visualizes the table by using side-by-side bar charts for treatment and control response model diagnostics, incremental response model diagnostics, cumulative incremental response diagnostics, and so on.

INCREMENTAL REVENUE AND PROFIT ANALYSIS

Constant Revenue and Constant Cost

Consider a simple direct marketing campaign for one product. Assume that the response is the product purchase, the quantity is only one per customer, the marketing cost is \$2.00, the product cost is \$10.00, and the product price that the customer pays is \$40.00. Under these assumptions, you want to decide which customers are in the most profitable group in terms of incremental response. Hereafter, the terms of revenue and profit are exchangeable because they depend on how the cost (expense) is defined. The property setting in Display 4 sets up the analysis.

| Revenue Calculation | |
|----------------------|------|
| Use Constant Revenue | Yes |
| Revenue Per Response | 40.0 |
| User Constant Cost | Yes |
| Cost | 12.0 |

Display 4. Properties for Revenue Calculation

The Sales and Cost variables shouldn't be used in the variable setting. As Display 5 shows, you can get profitable customers up to the top 40 percentiles by using the estimated incremental response model. The average profit is also shown for each percentile.

| Table: Average Incremental Revenue | | |
|------------------------------------|---------------------------|------------------|
| Percentile | Average Revenue Increment | Profit Indicator |
| 10 | 3.097417 | Profitable |
| 20 | 1.819569 | Profitable |
| 30 | 1.100485 | Profitable |
| 40 | 0.506864 | Profitable |
| 50 | -0.06921 | Not Profitable |
| 60 | -0.67331 | Not Profitable |
| 70 | -1.34373 | Not Profitable |
| 80 | -2.21642 | Not Profitable |
| 90 | -3.47217 | Not Profitable |
| 100 | -6.17777 | Not Profitable |

Display 5. Incremental Revenue Table with Constant Revenue and Constant Cost

The incremental revenue (R_{incke}) is formulated with the same notation used before. The estimated revenues from using the treatment and control models, \hat{R}_T and \hat{R}_C , respectively, are

$$\hat{R}_T = \hat{Y}_T R_{perResp} - C \text{ and } \hat{R}_C = \hat{Y}_C R_{perResp}$$

$$\hat{R}_{incke} = \hat{R}_T - \hat{R}_C$$

where $R_{perResp}$ is the revenue per response and C is the cost. Following are the actual score codes for the incremental revenue calculation:

```
EM_REV_TREATMENT = EM_P_TREATMENT_RESPONSE*40.0 - 12.0;
EM_REV_CONTROL   = EM_P_CONTROL_RESPONSE*40.0;
EM_REV_INCREMENT = EM_REV_TREATMENT-EM_REV_CONTROL;
```

Revenue Variable and Constant Cost

The incremental sales model usually has a revenue variable. For example, suppose a customer who receives the promotion campaign visits the store and buys other products that are not listed in the promotion. A marketing analyst wants to measure the incremental sales impact from the promotional campaign and also wants to find the group of customers who might produce the incremental revenue. The data set should have an interval target variable in addition to a response target variable.

Instead of using the constant revenue, the interval target is used as a revenue variable. In the example, Sales is a revenue variable. Set the promotion cost at \$35.00, which might include some special discounts on other products in addition to the campaign cost itself. Display 6 shows the property setting and estimated profitable percentiles.

| Revenue Calculation | |
|----------------------|------|
| Use Constant Revenue | No |
| Revenue Per Response | 10.0 |
| User Constant Cost | Yes |
| Cost | 35.0 |

| Table: Average Incremental Revenue | | |
|------------------------------------|---------------------------|------------------|
| Percentile | Average Revenue Increment | Profit Indicator |
| 10 | 43.7074 | Profitable |
| 20 | 29.51162 | Profitable |
| 30 | 22.39475 | Profitable |
| 40 | 16.33803 | Profitable |
| 50 | 10.72258 | Profitable |
| 60 | 4.990902 | Profitable |
| 70 | -1.2728 | Not Profitable |
| 80 | -8.11032 | Not Profitable |
| 90 | -16.0234 | Not Profitable |
| 100 | -26.9633 | Not Profitable |

Display 6. Incremental Revenue Table with Constant Cost and Revenue Variable

The incremental sales model has two predicted variables: \hat{Y}_{Resp} from the response (selection) model, and $\hat{Y}_{Outcome}$ from the outcome (interval target) model. So the incremental revenue is formulated as follows:

$$\hat{R}_T = \hat{Y}_{Resp,T} * \hat{Y}_{Outcome,T} - C$$

$$\hat{R}_C = \hat{Y}_{Resp,C} * \hat{Y}_{Outcome,C}$$

$$\hat{R}_{incke} = \hat{R}_T - \hat{R}_C$$

The actual score codes for the incremental revenue calculation from the score code are as follows:

