3. *ATTRIBUTE (MULTIVARIATE) EVALUATION*[1]

Page

3.1 INTRODUCTION

Attribute studies concentrate on stakeholder perception and preferences of performance of an organization and its offerings. These are based on general characteristics, benefits or attributes of the firms and its products and services.

3.1.1 STUDY OBJECTIVES

Typically these studies are undertaken to determine how well the organization is doing and where should it concentrate its efforts to improve? However, they can differ markedly based on the specific objectives of the study. Most surveys are limited and some compromise in coverage has to be made. What is excluded in terms of the number of attributes and measures should depend on the eventual use of the information. It should be noted that most of these studies are used for multiple purposes. However, usually a single purpose dominates.

3.1.1.1 Performance Benchmarking

The objective of performance benchmarking is to determine How well is the organization doing?

3.1.1.1.1 Stakeholder Satisfaction

Stakeholders include customers, employees, potential owners (stockholders) and other decision-makers. These are usually tracking studies where the client wishes to track progress and changes in the market. A key phrase here is to “Avoid Being Blind Sided.” Customer Satisfaction Studies are also used for acquisition and business evaluation.

[1] An on-line statistical text can be found at http://www.statsoftinc.com/textbook/stathome.html

3.1.1.1.2 Quality Studies (TQM)

Total Quality Methodology requires measurement of performance. As such, attribute studies are undertaken to provide a basis for business process re-engineering. The focus is what should we improve or work on.

3.1.1.1.3 Bonus Assignment

Management, sales and operations personnel bonuses are being linked to the results of customer satisfaction studies. Typically, these bonuses are determined by meeting minimum performance standards as measured by reported overall satisfaction.

3.1.1.2 Marketing

From a marketing perspective, these studies are the "Voice of the Customer." They provide insight into what customers consider important and what drive their decisions as well as his opinion of the firm and its competition.

3.1.1.2.1 Segmentation

Segments are groups of customers with similar characteristics that can be approached using a common marketing strategy. We seek out segments by examining the perception of attributes and what they consider important.

3.1.1.2.2 Positioning

From an attribute analysis perspective, organizational and product positioning relate to how groups of our customers perceive how products compared to the competition. It is a process of seeking competitive advantage.

3.1.1.2.3 Communications (Advertising)

Attribute analysis for communications focuses on identifying attributes and characteristics to emphasis for marketing purposes. Usually, studies designed for communications purposes will also collect information on media and selection of preferred communication channels.

3.1.2 TYPES OF DATA

There are four general types of data that are typically collected in these studies: (1) Performance, (2) Importance, (3) Behavior, (4) Demographics and Psycho-Sociographics. Most studies collect some of each of these types.

3.1.2.1 Perception (Attribute Performance)

Measuring attribute performance is the main objective of these studies. Attributes usually include overall evaluations and general attitudes. Rating scales are typically used and there may be up to 60 attributes. However, no more than 20 are usually considered with 10 to 15 typical with competitive studies.

Estimates of attribute performance are typically collected on the client's firm or product and one to five competitors. In tightly structure markets the other competitors are specified. In less structured situations, the respondent will identify the key competitors[2].

3.1.2.2 Preference (Importance)

In most cases the respondents will be asked to estimate the importance of the attributes on their actions. This is often some type of constant-sum exercise. With a large number of items, this is done with a nested structure where the attributes are grouped. The importances of components of the groups are estimated as well as the groups themselves. However, there are many other methods used including ratings and rankings.

3.1.2.3 Behavioral

These questions capture the behavior of the respondents. There are two types: (1) historical behavior such as having purchased specific products or brands, and (2) intent questions that focus on expected behavior. When they are included, behavior questions get to the "bottom-line" of marketing studies. Often purchase data may be added subsequently to the study based on business transaction records or other sources[3]. Behavioral data, particularly the identification of customers are used to qualify respondents and stratify the sample.

3.1.2.4 Demographics and Psycho-Sociographics (Value)

Demographic data focuses on the characteristics of the respondents. Behavior may also be included in this category. It answers who these respondents are. This data is usually used for segmentation and like behavioral data, to stratify the sample. These questions often focus on the channels and conditions of purchase as well as traditional distinctions (age, gender, education, income, and ethnicity). However, issues such as ethnicity are not recommended for "boiler-plate" questions. They can irritate the respondents and are usually not insightful unless required by product specificity or market targeting.

Psycho- and sociographic questions focus on underlying respondent values and interests. These are typically used for segmentation, particularly with consumer research and are associated with marketing communications issues. Usually these questions are not included in business-to-business (industrial) research.

3.1.3 DATA ANALYSIS

[2] There are many variations on this theme depending on the competitive structure.

[3] If a panel is being used, behavioral, demographic and sociographic data may be collected separately.

Data analyses of attribute evaluation studies are either univariate and bivariate procedures, that focuses on one or two specific attributes, and multivariate analyses that addresses the interactions of groups of variables. Most of the analyses discussed in this chapter will be multivariate analyses. Most univariate and bivariate analyses consist of tabulations of data. Tabulations and cross-tabulations are still the primary means of data analysis. Multivariate analysis should only be undertaken after tabulations have been investigated.

3.1.3.1 Visualization

The major purpose of multivariate data analysis is the "visualization" of the information. The large number of variables and possible combinations makes it difficult to "see the forest from the trees". A major side effect of this type of statistical analysis is that its results by their nature are imprecise. It is, therefore, always necessary to return to the univariate and bivariate, cross-tabular analyses to confirm results and to drill down into the details.

3.1.3.2 Overview

The following chart shows the outline for analysis of attribute data. Basically it follows from the two basic types of data: (1) perception or performance of attributes and (2) preference or importance of attributes. The other information is used mainly to provide alternative sub-groups in this analysis. The specific tools that are selected depend on the scaling of the data, the nature of the observed relationships, and the needs of the clients. The next section will look into the scaling of the data and subsequent sections in to the types of analyses starting from the clients' perspective.

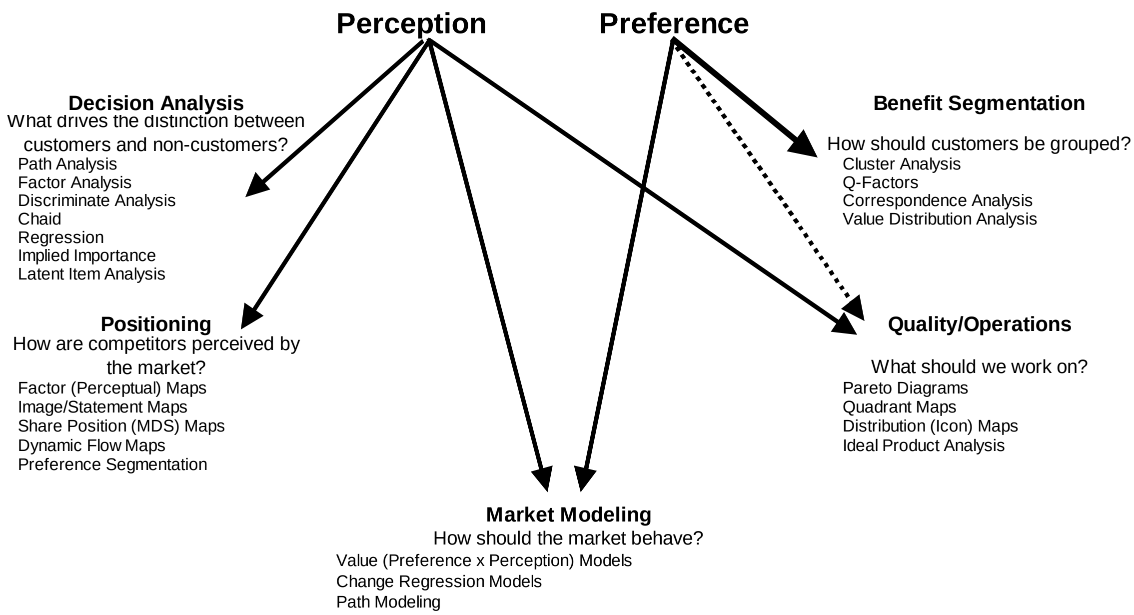
Analysis of Perception Data

Types of Data

Competitive Perceptions of Features, Attributes, Benefits, and Images
Preference (Importance) on Features, Attributes
Application and Use Information
Respondent Demographics and Psychographics

Types of Studies

Customer Satisfaction
Market Audits
Segmentation
Positioning
Communications Studies



3.2 TYPES OF QUESTIONS (DATA SCALING)

The nature of the responses to questions affects the range of statistical tools available and the accuracy (meaningfulness) of the results. There are four traditional classifications of "data scaling": (1) categorical, (2) ordinal, (3) interval, and (4) ratio scaling. Categorical or normative scaling consists of identifying a condition. It is a yes/no response. Ordinal responses are on a scale where we only know that higher values are consistently better or worse than lower numbers. Ranges from 1 to 10 are typical of this type of scaling. However, these scales may have other properties as well.

Interval and ratio scales are sometimes referred to as "quantitative" data since we can apply standard statistical tools (such as regression and factor analysis) to them. Interval scales are defined as having equal meaningful values between points. That is, on a 10-point scale the value of a change from 1 to 2 has the same value as a change from 4 to 5. Ratio scale data, in addition to consistent intervals, have a nature zero or standard reference value. Example of ratio scale data is size.

It should be noted that data can often be collected in any of the scaled types. Categorical data is usually the simplest. It is typically in the form of having the respondent select from a set of alternatives. Ordinal and interval data usually use a type of rating. Ratio data often requires the respondent to put in a number of values to a question.

3.2.1 CATEGORICAL (NORMATIVE) SCALED DATA

Typically most of the demographic and behavioral information is collected as categorical data. The key problem in this area is selecting the minimum set of all inclusive and mutually exclusive potential responses. This is often a non-trivial issue. Multiple responses can be used, but greatly increase the difficulty of analysis.

Content coding of open-ended questions can also be used but is not recommended. This is usually a difficult and time consuming process and often results in high levels of error. This approach is particularly problematic with written (mail) questionnaires where few respondents answer open-ended questions. Using telephone surveys, additional error is introduced using coded open-ended questions as the interviewer usually interprets the responses.

3.2.2 ATTRIBUTE RATINGS: (ORDINAL OR INTERVAL SCALED)

Most of the critical data for attribute evaluation is collected as ratings. Typically these include the performance of the target product and its competitors on a set of attributes. Recently, expectations are also included in these scales[4]. Though not recommended, importance scales are also typically collected in this manner.

---

[4] Total Quality Management methodologies sometimes call for customer expectation values as well as competitive performance.

3.2.2.1 Anchoring

The key problem in developing rating questions is anchoring the scales. Anchoring requires identifying the meaning of the scaled numbers. Under minimum conditions the extreme values are identified. It is usually advisable to also set the middle values if feasible[5]. It is critical, however, to anchor the "center" of the scale. This may be either at the logical center of the scale, such as 5 on a 1 to 9 scale, or at some reasonable position, such as a "satisfactory or expected" point of 7.

3.2.2.2 Bi-directional Axes (Semantic Differentials)

Bi-directional axes consist of scales with unique and specific extreme anchor labels. This type of scale is also referred to as a semantic differential. The major difficulty is in definition. The two extreme conditions must be opposites. In many cases, this is not obvious to all respondents.

3.2.2.3 Uni-Directional Axes

Uni-directional axes have common extreme values and are commonly used for attribute evaluation. These are the standard 1 to 10 scales. However, there are a number of ways to setting up these questions.

3.2.2.3.1 Satisfaction

The satisfaction scale is the most typical unidirectional axis. It requires the respondent to indicate satisfaction from delighted to grossly dissatisfied for specific services and product attributes.

3.2.2.3.2 Agreement Scales

Agreement scales typically focus on the characteristics of the supplier and consist of a number of statements. Good design often requires multiple statements (both positive and negative) to evaluate respondent opinions on a specific issue.

3.2.2.4 Number of Points

The general rule in determining the number of points is to use the minimum necessary. The larger the number of points on a scale the more inherent variability there will be in the results. The inherent variability comes from selected respondents' unwillingness to use the extreme values. However, there are two additional issues that should be considered: (1) the need for a center point, and (2) the need for competitive comparisons.

[5] There is a significant psychometric literature on selecting word-cues that allow for the construction of interval scales. However, they are normally considered to be too lengthy to be used on written questionnaires and not effective for the telephone surveys.

3.2.2.4.1 Odd and Even Scales

Respondents that are uncertain may tend to "take the easy way out" by selecting a center value. This loses the fine value of "slightly" positive verses "slightly" negative findings. Typically, we suggest an even set of options to avoid this problem. However, in some cases, the center value is meaningful, representing a "satisfactory" value and is, therefore, very appropriate.

3.2.2.4.2 Competitive Comparison

For the evaluation of a single product, very short scales (3 or 4 points) are very effective. However, they do not allow for the comparison between two similar products that differ only slightly in performance. Under these conditions, we have tended to use a long scale of 1 to 10 and in some cases even higher.

3.2.3 IMPORTANCE: CONSTANT SUMS (RATIO SCALED)

To get an effective measure of importance some type of "trade-off" needs to be imposed. Otherwise, the respondent will merely note that nearly everything is important. This is the key problem with using rating scales for importance.

