Notes on Estimating Probabilities of Success

1. Overview

The CDS business solution to Third Party Marketing (3PM) operations in the mail channel depends heavily on Variable Price (VP) calculations to eliminate the systemic mis-classification of profitable marketing opportunities as unprofitable. VP, in turn, depends heavily upon Customer(Mail Recipient) specific Probabilities Of Success (POS). It is reasonable to ask two questions:

1.1. How can the causal factors of Customer specific POS be estimated accurately, and
1.2. How effective are these factor estimates in creating new profitable marketing opportunities

We have developed an app that provides some insight into how these two tasks can be accomplished in the real world. This document outlines how that app works. The description below is primarily intended for non-technically oriented managers. For ease of explanation, there may be some slight differences between the description and what takes place in reality within the app. These are inconsequential and do not invalidate any of the arguments.

2. Overview of Customer Characteristics

Individuals in a population have many attributes such as:

2.1. Age
2.2. Income
2.3. Gender
2.4. Marital Status
2.5. FICO Score
2.6. Education
2.7. Home Ownership
2.8. Religion
2.9. etc. ...

Every one of these perspectives can be sub-divided into "bands". For example, Age Bands could be defined in bands of 5 years: 21-25, 26-30, 32-31 and so on. Education could be banded by years in school: 12 for a high school graduate, 14 for a community college degree, 16 for a four year degree, and so on. Banding is performed because it logically segments populations into categories of people that can be marketed by different means that presumably generate higher overall profits. Assuming beer companies know that single males between 21 and 40 generally consume more beer than other segments of the population it is reasonable to assume the companies would look for advertising opportunities that "reach" these segments such as sports events. Conceptually, VP simply takes the strategy of leveraging segmentation to its logical conclusion: segmenting to the level of the individual or household. These notes will focus on the individual approach in its examples. The structure of the argument, however, is equally applicable to households.

Once banding has taken place, every individual in a population can now be given a "profile" that describes his characteristics. Profiles are sets of numbers that are assigned to every band of every attribute. These numbers are not randomly assigned - they are assigned in a way that facilitates mathematical quantification procedures on just how much, on average, customer membership in attribute bands contributes to differences in POS. For dichotomous attributes like Marital Status, Gender, and Home Ownership where there can only be one of two values the relevant profile entries would only consist of: 0 or 1. For continuous attributes such as Income the profile might number might be based on band values of monthly income to the nearest $1000. For example, an individual with $1400 a month of income would be given a profile value of 1 in Income. An individual with $3900 in monthly Income would be given a profile value of 4 for Income and so on. For those with an interest, categorical attributes like religion would need to be expressed as a series of mutually exclusive dichotomous values. For example, Catholic (0 or 1), Methodist (0 or 1) ... and so on.

While Marketers are aware that bands in attributes of populations have different success rates (in the form of sales) they are not generally accustomed to computing customer specific POS and negotiating prices for that access to that individual customer. But OPTIM requires that these differences in success/sales rates be calculated so that customer specific POS can generate customer specific CAC (Customer Access Charges) using VP. The app provides that demonstration.

3. Outline of App Processing

In order to answer questions 1.1 and 1.2 above the app takes the following actions to develop a simple test population:

3.1. it allows for every customer (mail recipient) to have 3 attributes:
  3.1.1. gender with two bands representing Male and Female
    3.1.1.1. 0 (Male)
    3.1.1.2. 1 (Female)
  3.1.2. marital status with two bands representing Married and Single
    3.1.2.1. 0 (Single)
    3.1.2.2. 1 (Married)
  3.1.3. monthly income in 1000s with 12 bands
    3.1.3.1. 1 ($1K per month)
    3.1.3.2. 2 ($2K per month)
    ...
    3.1.3.3. 12 ($12K per month)

  With all attributes banded, each customer can be given a profile. For example:

  3.1.4. a male, married, with $5K income per month would have his profile vector as [0,1,5]
  3.1.5. a female, single, with $7K income per month would have his profile vector as [1,0,7]

3.2. it allows for user defined differences in theoretical impact on POS by each attribute. These differences are coefficients or "weights" that are used to multiply a customer profile to generate a single theoretical POS for that customer. For example, suppose these were the attribute coefficients (applicable to all customers):

  3.2.1. gender: 0.0020
  3.2.2. marital: 0.0040
  3.2.3. income: 0.0005

  Given these theoretical coefficients, we can develop a customer specific **<u>theoretical</u>** POS by computing the sum of the products of the profile elements multiplied by its associated coefficient. Using the two customers in 3.1.1 and 3.1.2 as examples we would get:

  3.2.4. POS = [0.0020,0.0040,0.0005]*[0,1,5] = 0.0020*0 + 0.0040*1 + 0.0005*5 = 0.0065
  3.2.5. POS = [0.0020,0.0040,0.0005]*[0,1,5] = 0.0020*1 + 0.0040*0 + 0.0005*7 = 0.0055

  Marketing to ten thousand customers like 3.2.2 has a **<u>theoretical</u>** expectation of 65 sales.
  Marketing to ten thousand customers like 3.2.3 has a **<u>theoretical</u>** expectation of 55 sales.