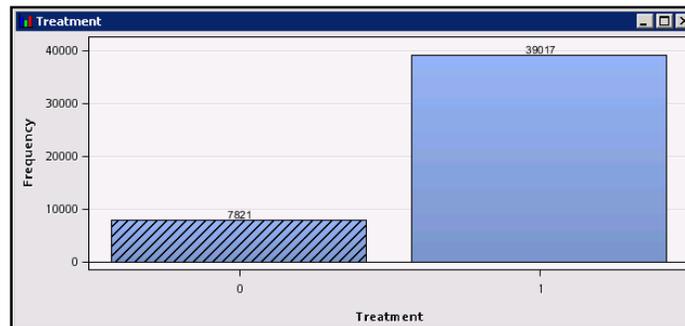
```
EM_REV_TREATMENT = EM_P_TREATMENT_RESPONSE*EM_P_TREATMENT_OUTCOME - 35.0;
EM_REV_CONTROL   = EM_P_CONTROL_RESPONSE*EM_P_CONTROL_OUTCOME;
EM_REV_INCREMENT=EM_REV_TREATMENT-EM_REV_CONTROL;
```

Cost Variable

Instead of constant cost, a cost variable can be used in both of the preceding cases. The cost variable should be specified as “Cost” in the variable editor, and the property of **User Constant Cost** should be set to No.

A CASE STUDY

A real data set with masked variable names from a bank is used in this case study. It contains 197 variables including one treatment variable that indicates whether the customer is targeted by a specified marketing action, one ID variable, and one binary target whose value equals 1 if the customer responds to the action and 0 otherwise. As seen in Display 7, the control group consists of 7,821 customers, and the treatment group consists of 39,017 customers.



Display 7. Exploration of the Treatment Variable

In SAS Enterprise Miner, the data source is created by the Advance Advisor, and it is added to the process flow diagram. Then a Data Partition node is added, which splits the data to create a training data set that has 60% of the observations and a validation data set that has 40% of the observations. More specifically, the variables Treatment and Target are each assigned the partition role of Stratification, as shown in Display 8. This ensures that the training data set does not contain a biased representation of either those who receive the treatment or those who do not receive the treatment.

| Name | Partition Role | Role | Level |
|-----------|-----------------------|-----------|--------|
| Treatment | Stratification | Treatment | Binary |
| Target | Stratification | Target | Binary |

Display 8. Partition Roles in the Partition Node

Then the Incremental Response node is connected to the Data Partition node, as shown in Figure 3.



Figure 3. Process Flow Diagram with the Input Data Source and the Partition Node

Most of the default settings in the Incremental Response node are used, except for the property changes that are shown in Display 9.

| | |
|----------------------------|------|
| Rank Percentage Cutoff | 10 |
| Two-Way Interaction | Yes |
| Revenue Calculation | |
| Use Constant Revenue | Yes |
| Revenue Per Response | 10.0 |
| User Constant Cost | Yes |
| Cost | 0.5 |

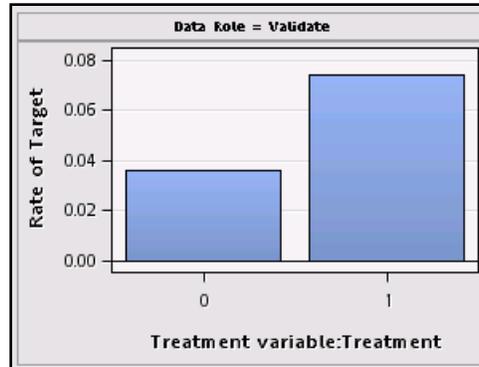
Display 9. Property Setting at the Incremental Response Node

The Train properties indicate that prescreening (variable selection) will select a subset of predictive variables based on the penalized NIV (because a validation data set exists), prior to building the model. Specifically, the variables in the top decile will be used to build the model; the top 10% of variables that have the strongest correlation to the model response will be used, and the rest will be rejected. The two-factor interactions will be included in the model, and the rest of model selection properties are default. The Revenue Calculation properties indicate that a constant revenue of \$10.00 per response will be assumed, with a constant cost per response of 50 cents. As a default, reports will display the results in deciles (10 bins).

Prior to building the predictive model, variable prescreening is performed to identify the variables that are most likely to maximize the incremental response rate. Display 10 shows the resulting table that contains the NIV, penalized NIV, and the selection status of input variables. It is available from the **View** menu under **Model** in the Results window. The variables are ranked by the penalized NIV because a validation data set is used. The top 10% of variables are shown as the preselected variables.