Importance measures can be collected as a "constant-sum" where the respondent is asked to distribute 100 points across the attributes based on their importance in deciding a key issue. These are often set up as "nested-sums" where the respondent distributes points across attributes in groups and across the groups. It should be noted that this can be a difficult and time consuming task. However, it is the recommended procedure for measuring stated importance.

3.2.4 MOST IMPORTANT: RANK ORDER (ORDINAL SCALED)

Alternatively, one could rank the attributes in the order of importance. This would also be a "trade-off" type response. It is often viewed as simpler than "constant-sum" evaluation, but does not provide the details for analysis[6]. There are also several problems using rankings:

- Respondents tend to confuse a ranking with rating scales; and
- Partial ranks are often generated.

This approach is used when the size of the questionnaire is viewed as excessive.

[6] Quantitative values are obtained either collective or individually by imposing a rank order statistical distribution, such as the "broken-stick rule."

3.3 DATASETS

Data for analysis is supplied electronically in a number of formats[7]. While we have preferred the data in an *EXCEL* format or one of the standard statistical or database formats, due to legacy systems, the typical data is provided as an ASCII text file. It should be noted that data transfer and preparation take a sizable amount of time and effort before analysis can start.

3.3.1 RECOMMENDED FORMS

There are three issues that we have found problematic: (1) multi-card format, (2) multipunch and (3) missing first column entries[8]. The legacy systems are based on the archaic 80-column card data entry. As such, data is often available as a number of cards or records per respondent. Most modern systems deliver the data as a single record. At present, we recommend a single record per respondent if feasible.

Old card based systems store data as a multipunched card. This allows for full use of the 12 punching positions (used for multiple selection questions). Unfortunately, ASCII can not interpret all of the (4096) combinations. It is therefore, important that the data be recoded for standard ASCII.

There can be problems with data that have inconsistent missing first column entries. Many of the tools that used to parse the file will miss the first column on this condition and incorrectly code the record.

3.3.2 CLEANING DATA

There are several problems with the data as coded which often need to be corrected before analysis. These include:

- Codings for missing data (often these are alpha-numeric codes that will conflict with computation using either *EXCEL* or one of the statistical packages.
- Comas in text fields are often interpreted as signaling new records.
- Alphanumeric codes not interpreted as numeric (&, -, +).
- Incorrect data entries or responses (i.e. Data out of range and constant sums that aren't).

---

[7] These formats include *SPSS*, *SAS*, Multipunch Text, and special formats as well as conventional database formats such as *Microsoft ACCESS* and *DBase* and Spreadsheets such as *Microsoft EXCEL*.

[8] We typically parse the data (split up the data into variables) by using *QuickBASIC* or within *EXCEL*. As such, there are a number of problems that an inconsistent first column entry produces.

These need to be corrected during the data transfer operation.

3.3.3 ARCHIVING

If it is desired to revisit the data it is important that it is archived in a form that can be used. This requires that the original data be saved along with electronic forms of the questionnaire and the coding sheets. Alternatively, *EXCEL* workbooks with the data can be used, but must be fully documented[9].

[9] Unfortunately, this is rarely done.

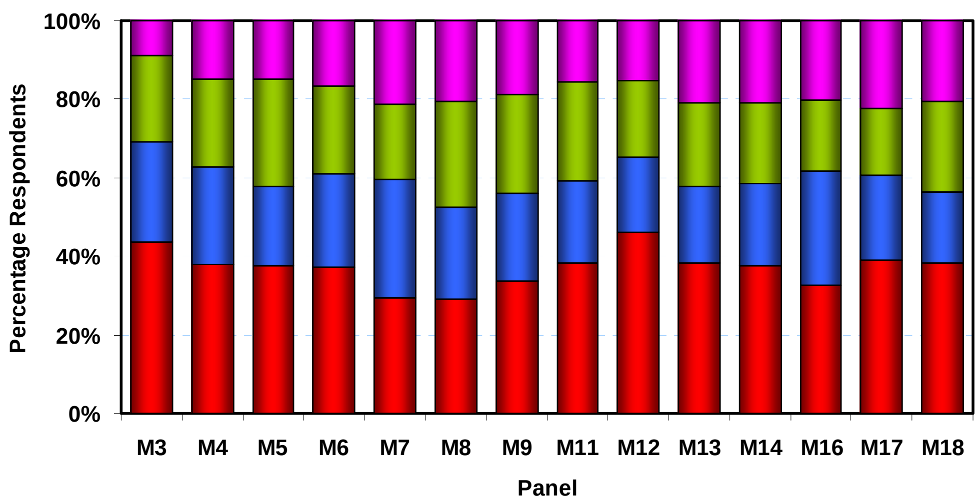
3.4 SAMPLE QUALITY

All survey studies assume that the sampled respondents represent the targeted population. The sample is assumed to be stable, consistent and represent the population from which it is drawn. For consumer, political and social research we have also assumed that the sampled respondents represent the general census population. This is referred to as the probabilistic sample assumption. However, such samplings are rarely the case. Telephone surveys, for example, have been found to be increasingly biased due to refusals and the incomplete phone listings (mainly due to cell phone use). Online panels have become increasingly used both for consumer research and for specialty surveys (business to business). These are clearly not probabilistic samples.

The present issue is can the quality of sources be controlled. They can and should be at least monitored and audited. Auditing can be done either internally in the executed questionnaire or as an independent exercise. Typically internal audits take up valuable questionnaire space and time and are generally not used. Independent audits should be provided by the source vendor or supplier covering key issues.

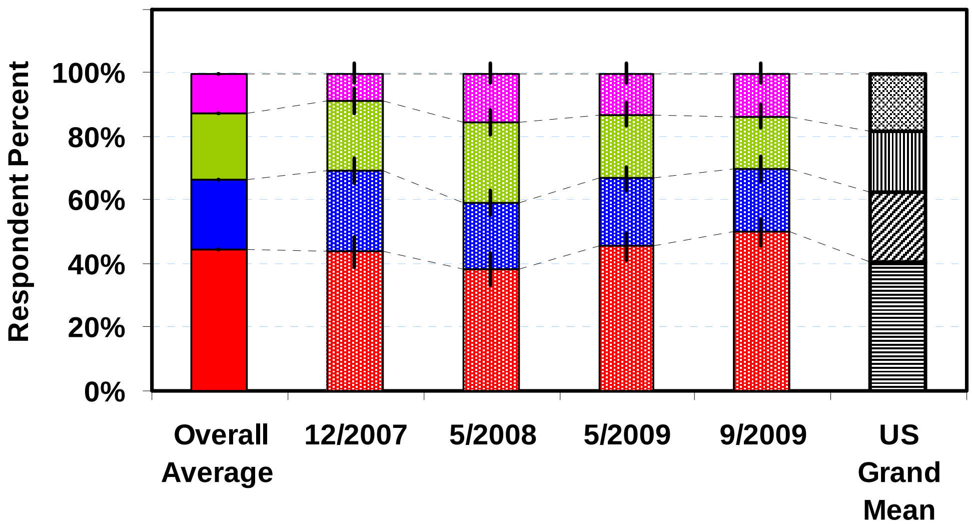
3.4.1 COMPARING QUALITY

The quality of the samples can be tested using a collective metric (over a large set of variables). In the chart below shows the variation of a collective metric[10] across a number of online consumer panels within a country using similar screen criteria.



This chart indicates significant variation between panels. However, there is also variation over time. This can be particularly a problem with respect to tracking studies where period to period consistency of the sample is assumed. Studies have indicated that there is some significant variation as noted below.

[10] This is a media use segmentation representing the frequency of various groups of consumers based on the responses to a set of behavior and attitude questions.



The above discussion of quality focused on the variation of the sample's population. In addition, there are issues regarding the appropriateness of the responses in the surveys.

3.4.2 FREQUENCY OF "ERRORS"

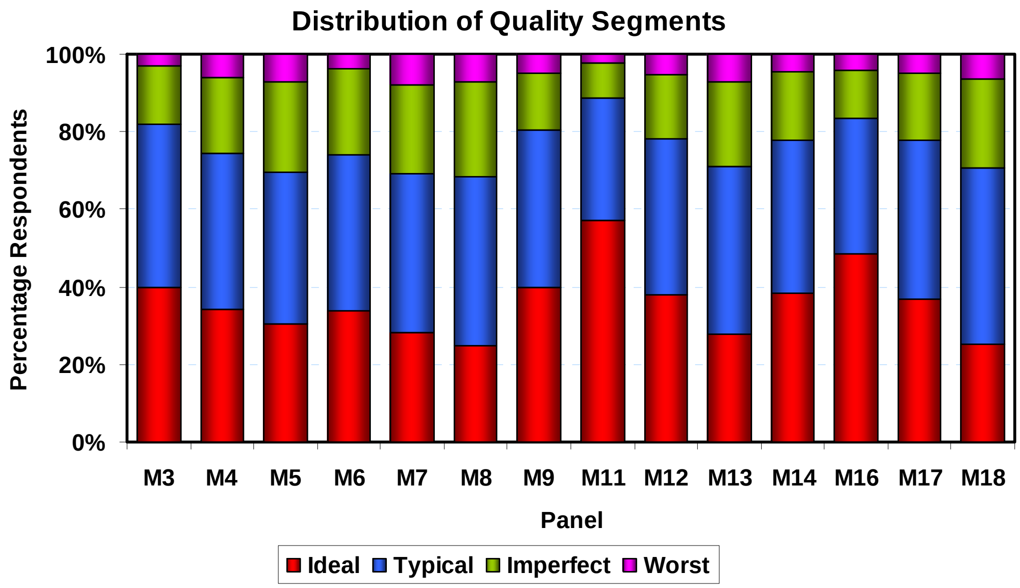
There are several types of panel issues that reflect the potential for respondents to generate erroneous responses. It is in the incidence of multiple errors that provides insight into the potential that these are systematic and reflect problem respondents. This is explored with the quality segments based on the number of errors that each respondent does. There are three types of potential errors and problems that are explored:

- The ***incidence of actual errors*** in the execution of a questionnaire reflects the quality of the panel. These are "checks" designed into the testing instrument. They include but are not limited by: (1) inconsistency in responding to multiple questions and (2) the failure to follow instructions

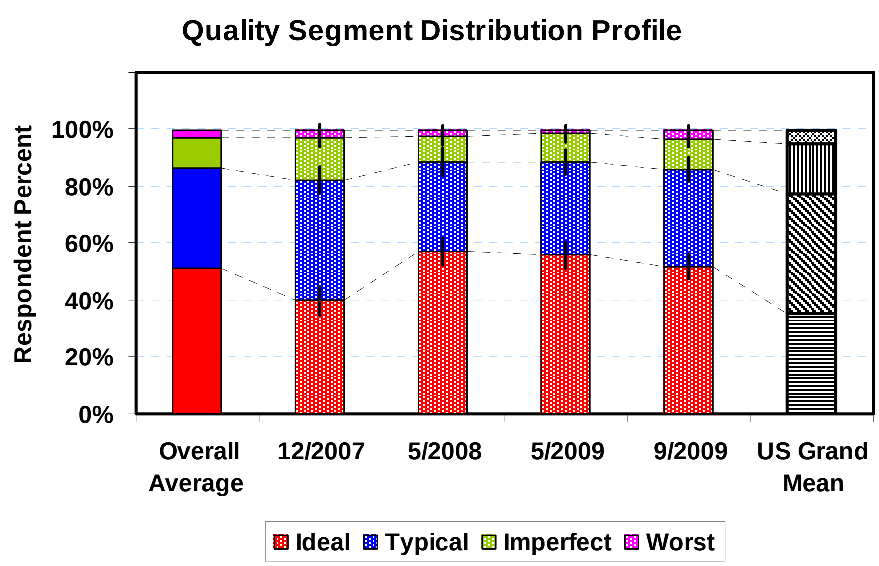
- ***Indications of "Professionalism"*** reflect the frequency of survey taking and the number of panels that the respondent belongs to. In general, these focus on issues and concerns with the long term source maintenance and in particularly the tendency of containing "professional" participants.

- ***Satisficing*** **–** Respondents occasionally show satisficing behavior. These are not errors, just extreme behavior which provides a potential warning of problems. These include: (1) "speeders" who finish their questionnaire in extraordinarily short time and (2) "straight-liners" who tend to give the same answer to a large number of questions.

The quality segments are based on the number of noted issues including errors, indicating "professional" behavior, or showing satisficing behavior. There can be any number of indicators, in the chart below six indicators are used: three indicating errors, one measure of professionalism, and the two measures of satisficing behavior. Four segments are used corresponding to: (1) no error (Ideal), (2) one error (Typical), (3) two errors (Imperfect), and (4) three or more errors (Worst). In this context, it is the Worst segment which is of the greatest concern since it represents those who are most likely to give erroneous responses. Below is the variation across panels in a single country.