3.3. it allows for a user defined approximate number of customers that will be simulated to receive marketing messages.

3.4. it allows for a user defined Gross Profit per successful response from a customer

3.5. it allows for a user defined Print Cost for the Insert

3.6. it allows for a user defined desired ROI by the marketer

To this point, the user has provide some information on the **<u>theoretical</u>** values of the coefficients that are associated with the profile components of each customer. However, when just starting a marketing campaign a Marketer may not know anything about what these values are. In point of fact, the Marketer can never exactly know what the true values of these **<u>theoretical</u>** coefficients are. However, the app will show that these **theoretical** values can be estimated sufficiently accurately to to allow their effective use by VP in future Marketing campaigns. The app demonstrates this in the following manner:

3.7. the app constructs a "core set" of customers that contains every possible combination of profile values. Since there are 2 gender values, 2 marital values, and 12 income values there are 48 possible combinations of profiles (2*2*12=48). This "core set" is a template that also can keep a running total of offers made and sales generated for each of the 48 unique combinations of profiles

3.8. given the number of customers in the "core set", the app computes the minimum number of "replicates" of the "core set" the app needs to generate at least the user defined number of test customers. For example, if the desired number of customers to be simulate is 1 million then the number of replicates is basically 1,000,000 divided by 48 = 20,834 (rounded up to the nearest integer). We will use this replicate value of 20,834 further in this example.

3.9. for each of the 48 customers in the "core set", the app computes the **theoretical** POS using the vector multiplication process described above in 3.2.4 and 3.2.5.

3.10. given the number of "replicates" of the "core set" the app simulates a test mailing to a large number of customers by repeating the following process 20834 times:

3.10.1. increment the offers count for this profile record by 1
3.10.2. spin a random number generator that returns a number between 0 and 1 (exclusive of the end points). if that random number is less than the **theoretical** POS for the unique profile record then increment the sales count for this profile record by 1
3.10.3. the app records all customer information (demographics, offer, and sale) by adding a line to a "marketing results" file

The basic idea is that over a large number of simulated test mailings that the sales rates for the profile records will be highly correlated with the POS. That is the profile records with higher POS will have higher sales and vice versa. But the POS are simply sums of the profile component numbers for the population. There is a well established mathematical procedure called "regression" that can estimate what the profile component numbers are for the population in these cases.

3.11. so the app it reads in the "marketing results" file and runs the "regression" procedure on the data to estimate what the true population profile values are for gender, marital, and income attributes. The "regression" software is freeware called "R" and is readily downloadable over the inernet/

The true population profile values are not exactly knowable but the practical question is whether or not the estimates derived from "regression" can be used in their place to do what VP is meant to do:

3.11.1. generate customer specific VP to expand the number of profitable marketing opportunities
3.11.2. to accomplish the expansion while maintaining the same ROI

3.12. once the app has the profile value estimates it runs through an iterative procedure similar to the test marketing simulation in that it will replicate processing of the "core set " 20,834 times. For each core set record it will:

3.12.1. generate a customer specific POS estimate based on the regression estimates
3.12.2. "back out" a VP estimate based on the POS estimate, Gross Profit, and Print Cost
3.12.3. if the VP is > 0 then increment the offers count by 1
3.12.4. spin a random number generator as before, if the random number is less than the **theoretical** POS then increment the sales count by 1

It is important to understand what the app is doing at this point. It is running a simulated sales campaign based on its **estimates** of the population profile parameters but the simulation of actual sales is based upon the **theoretical** population profile values. This is exactly what is needed.

3.13. at this point, each of the 48 "core set" records (market segmentations) has enough information to compute ROI (total CAC, total sales, total Gross Profit, total Print Costs, etc.) Further, judiciously summing these quantities is enough to calculate effective ROI at higher level aggregations. These are shown in the charts of the app. Below is an instance of the app with various different impacts on individual POS by attribute band ("betas" are the profile coefficient estimates).

