The CHAID Analysis

The CHAID Analysis (Chi Square Automatic Interaction Detection) is a form of analysis that determines how variables best combine to explain the outcome in a given dependent variable. The model can be used in cases of market penetration, predicting and interpreting responses or a multitude of other research problems.

CHAID analysis is especially useful for data expressing categorized values instead of continuous values. For this kind of data some common statistical tools such as regression are not applicable and CHAID analysis is a perfect tool to discover the relationship between variables.

One of the outstanding advantages of CHAID analysis is that it can visualize the relationship between the target (dependent) variable and the related factors with a tree image.

CHAID Analysis for Surveys

Most survey answers have categorized values instead of continuous values. Finding out the statistical relationship in this kind of data is a challenge. CHAID Analysis is a good solution for such survey studies.

Case Study: Composter Usage in a Major Canadian City, 2007.

CHAID Analysis for Customer Profiling

Based on historical customer data, CHAID Analysis can be used to analyze all characteristics within the file, for example, product/service purchased, dollar amount spent, major demographics and demography of the customers, and so on. A blue print can be produced to provide an understanding of the customer profile: strong or weak sales of products/services; active or in-active customers; factors affecting customers' decisions or preferences, and so on.

Such customer profile will give the Sales & Marketing Team a clear picture of which type of person is most likely to buy the products and services based on factual purchase history, geo-demographics and lifestyle attributes.

Case Study: "Health Insurance Customer Profiling" for a national insurance company, 2007.

CHAID Analysis for Customer Targeting

Recruiting new customers via direct contact (phone or mail) is a time consuming and costly effort. For most products or services, the hit rate is less than 1%. That means, in order to get a new customer, over one hundred contacts are required.

By mapping the current customer list to a general population database (e.g., SMR Residential Database that contains 12 million listed households), CHAID Analysis can find the household clusters that have much higher incidence rates than the average. By concentrating on these household clusters, the actual hit rate can be dramatically raised. The result is "Fewer phone calls or mail pieces with higher sales returns!"

Case Study: "Potential Customer Indexing" for a company providing residential services, 2007, 2008.

Explanation of CHAID Analysis of Company X Contracts

Number of records in node. One record = one household which had at least one existing person with a Company X contract.

Average penetration of study area as percent of households. Note the "study area" is made up of all FSA's where existing penetration is at least 1%. We do this so as to not over-consider areas where promotion may not have been significant to date. In other words, the 1% penetration rule gives us a study area with evidence of existing product awareness.

Penetration

Node 0

Mean	3.5135
Std. Dev.	25.4518
n	189667
%	100.00
Predicted	3.5135

Variable selected by CHAID as the strongest correlation to sub-divide data into the next level with more predictable penetration, either high or low.

Highrise_pct
Adj. P-value=0.0000, F=1214.4444, df=2,189664

<=0

>0

Value that defines criteria for sub-tree. So this sub tree is represents all neighborhoods where the percentage of high-rise dwellings (i.e. apartments) is greater than zero.

Node 27

Mean	2.7190
Std. Dev.	11.6728
n	169504
%	89.37
Predicted	2.7190

Node 28

Mean	13.5496
Std. Dev.	82.2928
n	14246
%	7.51
Predicted	13.5496

Node 29

Mean	2.1119
Std. Dev.	17.1834
n	5917
%	3.12
Predicted	2.1119

Predicted penetration of prospects residing in neighborhoods meeting "node criteria".

This number also defines "lift". The predicted uptake to prospects is 13.5496 and the base rate is 3.5135, so the lift is 13.5496/3.5135 = 386%. This means that DM to prospects meeting this criteria should result in 386% more business than random DM.