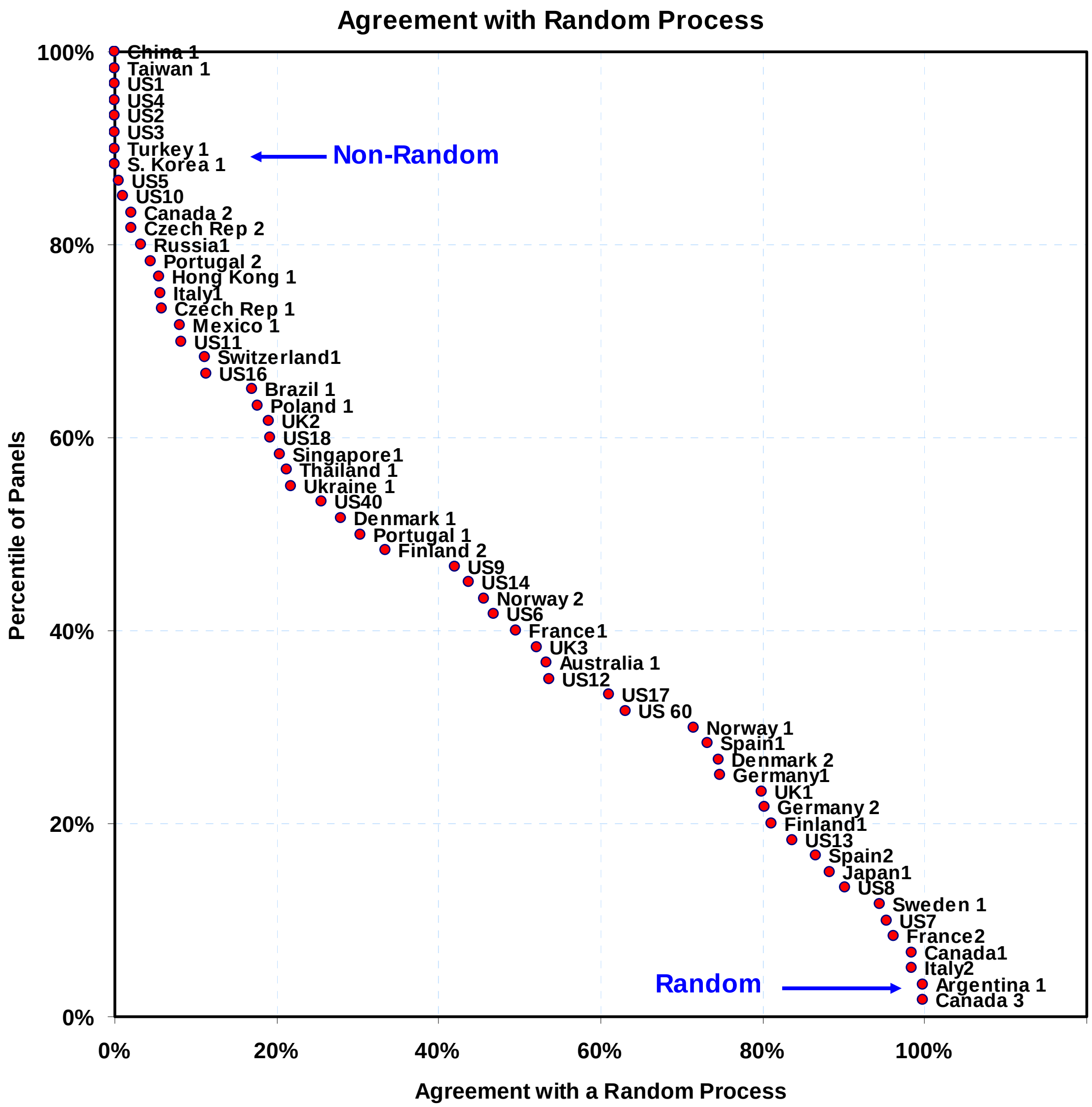
As in the case of consistency, these quality segment distributions can vary significantly over time as indicated in the following chart.



3.4.3 Faulty Respondents

As previously noted these measures are used as predictors for potential faulty behavior. Some vendors use this and other procedures to simply remove all offending respondents which often can be as high as 50% or more. However, that may also result in producing a biased sample.

Another way of looking at this is to assume that the error could be random. That is that there is an independent chance that any respondent could produce any of the errors. If it is a random process, however, that would produce a specific distribution of quality segments. With that as a reference we can compute the likelihood that any particular quality segment distribution is random. Below is the distribution of international panels indicating the agreement with a random process. Note that this covers a broad range indicating that some of these may be random while others are more likely to be systematic.



Methods of handling the issues of sample quality are still evolving. What is clear is that these issues need always to be considered as part of the error assessment of any survey.

3.5 SAMPLE SIZE

There are two general types of error that need to be considered in designing attribute evaluation studies: (1) accuracy issues including measurement and execution error and (2) precision or sampling error. Both of these assume that the data sample is consistent and representative of the targeted population. This issue has been previously covered, but it still as a major concern.

Accuracy is the more difficult and less tractable problem and consists of all inherent and operational problems associated with capturing the actual opinions of the respondents that are interpretable by the client[11]. Precision is a statistical error that reflects the uncertainty in assessing population characteristics based on samples. Here again, we are assuming that the sample is table, consistent and representative of the targeted population. This is referred to as sampling error. However, it needs to be understood that measurement, quality and execution errors usually are more severe than precision unless very small samples are used.

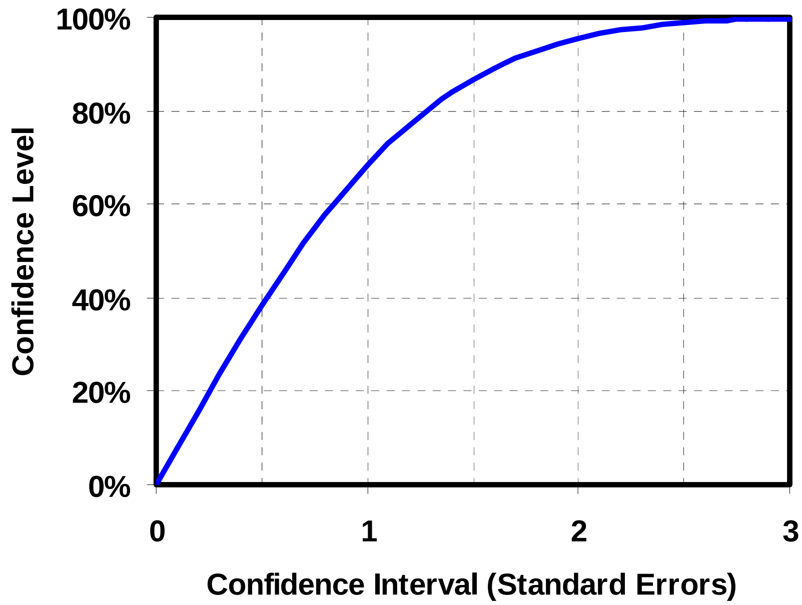
3.5.1 SAMPLING ERROR

As previously noted, sampling error is stochastic. It is an estimate of the uncertainty around a statistic. Unless there is a design problem, the larger the sample size the smaller the uncertainty between the estimate of a characteristic of the population and its actual value. Note that this tendency is independent of the quality of the sample source. It only refers to the likelihood or uncertainty that the statistic of the sample differs from that of the larger population. We select the size of the sample (or more often accept the size of the sample) based on the uncertainty that we can accept in the key parameters of interest. We, therefore, start with "guesstimates" of key variable statistics and the precision needed to make reasonable decisions. We will consider two kinds of data: (1) value data that are usually the average values of ratings and (2) percentage data such as percentage dissatisfied.

3.5.1.1 Values

We usually assume that average values of ratings are Normally (Gaussian) distributed. As such, the "Standard Error" around the mean represents variation due to sampling. For example points within one standard error contain 68% of all variation of the mean. Points within two standard errors represent approximately 95% of the variation. The chart below shows this relationship.

[11] These are covered in detail by standard marketing research texts and not covered here in general.



The issue is then computing the expected standard error around the mean. It is a function of the Standard Deviation of the data and the sample size by the following relationship.

$$\textbf{Standard Error} = \textbf{SQRT}\ [\sigma^2/N] = \sigma \bullet \textbf{SQRT}(1/N)$$

The appropriate sample size based on a 68% confidence would be:

$$N = [\sigma/\text{Acceptable Error}]^2$$

where σ is the standard deviation of the data and σ^2 is the variance. For other confidence levels multiple values of the standard deviation are substituted.

3.5.1.2 Percentages

In a similar fashion, estimates around percentages can be computed. The Standard Error around a percent is:

$$\textbf{Standard Error} = \textbf{SQRT}\ [P \bullet (1 - P)/N]$$

where **P** is the percent dissatisfied, **N** is the sample size. Once again, the sample size corresponds to an acceptable error.

$$N = P \bullet (1 - P)/\text{Acceptable Error}^2$$

This formula is based on a normal (Gaussian) distributed mean error. Unfortunately, percentages tend to be binomial distributed not Gaussian. The following *EXCEL* formula is to compute the error bounds based on the binomial distribution. The appropriate values of N can be then computed by approximation[12].

$$\text{Error Bound} = \text{BETAINV}(\text{Level},[P \bullet N],[N - (P \bullet N)+1],0,N)/N$$

Where **Level** is the probability level. It should be noted that for most cases these two methods produce similar results. As such, we normally use the Gaussian form for estimation of sample size though it is technically incorrect.

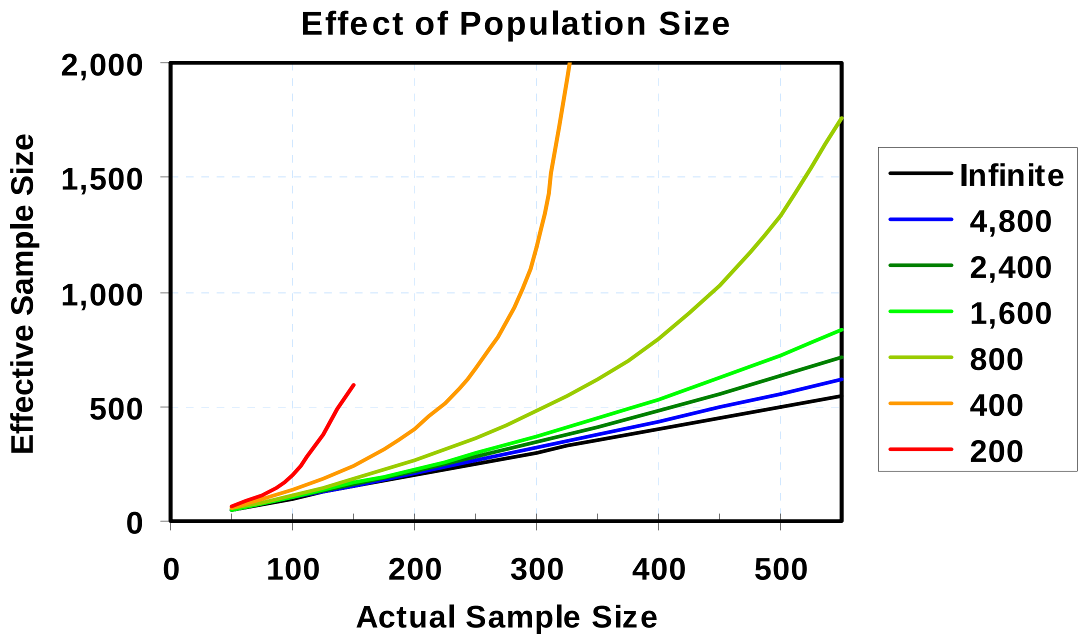
3.5.2 Small Populations

For small populations the effective sample size has to be adjusted for the total population size. The following relationship is the adjustment for total population, **n**, and the sample size, **N**.

$$\# = N \bullet (n-1)/(n-N)$$

It should be noted that as **n** becomes very large the effective sample size equals the actual sample size. If n becomes small, however, the effective size can become extremely large. The following chart shows the range of effective sample sizes given actual samples and population sizes. Note that the effective sample is always larger than the actual sample with small populations. This effect can be very large reducing the need for large samples.

[12] This formula uses the beta distribution that is a continuous version of the binomial and allows inversion. The solution for the sample size can be done using the SOLVER capability in *EXCEL*. SOLVER searches for the solution; in this case, it will seek a value of N to match the acceptable error.



3.5.3 Stratification and Weights

An additional problem exists when weights are to be used with the sample. This is particularly the case when specific sub-populations are sampled by quota. This makes the various groups and segments not represented statistically in the sample. Under this condition, the sample size of each group is computed separately. The statistics for the total sample are then computed using weights that correspond to the proportionality in the population. However, there is a change in the effective sample and the corresponding precision of the average results and other sample statistics.

Sample stratification is done to increase the precision of the sub-population estimates. The use of weights is used then to improve the accuracy of the total sample statistics and remove the sampling bias introduced by stratification. Weighting these statistics increase the accuracy but at the expense of precision. The choice of using weights is therefore one of balancing the potential improvement in accuracy against the loss in precision.