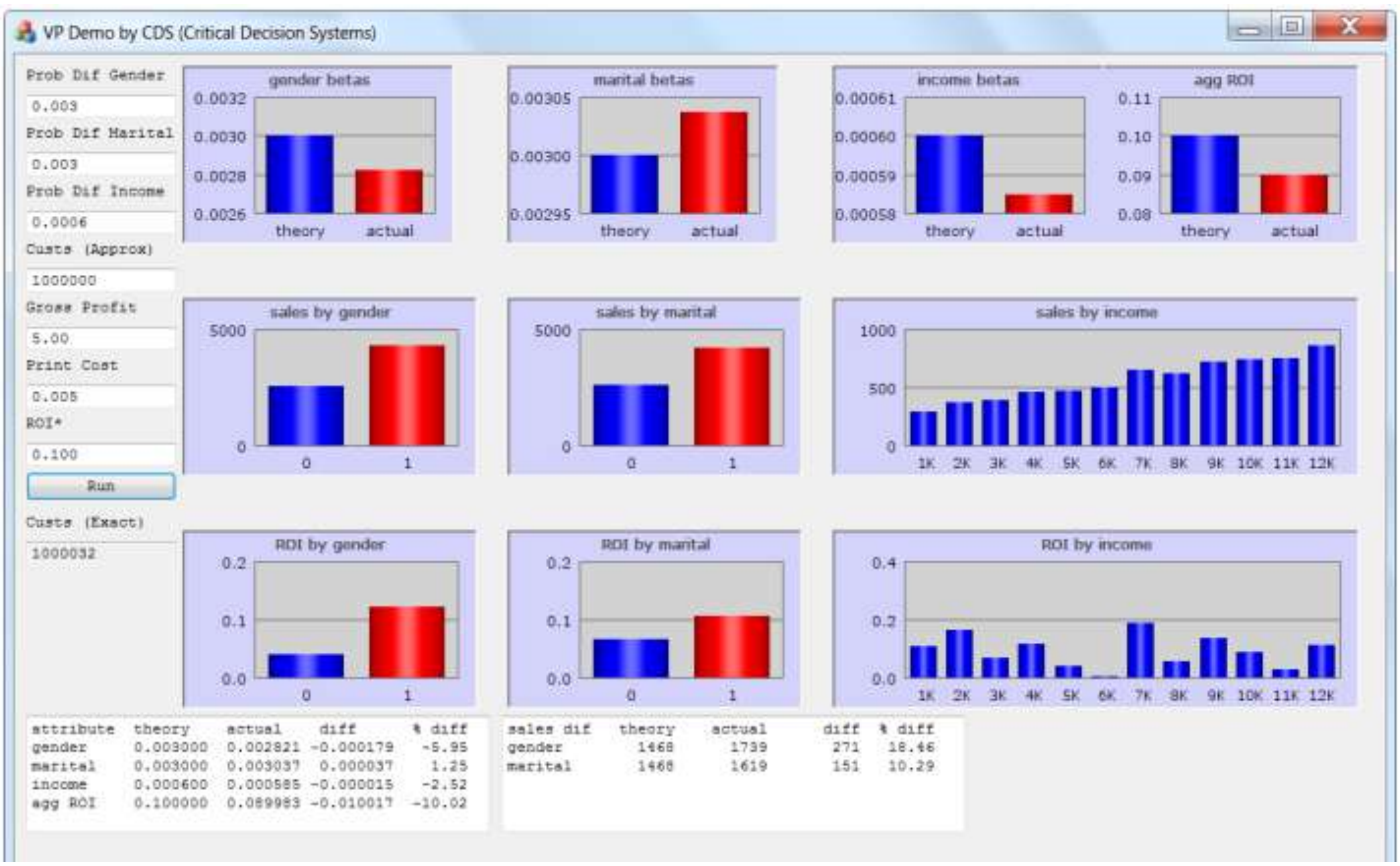
4. Quick Conclusions

The takeaways should be that:

4.1. "regression" is quite capable of accurately estimating the **theoretical** population profile values
4.2. more importantly, ROI on the expanded base of customers meets expectations and needs, and
4.3. the ROI by income band reassuringly fluctuates around the desired ROI of 0.100.

We have run repeated simulations of the app with different settings. There is almost uniform replication of results similar to these. At some point in the future we will provide more clinically detailed results of a wide ranges of runs to replace our sense that this works with clinical proof. Additional runs under different settings are provided below. These display similar results.


VP Demo by CDS (Critical Decision Systems)
Prob Dif Gender
0.004
Prob Dif Marital
0.002
Prob Dif Income
0.0005
Custs (Approx)
1000000
Gross Profit
5.00
Print Cost
0.005
ROI*
0.100
Run
Custs (Exact)
1000032
gender betas
marital betas
income betas
agg ROI
theory
actual
sales by gender
sales by marital
sales by income
ROI by gender
ROI by marital
ROI by income
1K 2K 3K 4K 5K 6K 7K 8K 9K 10K 11K 12K
attribute theory actual diff % diff
gender 0.004000 0.003919 -0.000081 -2.03
marital 0.002000 0.002195 0.000195 9.73
income 0.000500 0.000487 -0.000013 -2.66
agg ROI 0.100000 0.128133 0.028133 28.13
sales dif theory actual diff % diff
gender 1916 2196 280 14.61
marital 958 1064 106 11.06


VP Demo by CDS (Critical Decision Systems)
Prob Dif Gender
0.003
Prob Dif Marital
0.003
Prob Dif Income
0.0006
Custs (Approx)
1000000
Gross Profit
5.00
Print Cost
0.005
ROI*
0.100
Run
Custs (Exact)
1000032
gender betas
marital betas
income betas
agg ROI
theory
actual
sales by gender
sales by marital
sales by income
ROI by gender
ROI by marital
ROI by income
1K 2K 3K 4K 5K 6K 7K 8K 9K 10K 11K 12K
attribute theory actual diff % diff
gender 0.003000 0.002821 -0.000179 -5.95
marital 0.003000 0.003037 0.000037 1.25
income 0.000600 0.000585 -0.000015 -2.52
agg ROI 0.100000 0.089983 -0.010017 -10.02
sales dif theory actual diff % diff
gender 1468 1739 271 18.46
marital 1468 1619 151 10.29