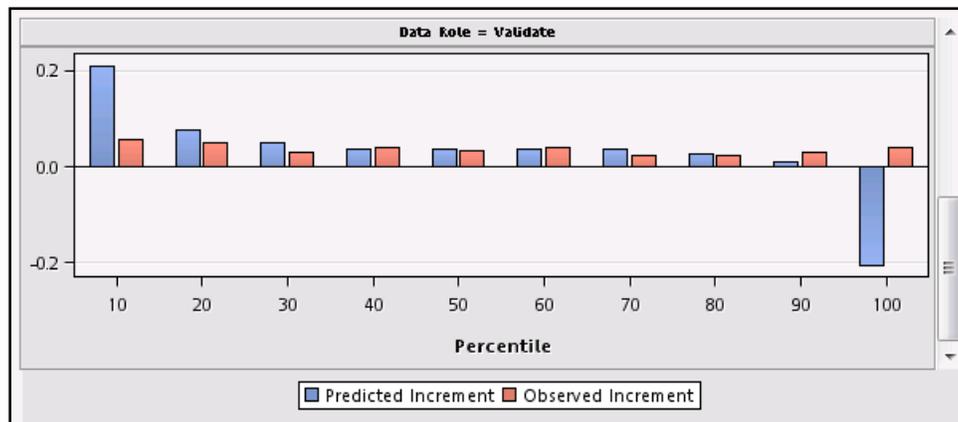
| Variable | Penalized Net Information Value ▼ | Net Information Value | Rank Percentile | Selection |
|----------|-----------------------------------|-----------------------|-----------------|-----------|
| Input100 | 96.65104 | 195.751 | 0.515464 | Yes |
| Input158 | 54.34022 | 60.3497 | 1.030928 | Yes |
| Input143 | 40.15586 | 72.36653 | 1.546392 | Yes |
| Input27 | 22.94669 | 162.4548 | 2.061856 | Yes |
| Input116 | 22.31692 | 70.03406 | 2.57732 | Yes |
| Input69 | 20.2802 | 25.54924 | 3.092784 | Yes |
| Input31 | 19.06357 | 44.14563 | 3.608247 | Yes |
| Input25 | 14.53473 | 81.73879 | 4.123711 | Yes |
| Input73 | 13.84242 | 27.7278 | 4.639175 | Yes |
| Input52 | 13.05852 | 29.73999 | 5.154639 | Yes |
| Input187 | 12.18263 | 53.20336 | 5.670103 | Yes |
| Input118 | 11.58659 | 70.39867 | 6.185567 | Yes |
| Input139 | 11.0304 | 34.18164 | 6.701031 | Yes |
| Input81 | 9.559246 | 66.40947 | 7.216495 | Yes |
| Input18 | 7.635334 | 24.9277 | 7.731959 | Yes |
| Input164 | 6.308685 | 25.59682 | 8.247423 | Yes |
| Input78 | 6.183295 | 7.671384 | 8.762887 | Yes |
| Input183 | 5.696995 | 35.15615 | 9.278351 | Yes |
| Input161 | 5.485202 | 9.609462 | 9.793814 | Yes |
| Input65 | 5.414703 | 5.563288 | 10.30928 | No |
| Input46 | 4.970878 | 9.434491 | 10.82474 | No |

Display 10. Top 10% of Input Variables Ranked by the Penalized Net Information Value Score



Display 11. Response Outcome Summary Plot

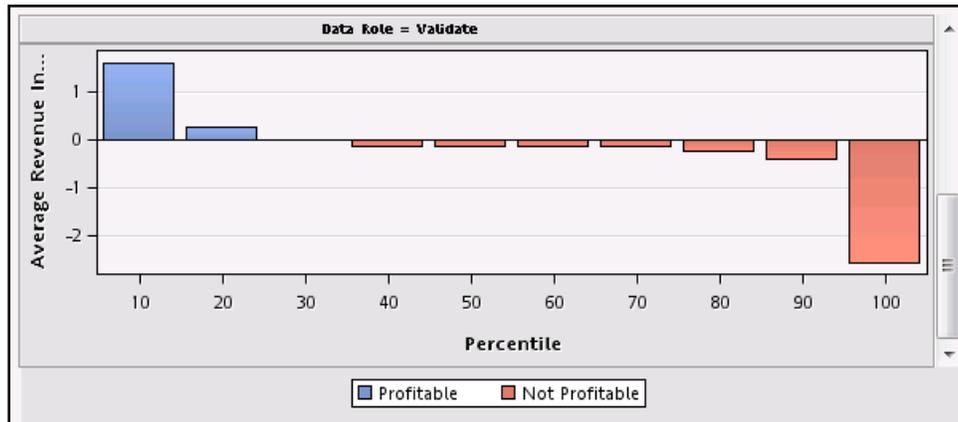
The response outcome summary plot in Display 11 shows that in the control group (Treatment=0), 3.6% of the customers in the validation data set responded (Target=1). These customers responded even though they did not receive the marketing incentive. In the treatment group (Treatment=1), 7.4% of the customers in the validation data set responded (Target=1). These customers include those who will respond regardless of the marketing incentive and those who are likely to respond only when they receive a marketing incentive. To maximize the return on investment, target the second type of customers. Recall that the difference in response rate between the treatment and control groups is the incremental response rate overall. According to these basic statistics, it is expected that the average incremental response rate from these customers is $7.4\% - 3.6\% = 3.8\%$.