For example, consider a sample of 1000 respondents, with only one individual being important, all others have weights of zero. In this extreme case, the actual sample size, for statistical purposes, is only one not 1000. On the other hand, if we equally weighed the data, the effective number would once again be 1000. The effective sample size can be computed as:

$$\eta = 1/\sum_{i=1}^{N} W_i^2$$

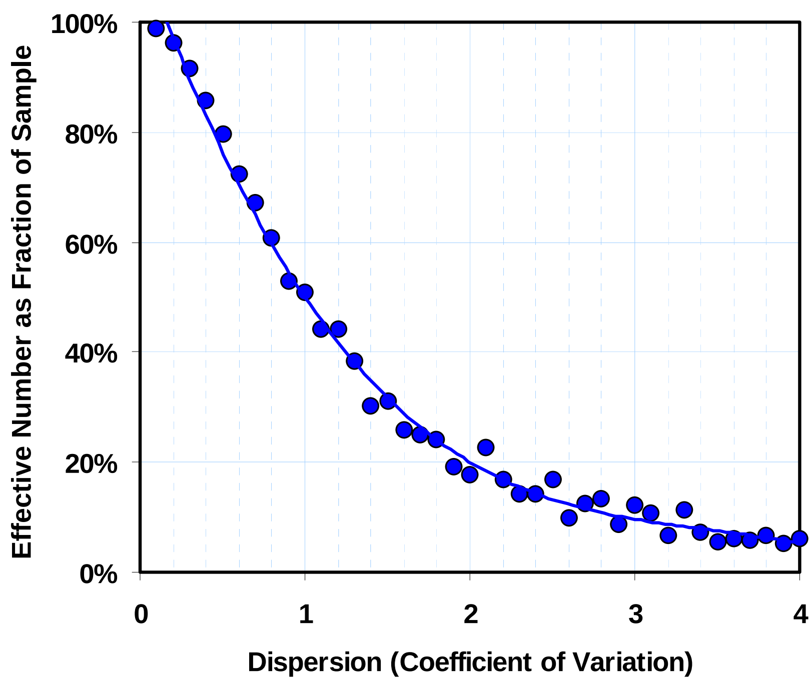
Where η is the effective sample size, **N** is the actual sample size, $\mathbf{W_i}$ is the normalized weight for respondent **i**. Another way of looking at this is the greater the variation in weights the lower the effective sample size. In terms of general weights this expression becomes:

$$\eta = \left[\sum_{i=1}^{N} \lambda_i\right]^2 / \sum_{i=1}^{N} \lambda_i^2$$

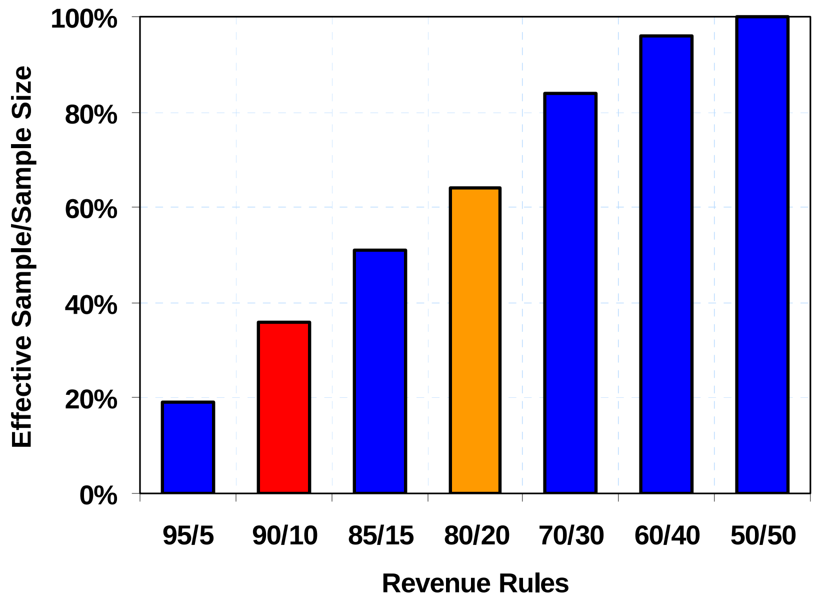
Where λ_i is the weight applied to item **i**. The numerator of this expression is the square of the sum of the weights. This is approximately equal to:

$$\eta = N/[\nu^2 + 1]$$

Where ν is the coefficient of variation (standard deviation/average) of the weights λ_i. With no variation in the weights, the effective sample size is equal to the actual, while at an infinite variation the sample size goes to zero. The following graph shows the results of the variance of weights on the precision of results based on simulations using normally distributed weights.



Below are the results for the effective sample size for a number of commonly used market importance ratios. An 80/20 rule, for example, means that 80% of the sales come from 20% of the customers. This is a fairly common rule for general businesses. However, industrial suppliers often see 90/10 and even 95/5 rules where as much as 95% of the earnings come from only 5% of the customers.

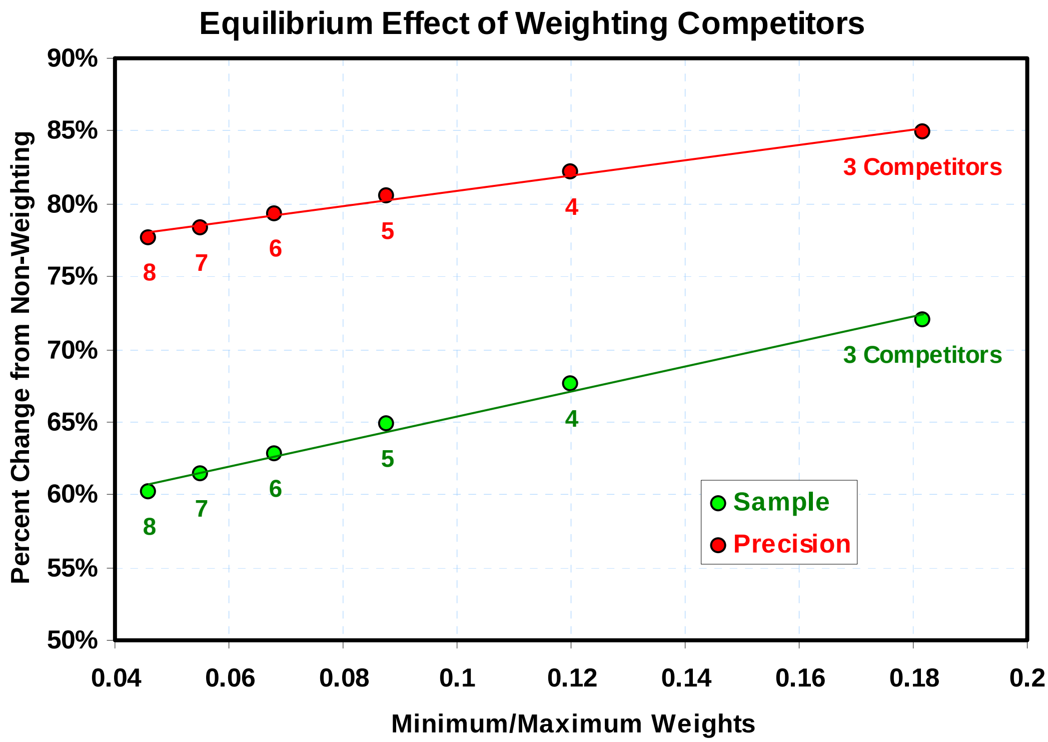


For an 80/20, the effective sample size is slightly over 60% of the original sample. For a 90/10 rule, it is less than 40%. Because of this effect, we have tended to resist using weights particularly when multivariate analysis will be used.

Typically stratification is used to capture a prior condition like product ownership. That is sampling is done to provide equal samples of groups of existing customers. The impact of this is shown below in terms of percentage sample (effective sample size) and resulting precision assuming a standard random size distribution[13].

---

[13] The "Broken Stick Rule" is used here to compute the shares given the number of competitors and the product position. The Broke Stick Rule is a rank order ergotic distribution which is discussed in the Forecasting methods section of these notes.



The horizontal axis on this chart represents the ratio of the smallest to the largest weights and measures the variation in the weights. The higher this ratio is the more similar the weights and the higher the equivalent sample size. Note that as the number of competitors rise, the more severe is the impact of the weights. The precision or the expected error goes with the square of the sample size. As such, it should be noted that the total impact on precision is expected to be modest even though there may be major reduction in the effective sample size.

3.5.4 BONUS QUALIFICATION

One of the more recent applications for customer satisfaction studies is determine bonus qualifications of employees with direct contact with customers. Qualifications are usually set based on maintaining a given, threshold level of customer satisfaction. That is, the percent of dissatisfied customers should not exceed a given level, or threshold. A satisfied customer is defined based on his response being over some criteria on a scale of overall satisfaction. The problem arises from the nature of sampling. If all customers were sampled the calculation would be straightforward. However, with random sampling, there is a probability that the true value will be below the threshold while the sample is above it.

It should be noted, that with sampled data, only an estimate can be obtained. The question is at what level of confidence is sufficient for refusing a bonus. The computation of qualification is based on the confidence interval around a percent as discussed above but with a twist to adjust for small populations. As previously noted, we can assume that the percent values are binomial distributed. For *EXCEL* the formula for the lower bound of the confidence interval is:

Lower Bound = BETAINV(Level,[P • #],[# - (P • #)+1],0,#)/#

where **#** is the effective sample size. For small populations the effective sample size has to be adjusted for the total population size. The following relationship is then adjusted for total population, **n**, and the sample size, **N** as noted above [**N •(n-1)/(n-N)].**

It should be noted that as **n** becomes very large the effective sample size equals the actual sample size.

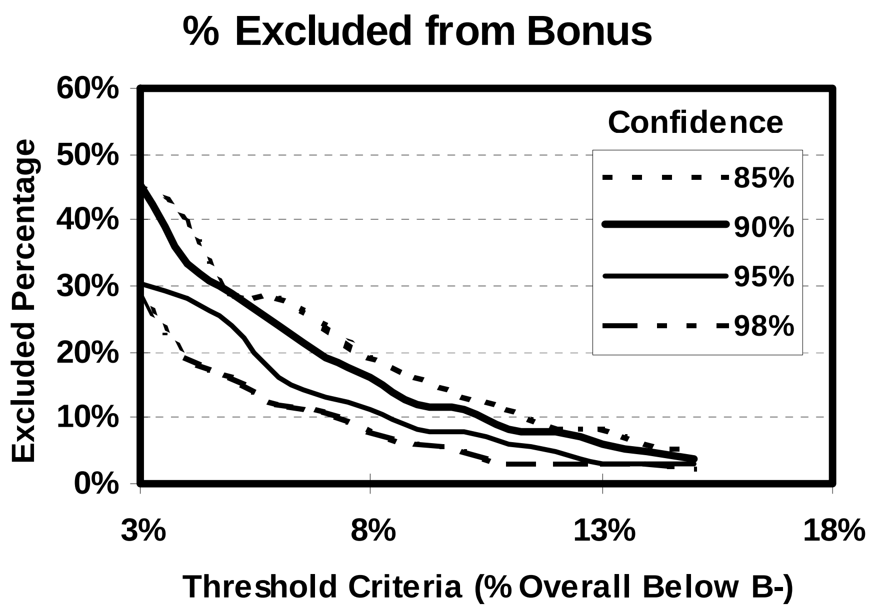
3.5.4.1 Exclusions

The determination of qualification for the bonus is, therefore, dependent on:

- The criteria to define dissatisfaction,
- The fraction of dissatisfied relevant respondents,
- The sample size of relevant respondents,
- The number of potential relevant respondents, and
- The level of confidence desired to assume that fraction of dissatisfied respondents' estimates exceeds the threshold.

The sample sizes and the population universes are set by the customer satisfaction study. As such, the percentage of excluded bonus candidates is a function of the distribution of dissatisfaction rates, the threshold and the confidence level. Typical results are shown on the chart below. Levels of the threshold and confidence levels can be selected to yield a desired exclusion rate[14].

[14] It is often desired to have some respondents excluded from the bonus to encourage improvement and to maintain the integrity of the bonus system. However, it is also desirable that the exclusions should be minimal to maintain morale in the organization. Rules for partial bonuses can also be developed in a similar but more complex fashion.



3.5.4.2 Simulation

To explore the full flexibility of selecting the appropriate bonus policy a simulator can be constructed. This allows varying, the criteria, the threshold, and confidence levels to determine both the exclusion rate and the individual assignments of bonuses. The *EXCEL* user interface is shown below[15].

[15] This model is based on several thousand respondents evaluating a couple of hundred employees.

Bonus Exclusion Model

Minimum Performance: A, A-, B+, B, B-, C+, C

Percent Excluded
11.4%

Confidence Interval: 98%, 95%, 90%, 85%

Threshold 10.0%

RESULTS		
PCP	Below Target	No Bonus
"15R098"	3.7%	0
"MHSE72	3.8%	0
"68YE47"	37.0%	1
"783F25"	11.1%	0
"70NA46"	0.0%	0
"014E36"	14.8%	0
"69AE49"	14.8%	0
"MHSE79	3.7%	0
"032451"	7.4%	0
"79GF08"	7.4%	0
"FFC141"	7.4%	0
"18T383"	7.4%	0
"57PE98"	22.2%	1
"053270"	14.8%	0
"57PE99"	3.7%	0
"783F26"	3.7%	0
"177323"	7.4%	0
"26K289"	7.4%	0
"70N675"	0.0%	0
"66EC97"	7.7%	0
"60G456"	3.7%	0
"FFCA57"	33.3%	1
"72WA34'	22.2%	1
"093541"	3.7%	0
"783F64"	7.4%	0
"MD1023'	7.7%	0
"MCD124	0.0%	0
"FFCE27'	7.4%	0
"MHSE78	0.0%	0
"NBMF69	3.8%	0

3.6 DATA DISPLAYS

The purpose of the data presentation is to drive decision making. Data is analyzed and presented in various ways to provide insight. A classic way of looking at data analysis and presentation is as a linear process by which data is reduced to more meaningful and targeted information and intelligence which is then used for decision making and finally action.

Data ⇒ Information ⇒ Intelligence ⇒ Decisions ⇒ Actions

However, this is far simpler to say than to do. There are a number of conditions that makes the information salient and useful. The goal is to make the information engaging in the creative planning process and compelling for action. At the same time, it is critical that the intelligence properly conveys the uncertainty in the data.

3.6.1 PRINCIPLES OF PRESENTATION[16]

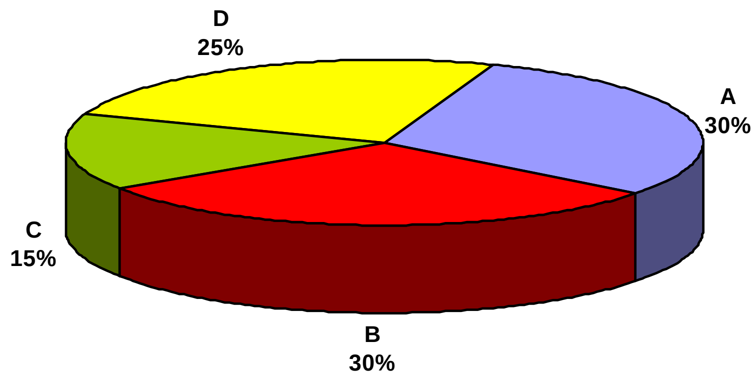
The function of charts and data presentations is to work with the decision makers. They are psychological tools. Information overload is the problem. Too much information can be as much as or more of a problem than insufficient information. Information overload results in overly simplistic and often erroneous approaches to decision making. This sets *the first principle of presentation that is the requirement to simplify the situation and focus only on those issues and factors that must be considered***.**

Data presentation must take a "Gestalt" approach in that the whole is greater than its parts. The presentation itself tells a story. How charts are constructed implies types of solutions. Their construction is based on a set of assumptions and principles that influence the way the information is viewed. *The second principle is that of the need for multiple views of the information.*

3.6.1.1 Graphic Design[16]

Not all graphical presentations give fair and reliable representations. Therefore choosing how data is presented can be critical. For example, three dimensional (perspective) pie charts are notorious for providing false perceptions of values. This is due both to the reduced visual acuity to area measures and the distortion due to the perspective as shown below. Note that segments A and B are of the same value though segment B appears larger. This is due to the rotation of the chart.

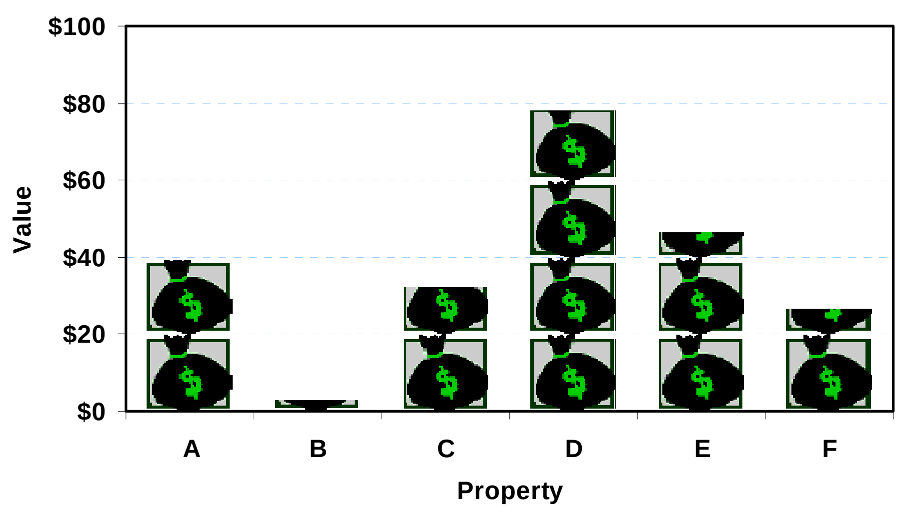
[16] Leland Wilkinson, "*Cognitive Science and Graphic Design*", Systat 12 Graphics, Systat Software, Inc. (2007)



As indicated above psychometric visual studies have indicated that there is a hierarchy in the reliability of measures:

1. Position along common scales
2. Position along not aligned scales
3. Length
4. Angle
5. Area
6. Volume
7. Color/Hue/Saturation
8. Image

Notice that the simple linear measures are the most reliable (bar charts for example). Area measures are far less reliable (pie charts), color and images are even worse. Furthermore, complex graphs can be come increasingly distractive. The use of images (dollar signs in the case below) tends to obscure the relationship among points.



Also note that all people do not perceive color in the same manner. Typically different saturations or markings are used with colors to avoid this problem. Based on these graphic restrictions the most critical information should be provided in the most direct and simple fashion. *The third principle of presentation is need for graphical simplicity.*

Keep the charts simple. The more information and the complex the presentation is the more difficult it will be to understand it.

3.6.1.2 Dimensions and Coordinates[16]

There is a long standing principle that a decision maker can only handle 5 ± 3 concepts simultaneously. Experience has tended to favor the lower limit of two concepts[17]. From a data presentation perspective, these concepts represent dimensions or categories of the data that must be viewed together to draw conclusions and make comparisons. As the number of dimensions and categories increase, the ability to draw consistent conclusions decreases. As such, *the fourth principle is to keep the number of dimensions and comparisons to a minimum*.

The linear coordinate system is the simplest and most universal one. Even “CEO’s” understand linear dimensions. They are the “standard” way of seeing things. Unfortunately, some phenomena are not readily presented in linear form. Things that grow exponentially (constant percentage growth) are not easily viewed on linear graphs. Furthermore, things that change over great ranges are likewise ill-suited for linear presentations. Logarithmic coordinates are useful for these cases. Furthermore, other coordinates[18] are useful when the underlying mechanisms are more complex. The problem using these coordinates is that the users may not be able to understand them. *The fifth principle is to provide the coordinates in a form that the user understands.*

3.6.1.3 Data Variation

Marketing rests on the variation in data. While grand average values are useful, they rarely represent the totality of the situation nor likely to provide the insight necessary for creative decision making. In almost all cases, it is necessary to look at groups of customers and the variation in the data. Furthermore, average values are often deceptive measures of central tendency. The term “typical” value is used to capture the “true” center of the data. This may be the median or the mean or some other measure of the center. The major problem is showing the variation in the data. Typically a standard error can be used. But even here it may not show the opportunity than viewing the distribution of data. Note, however, that showing data variation greatly increases the complexity of the presentation. Often different charts are used to this purpose other than for capturing product comparisons.

3.6.1.4 Presentation Dynamics

Presentation dynamics consists of user controlled options within a decision support

---

[17] When this problem was presented to several of my classes, the general opinion was that for senior management, they can only handle one single concept at a time.

[18] These include double logarithmic scales (log-log) for power law functions and normal (Gaussian) scales for normal data distributions.

system. These are options for the charts and graphs that provide for changes in the mode of presentation, the existence or absence of objects and information on the charts, or changes in the basis for computing the charts. Spreadsheet programs such as *Microsoft Excel* allow for a broad range of these options to be built into charts. These options allow users to view increasing complexity in the data presentations. These options allow for both a simplified information presentation and provide the in-depth intelligence required to explore creative possibilities.

3.6.1.5 Planning Paradigms

In order to create effective marketing intelligence, charts need to be simple and on target. These charts are constructed based on a set of assumptions regarding how markets and businesses should behave. For example, we might use average performance values to reflect the product positions. This is based on the assumption that the average values properly measure the overall market impact of the attributes. Different tools, doctrines or approaches can be used. Each doctrine has its own set of assumptions of what is important, and how they need to be presented. It is through these doctrines that market intelligence is applied to decision making. While there is no universally powerful market doctrine that solves all problems, they can all be useful depending on the situation[19].

[19] The various doctrines and their tools are discussed in the Strategic Planning Section of these notes.

3.6.2 EXAMPLES OF DATA CHARTS

There is any number of ways attribute data can be presented. Below are standard approaches used with quantitative market intelligence data[20]. These are viewed as displays as opposed to maps which reflect an underlying structure or assumptions. Competitive maps are discussed in the next section.

3.6.2.1 Summary Data Table

The Summary Table or Matrix displays the average values of the attributes across the key distribution variable, in this case products. This type of chart is used as an overall view of the data. It is usually considered too detailed to highlight relationships. In this case variation around the grand mean is also shown in colors with red being one standard error below and green as one standard error above the grand mean. Also included are measures of importance.

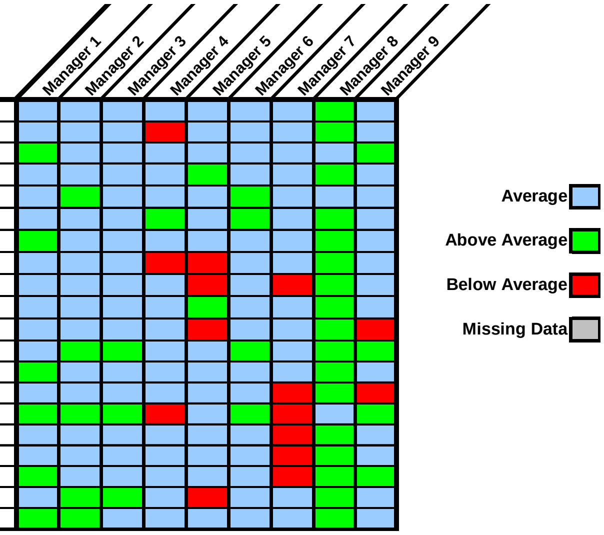
		Brand	Product A	Product B	Product C	Product D	Concept X	Concept Y	Combined Importance is the average of Stated and Derived		
	Describer	Characteristic							Stated Importance	Derived Importance	Combined Importance
01	Effect	Effectiveness	7.2	6.6	4.7	7.9	6.0	7.0	11.2%	4.4%	7.8%
02	Waste	Waste Reduction	8.7	8.0	6.7	9.2	7.0	8.0	7.8%	4.2%	6.0%
03	Yield	Production Yield	5.7	5.6	5.5	6.3	6.8	8.0	3.6%	2.1%	2.8%
04	Quality	Consistency (Quality)	7.5	6.7	5.8	8.1	7.0	8.0	6.8%	4.4%	5.6%
05	Strength	Mechanical Strength	2.9	8.2	9.5	7.9	5.0	6.0	8.8%	0.0%	4.4%
06	Temp	Temperature Stability	5.2	7.2	6.9	5.9	8.0	6.0	5.8%	1.9%	3.8%
07	Run	Runability	8.0	6.1	6.4	2.8	8.0	7.0	7.7%	1.1%	4.4%
08	Env	Environmental Impact	2.5	6.1	5.4	6.9	8.5	8.0	5.0%	0.5%	2.7%
09	Safe	Safe to Use	7.5	8.9	8.4	9.2	8.5	8.0	11.4%	2.8%	7.1%
10	Design	Design	5.2	9.3	9.2	8.6	6.0	7.0	2.8%	1.4%	2.1%
11	Easy	Ease of Use	8.0	8.5	8.3	8.9	6.0	8.6	3.5%	2.8%	3.1%
12	Recommended	Recommended	6.5	7.9	5.5	8.7	8.5	7.0	1.3%	4.9%	3.1%
13	Support	Support	6.6	8.6	6.5	8.0	8.5	7.0	3.4%	3.3%	3.4%
14	Stands	Stands	7.4	8.6	5.7	8.3	8.5	7.0	6.2%	3.7%	5.0%
15	ommunications	Communications	6.0	7.8	4.5	7.6	8.0	7.0	1.8%	3.6%	2.7%
16	Profit	Profit	6.1	6.9	6.3	6.5	8.5	7.0	4.2%	3.5%	3.8%
17	Cheap	Cheap	4.6	5.2	6.6	4.4	7.0	8.0	0.7%	0.7%	0.7%
18	Price	Price	7.6	6.5	6.7	5.8	6.0	7.0	4.8%	0.0%	2.4%
81											
82											
83											
100		Target Price	$16.00	$18.00	$13.50	$24.00	$70.00	$60.00			
101		Share	5.2%	24.3%	5.0%	16.6%	9.9%	2.9%		Data Color Codes	

3.6.2.2 Flash Displays

The flash displays are similar to the summary tables but only shows the variation of the values. In this case the color box is used to show deviation from the grand average thereby highlighting advantages and disadvantages. Note here that the key distributed variable is region controlled by managers. While the product identity is usually the key variable in competitive market analysis, regions and product groups are also used.