Display 12. Incremental Response Model Diagnostics Plot

The incremental response model diagnostics plot in Display 12 shows both the predicted and observed incremental response rate by decile. The top decile has the highest incremental response rate, with a predicted increment of 21.2% and an observed increment of 5.8%. This predicted increment is approximately five times higher than the average incremental response rate of the data (3.8%). The predicted increment is calculated as the difference in the predicted propensity of response between the treatment and control groups. The customers are then ranked according to the predicted increment values so that the top deciles of customers are those who are more likely to respond if they receive the marketing incentive. Within each decile, the observed increment is calculated as the difference between the averages of the actual response rates of the treatment and control groups. The observed incremental response rate examines whether the model does identify the optimal customers, as described earlier. This plot, along with the average incremental revenue plot (Display 13), helps to make decisions about how to target a small portion of the whole population while still generating a higher incremental response than that achieved by targeting the whole population. The plot also identifies customers who should not be contacted. In this example, the last decile has the predicted increment of -20.4% ; this could mean the marketing incentive has a negative impact on these customers and could reduce the response rate if they are targeted.

The average incremental revenue plot assumes a constant revenue of \$10.00 per response and a constant cost of 50 cents per incentive sent. From these values, the expected incremental revenue by decile can be calculated, as shown in Display 13.



Display 13. Average Incremental Revenue Plot

If the predicted response rate multiplied by the revenue per response less the cost per contact is greater than 0, then the customer is considered profitable. Display 13 shows that the first three deciles contain profitable customers (under these cost and revenue assumptions).

In conclusion, to maximize the return on investment for this particular marketing incentive, the top 30% of customers, as ranked by the average incremental revenue plot, would be the profitable customers under the assumptions of a constant revenue of \$10.00 and a constant cost of 50 cents. This can also be seen in the corresponding table in Display 14.

| Average Revenue Increment | Percentile | Data Role ▼ | Profit Indicator |
|---------------------------|------------|-------------|------------------|
| 1.62118 | 10 | Validate | Profitable |
| 0.285614 | 20 | Validate | Profitable |
| 0.012646 | 30 | Validate | Profitable |
| -0.1316 | 40 | Validate | Not Profitable |
| -0.13861 | 50 | Validate | Not Profitable |

Display 14. Average Incremental Revenue Table

SUMMARY

This paper illustrates the importance of the incremental response model in direct marketing campaigns and shows how to use SAS Enterprise Miner to build the model. The modeling techniques include a specialized variable selection method through the net information value or the penalized net information value. Heckman's two-step method is also demonstrated for the incremental sales model to correct the selection bias. Because of the nature of data in this area, traditional modeling techniques cannot be applied to this problem, and no one method among several well-known methods is superior to the others. Many marketing analysts still seek better modeling techniques. SAS Enterprise Miner will continue to improve both the modeling techniques and the diagnostic methods.

REFERENCES

- Hansotia, B. and Rukstales, B. (2002). "Incremental Value Modeling." *Journal of Interactive Marketing* 16:35–46.
- Heckman, J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–161.
- Larsen, K. (2010). "Net Lift Models: Optimizing the Impact of Your Marketing Efforts." SAS Course Notes. Cary, NC: SAS Institute Inc.
- Lo, V. (2002). "The True Lift Model: A Novel Data Mining Approach to Response Modeling in Database Marketing." *ACM SIGKDD Explorations Newsletter* 4:78–86.
- Radcliffe, N. and Surry, P. (1999). "Differential Response Analysis: Modeling True Response by Isolating the Effect of a Single Action." *Proceedings of Credit Scoring and Credit Control VI*. Edinburgh: Credit Research Centre, University of Edinburgh Management School.

Siddiqi, N. (2005). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. Vol. 3. New York: John Wiley & Sons.

ACKNOWLEDGMENTS

The authors thank Jared Dean and Wayne Thompson for their encouragement and support for this paper.

CONTACT INFORMATION

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

Taiyeong Lee
E-mail: Taiyeong.Lee@sas.com

Ruiwen Zhang
E-mail: Ruiwen.Zhang@sas.com

Xiangxiang Meng
E-mail: Xiangxiang.Meng@sas.com

Laura Ryan
E-mail: Laura.Ryan@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.