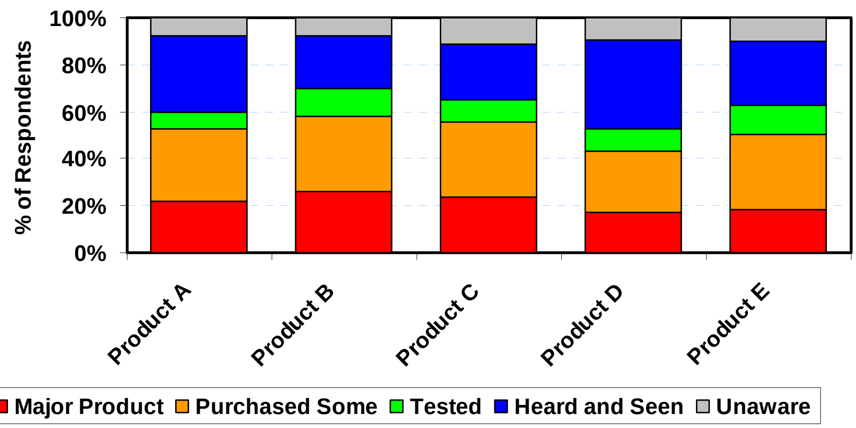
[20] These charts are used as dynamic graphs in a proprietary toolkit system (*Your-Marketing-Toolkit*)

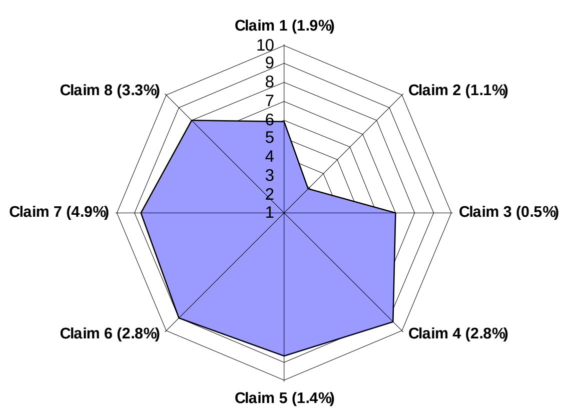
		Weight
Product Performance	Products Work With No Complaints or Resprays	4.7%
Rebates	Rebates	2.9%
SRComplaints	Sales Rep - Grower Complaints Easy to Contact, Comes Quickly, Han	2.8%
Inventory Protection	Inventory Protection	2.4%
Innovative	Develops Innovative New Products	2.3%
Product Available	Product Availability	2.8%
SRKnowledge	Sales Rep Aware Whats Going On and Gives Lot of Support/Technica	5.4%
ProfitProd	*Profitable Products (retail margin to grower)*	8.5%
Good Value	*Good Value to Growers*	4.4%
Stands Behind	*Stands Behind Products*	5.8%
Programs Simple	*Simple Retailer Programs*	4.9%
ThroughRetail	*Works Through Retailers; Not Around Them (Driver in Diehl, Wenning*	5.3%
SRBuilds Business	*Sales Rep Helps Build Business (Driver in Wenning, Som, Morris Are*	4.6%
Retail Mgt Supports	*Retail Management Encourages Support of Company Products*	3.5%
SRContact	*Sales Rep Contacts As Often As Necessary*	4.4%
Product Promote	*Market and Promote Products*	3.6%
Product Best Selling	*Best Selling Products - In High Demand From Grower*	4.1%
SRCustomers	*Comfortable Taking Sales Rep to Customers*	3.2%
Broadline	*Broad Product Line*	3.3%
SRTrains Staff	*Sales Rep Provides Product, Stewardship and Selling Skills Training*	2.9%



3.6.2.3 Percentage and Average Comparisons

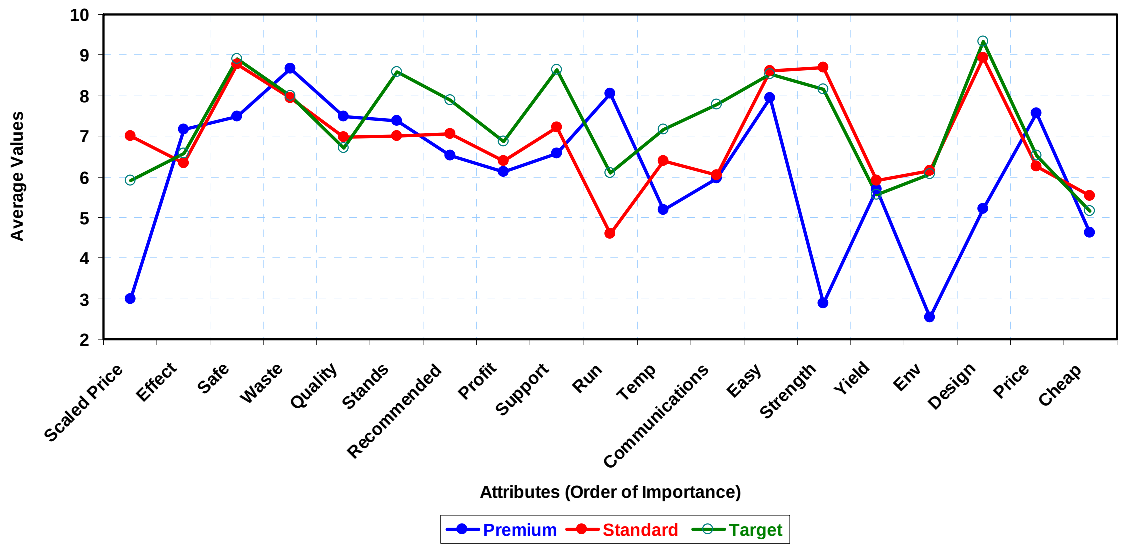
Simple charts are often used to show comparison among products. Two examples are shown below. In the first is a staked bar chart is used to show differences in the awareness of products. The second is a simple star diagram used to show product or region profiles in this case of promotional claims.





3.6.2.4 Competitive Profiles

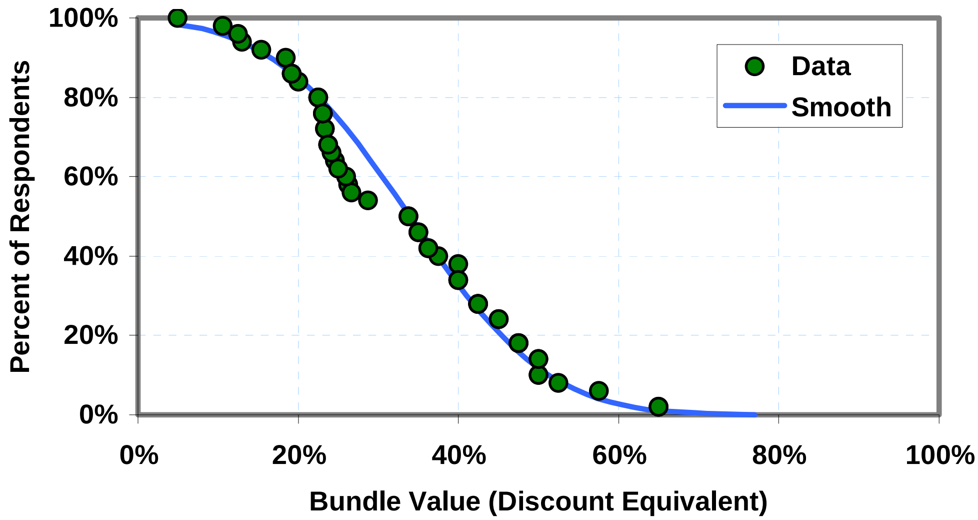
There are a number of graphical methods of displaying the competitive profiles of products. Below is a "Strategic Canvas[21]" approach showing the profiles of groups of products. The objective of this approach is to highlight areas of opportunities for developing competitive advantage.



3.6.2.5 Distribution Display

The above charts have all dealt mainly with average values or "measures of central tendency". The distribution of the data has been captured by noting the standard deviation around these means. However, as previously noted, effective marketing focuses on the variation. The use of the standard deviation assumes some type of smooth symmetric or normal distribution of the data. Values are often clustered or grouped together and the distributions may be asymmetric. Furthermore, one usually does not market to the median or the mean. Often we are more interested in the high and low values. The distribution of values can be displayed as a scatter diagram as shown below. As shown, it is often useful to indicate the "smoothed" curve. This is used to capture of overall relationships in the data and to reduce the effect of the "noise".

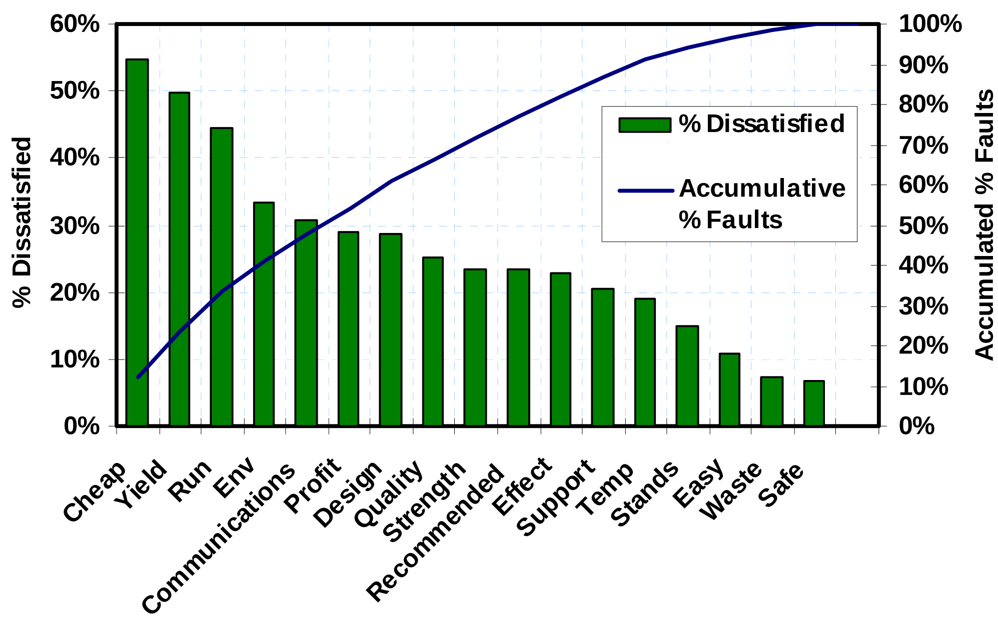
[21] This type of presentation has been proposed by the "*Blue Ocean Strategy*" approach to identify areas of low competitive action and potential opportunities for unhindered growth.



Smoothed Normal Distribution R-Square = 93%

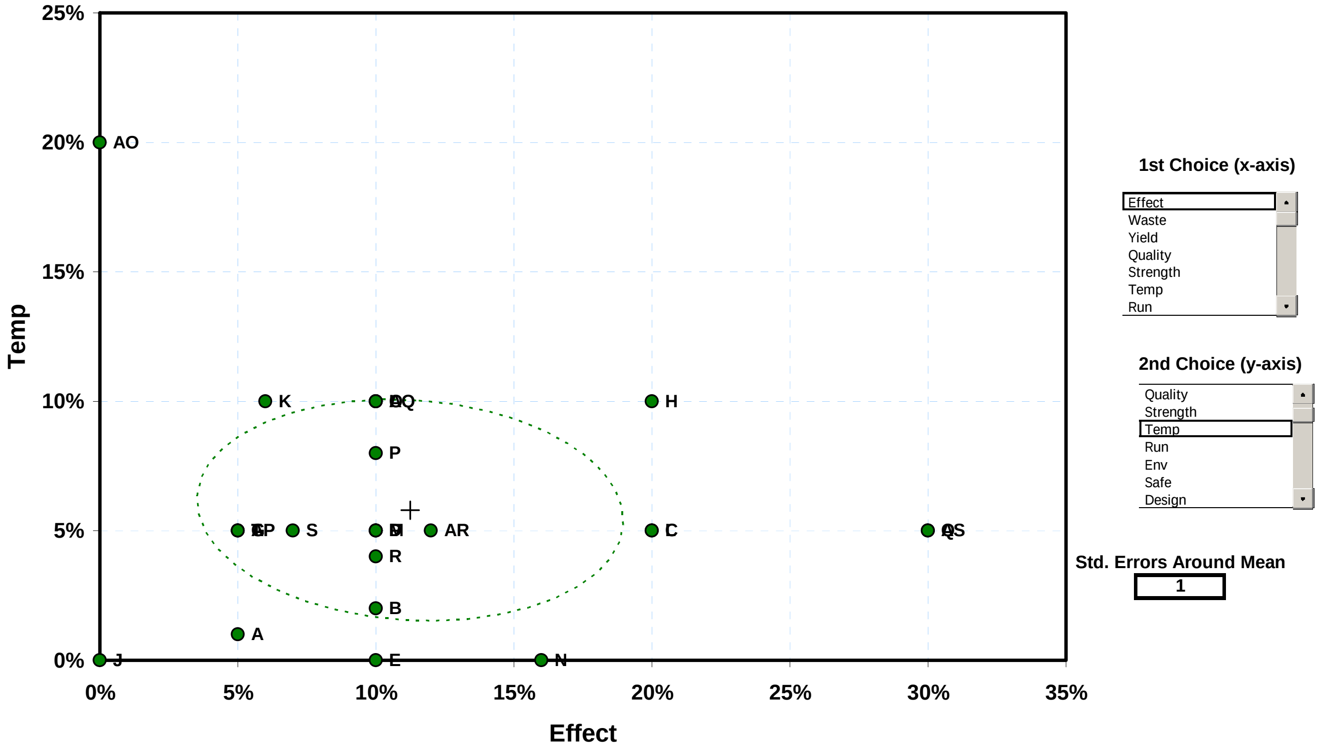
3.6.2.6 Pareto Charts

The Pareto Chart is a standard tool of Quality and "Six Sigma" management. Here the responses to questions are converted to a measure of faults, where a fault is defined as an incidence of values below some failure point. Since the failure points can be varied, this type of graph reflects the underlying data distributions.



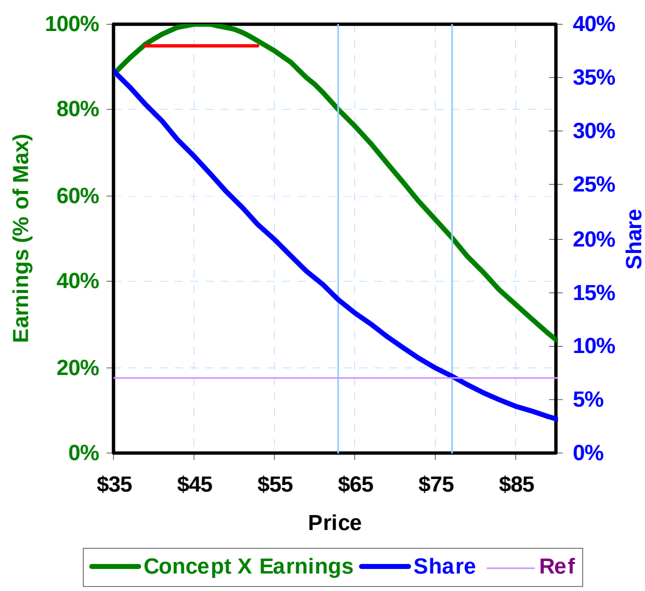
3.6.2.7 Position Displays

It is often useful to actually see the association of data points. Position displays show the actual data, typically on two dimensions as shown below. It is also useful to identify the points either as groups or as the identity of the respondents. The oval on the graph indicates the variation around the average. The average value is indicated by a "+" on the chart. This type of charts can be used for sales activity planning.



3.6.2.8 Model Charts

The last type of display is the result of an underlying market model. These are typically line charts showing the results of model parameter choices. The chart below shows the results of a pricing model based on choice exercise respondent data.



3.6.3 COMPETITIVE MAPPING

Competitive mapping is designed to help visualize the market situation and potential changes in that situation. It is a multivariate visualization process. As such, it is usually an approximation where we wish to see the "forest from the trees."

Competitive mapping and Quality Analysis focus on the relative position of **averaged** values. Other analyses such as segmentation, key driver analysis, and simulation focus on the variability on the respondent level. While similar tools are used, their functions are very different. Analysis of averaged data, is a visualization process only. Analysis on the respondent level is seeking understanding of underlying relationships.

3.6.3.1 The Function of Positioning

The concept of market positioning deals ultimately on what is going on in the mind of the respondent. We seek to understand how our firm and products differ in the minds of respondents from our competition. We also seek to formulate a strategy that could change that position at our advantage.

3.6.3.1.1 Perceptual Maps

The term "Perceptual Map" was developed as a means of viewing psychometric data and adopted for marketing and advertising research. In attribute evaluation, it refers to competitive performance mapping, where the average performance measures are displayed in a compressed format on two dimensions.

3.6.3.1.2 Importance Maps

Stated importance values can also be mapped in a similar fashion. However, this is an unusual analytical approach and rarely used.

3.6.3.1.3 Strategic Positions

The ultimate objective of mapping is to help develop marketing strategies that are implementable and should result in improved profitability. To do so, we are seeking to improve the "position" of the products against groups of potential customers.

3.6.3.1.3.1 Role of Segments

Groups of potential customers are viewed as "segments." These may be predetermined (ad hoc) or identified through the analysis of their preferences (benefit segments). We focus on segments, not on individual customers (unless they are very large).

3.6.3.1.3.2 Product Differentiation

Strategically, we might wish to change the product position in order to differentiate from the competition. However, we must do so in a way that makes our product favorable to its potential customers. This is viewed in terms of position against identified market segments. Product differentiation is always a key element in market strategy.

3.6.3.1.3.3 Market Segmentation

Alternatively, it might be possible to identify a new type of market segment that would look upon a modified product favorably. This is referred to as market segmentation where we target groups of customers that would favor the new product.

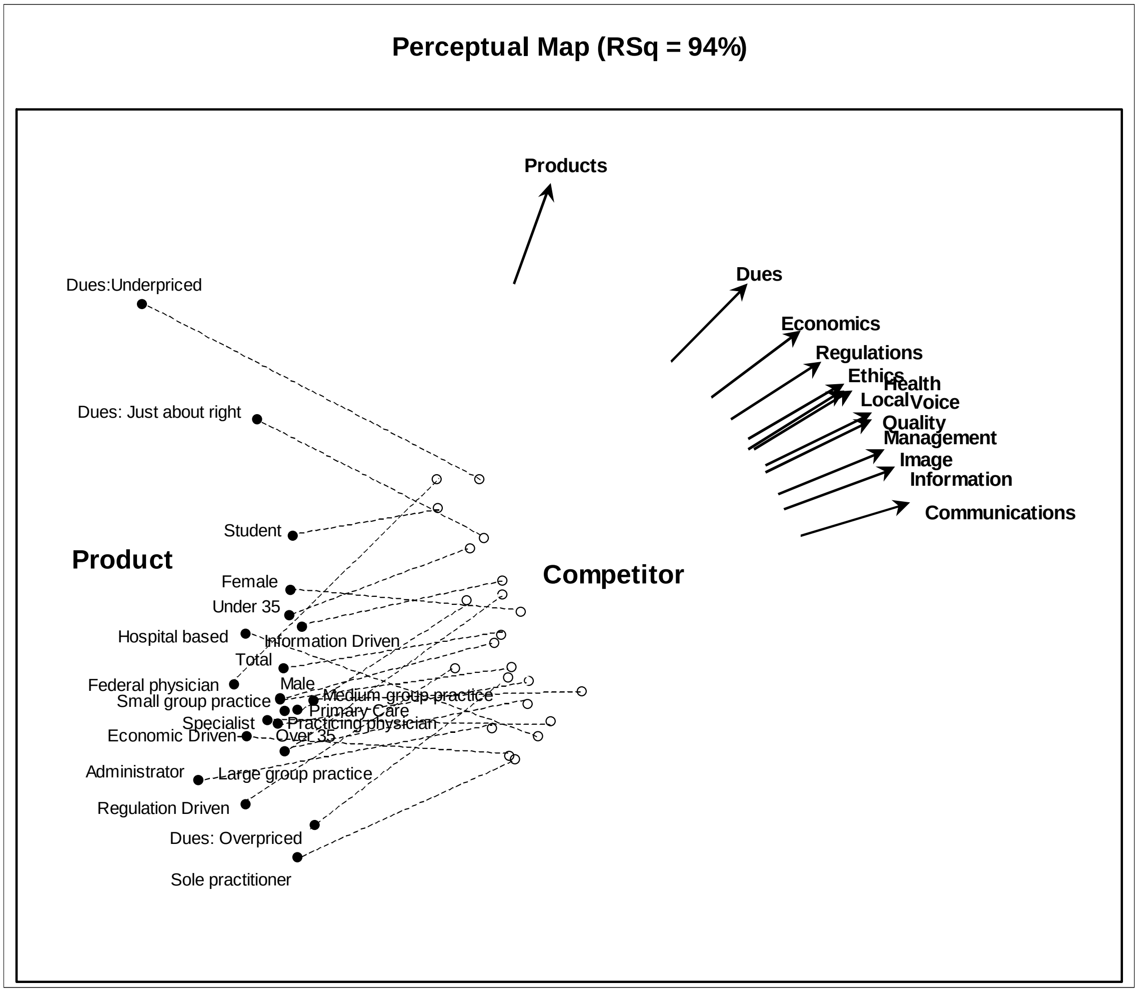
3.6.3.2 Methods of Position Mapping

The purpose of position mapping is to collapse many variables representing many dimensions into a flat (2-dimensional) picture. This will always be an approximation. Information will almost always be lost. There are two types of maps that are generated based on the relationships between points representing groups of respondents (segments) and the variables. There are two methods of computing them: (1) Factor Analysis and (2) Multiple Dimensional Scaling (MDS).

3.6.3.2.1 Position Vector Maps

Position-Vector or point-vector maps indicate the segments as points and the variables as arrows or vectors. This is the most common method of displaying position maps[22]. In the perceptual map below, groups of respondents are shown indicating their preference by properties of two organizations.

[22] Position-vector maps also require fewer critical assumptions than point-to-point maps.

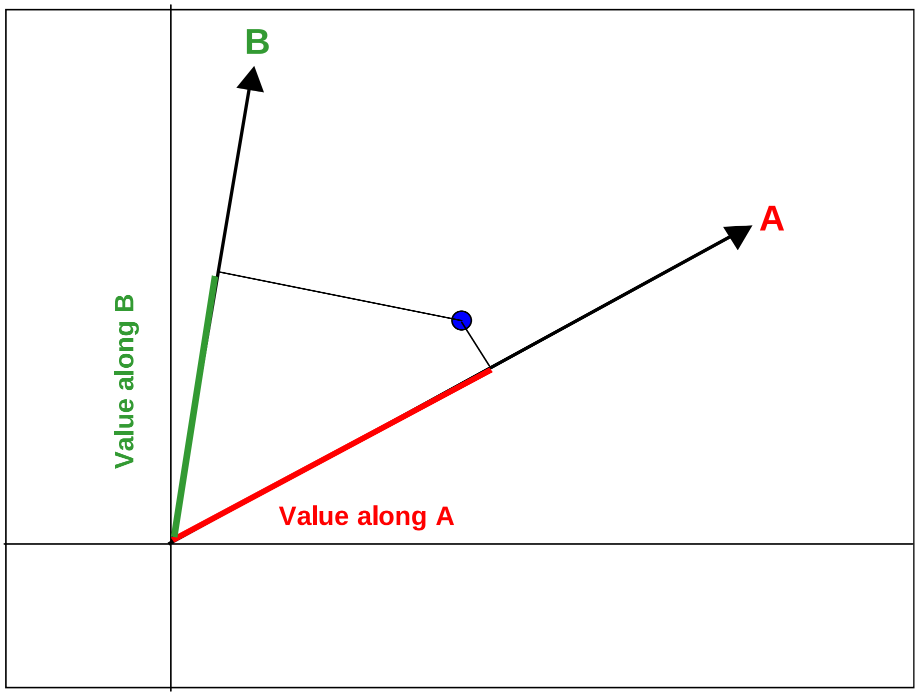


The position of the groups of respondents presents the "best" representation of the their attribute values. The vectors represent the attributes compared to the geometric horizontal and vertical positions. The length of the arrows represents the importance of the attribute. The importance may either be the impact of the variable in forming the map or alternatively, the average stated importance of the total respondents[23].

In the traditional display, the arrows are drawn to the origin of the graph (in the middle). This tends to make the map difficult to read. In the above graph the vectors have been pushed out beyond the range of the segment positions. The distance between points represents the similarity of the segments in regards to their perception of the attributes. The approximate values of the variables for each point are proportional to the length along the vector from the origin to a position perpendicular to the point. This is shown below. However, it must be recognized that this is only an approximation. The actual

[23] In standard perceptual maps the lengths of the arrows are proportional to the map "loadings." However, if this is used, it is incorrect to interpret them as "importance" measures. They only represent the agreement with the data representation.

values need to be read from the data and are available.



3.6.3.2.1.1 *Factor Analysis*

The most common method of mapping is using Factor Analysis. This method identifies a series of linear combinations of the variables in such as way that these combinations or factors are independent of each other. The basic method, referred to as principle components,[24] finds as many factors as there are either variables or cases (segments) which ever is smaller.

3.6.3.2.1.1.1 *Goodness of Fit*

If all factors are used, the resulting "factor model" captures all of the variation in the data. However, for mapping purposes only the two factors that capture the largest portion of the variance or (R-Square) is used. This may vary from less than 50% to almost 100%[25].

3.6.3.2.1.1.2 *Rotated Factors*

The factors can be rotated around the origin. This is done to make the variable vectors better coincide with the vertical and horizontal axes. There are several methods of orthogonal rotation that can be used. Usually "Varimax" rotation is used since it

---

[24] Principal components is a least squares error method of fit.

[25] If the factor model captures less than 50% of the R-Squared (less than 70% of the multiple R), it is usually not used. The variables are considered too independent for this type of mapping.

maximizes the alignment of variables and still maintains the independence of the factors. Alternatively "Oblique" rotation can be used this forces the alignment of variables but does not maintain independence.

Rotating the factors changes the position of the segments (factor scores) and the meaning of the axes (factor loadings). Rotating factors facilitates the interpretation of the axes. It is often useful to apply some general interpretation or name to the axes for discussion purposes. Oblique rotation makes this easier. However, the distance between segments becomes confused and may lead to incorrect conclusion regarding positioning.

3.6.3.2.1.2 Multiple Dimensional Scaling

An alternative means of forming maps is using Multiple Dimensional Scaling (MDS). This is an iterative method by which the best fit between the distances between points for the two dimensional map agrees with those from the multiple dimensional data[26]. It should be noted that if there is a very high degree of intercorrelation in the data both factor analysis and MDS will produce similar maps[27]. Because MDS maps conserve the inter-point distances, MDS is theoretically more accurate for judging position. However, the interpretation of the axes is more difficult and can be problematic[28].

3.6.3.2.1.2.1 Distance Measures and Non-metric MDS

There are multiple options in using MDS including the definition of distance and the method of optimization. For the most part, these variations do not greatly effect the results. MDS, however, does provide a means of estimating maps based on rank order data. This non-metric mapping can be useful with rating information.

3.6.3.2.1.2.2 Vector Meanings (Regression)

A major difficulty of using MDS is that the meaning of the axes is not computed directly in the process. Typically the meanings of the "dimensions" are obtained by regression in terms of the variables. Vectors can then be developed from the regression coefficients and the lengths estimated by the fraction of variation explained.

3.6.3.2.2 Space-Position Maps

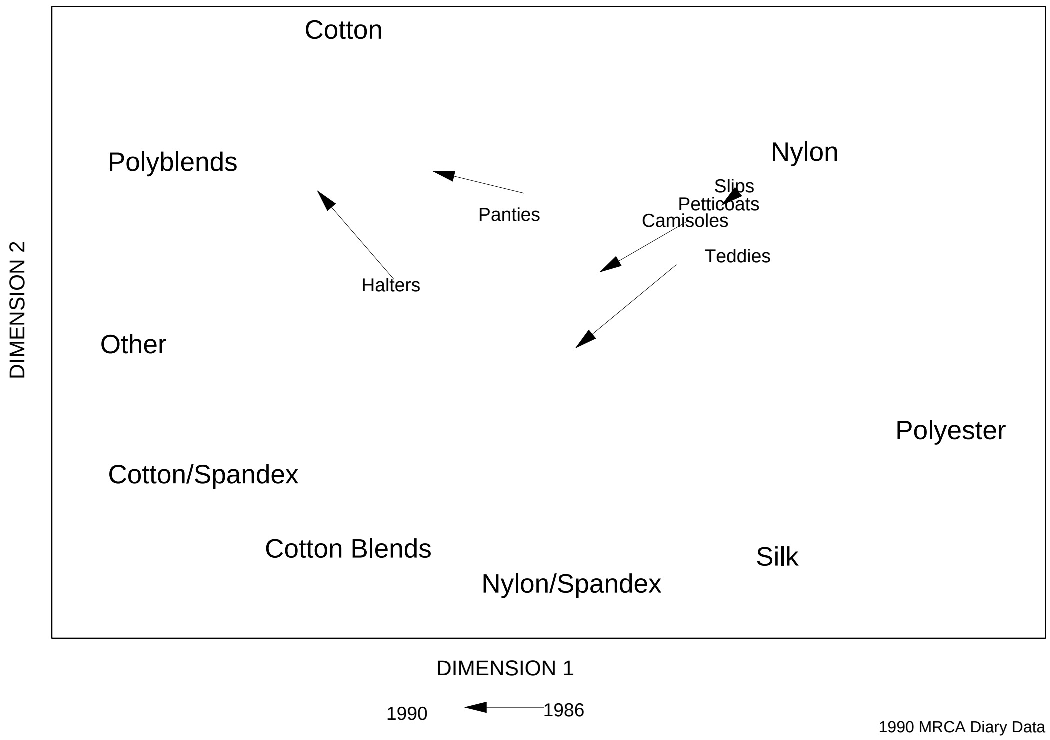
The Space-Position or point-to-point maps indicate both variables and segments as points. The distance between them is an indication of similarity and agreement. The following chart shows a typical point-to-point map on content. This chart shows the changes over time. Both vector-point and point-to-point maps can be used to display

[26] The mechanism of computation is similar between hierarchical clustering and MDS as factor analysis is similar to regression.

[27] The resulting maps may need to be rotated and flipped to be the same.

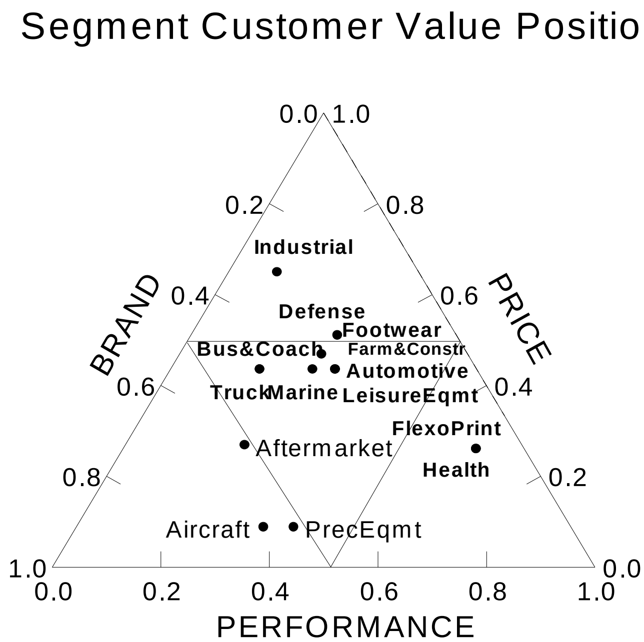
[28] Point-to-point maps are typically used with agreement statements rather than ratings of attributes.

dynamic data. Usually this type of analysis is done with normalized or share results. These can be computed for rating data. However, results using rating data tend to force segments to the center of the map.



3.6.3.2.2.1 Triangular Maps

The simplest form of point-to-point data is using triangular maps of three variables. Below is a typical triangular map for planning where the variables have been combined to form three groups that are then normalized. The positions of the variables are located on the vertices of the triangle. Since only three normalized variables are used, the map totally describes the data. This is not the case for larger numbers of variables.



3.6.3.2.2.2 Multiple Dimensional Scaling

Multiple Dimensional Scaling can be used to form point-to-point maps. Two methods are used: (1) fixed objects, and (2) MDS Unfolding.

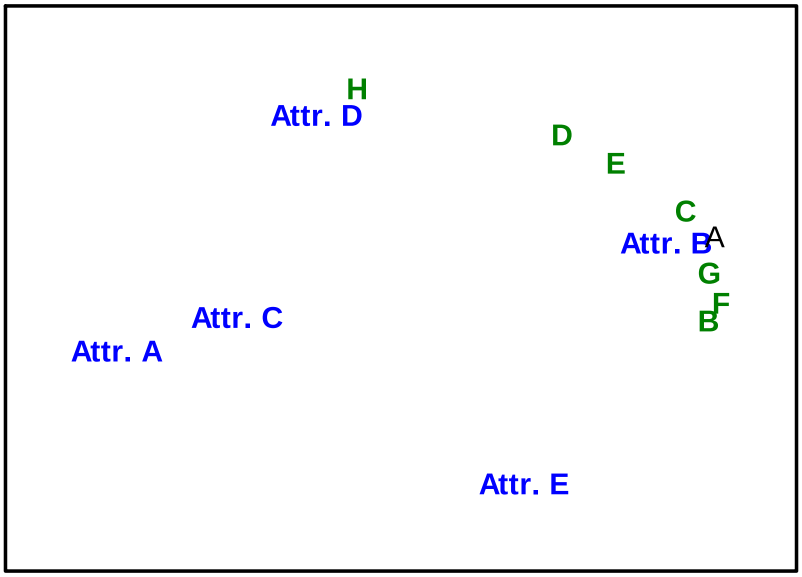
3.6.3.2.2.2.1 Fixed Objects[29]

To use Fixed Objects MDS the data needs to be normalized. The fixed object involves introducing pseudo-cases for the variables into the distance matrix. This is an identity matrix placed on the bottom of the distances. In effect, this assumes that these artificial cases will have 100% loading by one of the variables. MDS then computes the best fit map[30].

[29] This is not a standard MDS procedure nor a standard name for this proprietary tool

[30] There are two maps that can be produced: (1) with the variables around the segments and (2) the segments around the variables. Typically the variables around the segments are desired. This sometimes requires starting MDS with an initial configuration of the variables specified at the extremities.

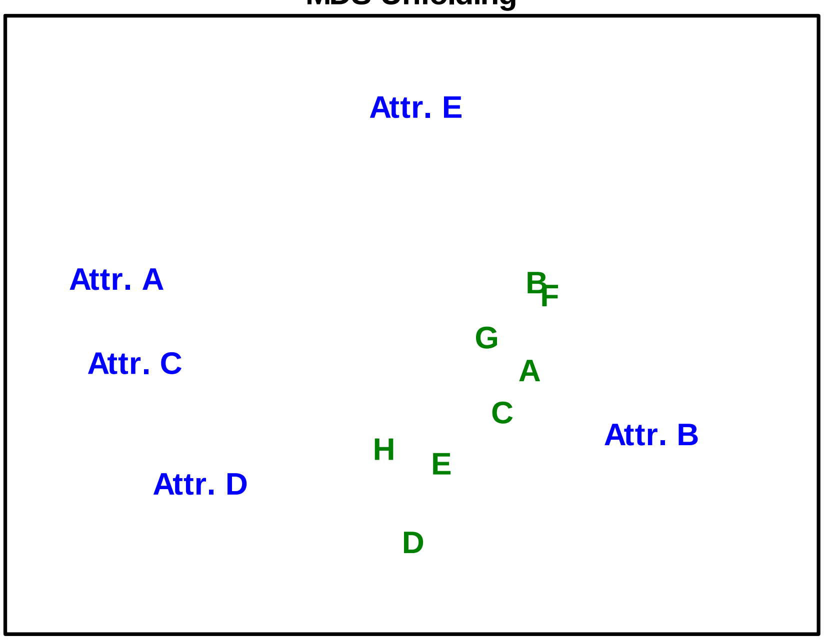
MDS Fixed Objects



3.6.3.2.2.2.2 Unfolding

MDS Unfolding consists of estimating the distance between the segments and variables and between variables. The same distance definition is used for all measures. However, since the number of cases typically is larger than the number of variables, this can tend to distort the geometry. It should be noted that the form of the map is very different from using Fixed Objects.

MDS Unfolding



3.6.3.2.2.3 Correspondence Analysis

Correspondence Analysis is a form of Factor Analysis that is capable of estimating point-to-point maps. The process involves doing factor analysis on the data and on the transpose of the data (Q-Factors). The results are then scaled to be superimposed. Correspondence Analysis is particularly useful with tabular results of categorical data[31]. Unfortunately, Correspondence Analysis requires a large number of fairly "heroic" assumptions and is therefore, not recommended unless there is little alternative.

[31] Typically Correspondence Procedures first scales the tabular data based on χ^2 distributions (similar to contingency table analysis) and then performs the two factor analyses and scaling.

3.7 SEGMENTATION

A key concept in marketing is the market segment that is a group of customers for which marketing programs, campaigns and products are targeted. Attribute studies are both analyzed by market segment and are used to identify them.

3.7.1 THE CONCEPT OF MARKET SEGMENTS

Market segments are defined in multiple ways and serve multiple functions. They should be viewed as views or perspectives on the market rather than inherent characteristics. Depending on what they are to be used for, there may be several sets of effective segmentation schemes. Individuals may fit into different segments depending on what is intended those segments to represent. The “best” scheme is that which allows the marketer to identify an opportunity and implement a plan. As such, the segments should have the following characteristics: (1) homogeneous characteristics, (2) common significant behavior, (3) be identifiable and (4) be useful in forming a marketing program.

3.7.1.1 Useful

Effective segments must always be useful. That is, the segments should facilitate the development of marketing strategy through the ability to target actions. The identification of segments should therefore allow for the influencing of buyer behavior.

3.7.1.2 Predictors of Behavior

Ultimately, segmentation is intended to help focus programs on things that will change behavior. As such, the segments should represent tendencies toward common and consistent behavior.

3.7.1.3 Significant and Important

The purpose of segmentation is to provide a useable and reliable alternative to an individualized sales perspective. As such, the size of the market segments must be large enough for the average values to be statistically significant. But beyond that, they need to be large enough to be important and worth consideration.

3.7.1.4 Homogeneous

Market segments should be fairly homogeneous in regards to key characteristics. By this we seek groups of customers who have more in common than members of other market segments.

3.7.1.5 Distinct and Reliable

Market segments should be significantly different or distinct from each other. That is the characteristics of each group should allow for unambiguous assignments to the segment.

3.7.1.6 Identifiable and Targetable

In order to directly use segments, the customers within them must be independently identifiable in means other than in terms of their beliefs and perspectives. One should be able to target programs to them. For that purpose they must be reachable or targetable.

3.7.1.7 Isolated

Finally, in order to apply independent marketing strategies to segments they need to be isolated. That is, actions in one isolated segment will not impact other isolated segments.

3.7.2 USES OF SEGMENTATION

As previously noted, segments are used for developing strategy. It is not necessary that the segmentation structure be common for all parts of a strategy. However, the more consistent that segmentation structure is, the simpler and more consistent will be the marketing strategy.

3.7.2.1 Structuring Market Data

Market segments become a basic element for structuring market and marketing data. The characteristics that are typically used to form data tables are the market segments. Market segmentation in this sense allows the examination of details of the market without expanding down to the individual respondent level. Segmentation allows the marketers to have a broad view of the variation in the market.

3.7.2.2 The Role in Positioning

Position and mapping provide a means of viewing the competitive situation. Segmentation presents the "battle field" in various perspectives. Because of the high variability, the respondent level usually shows a blur in position. The aggregate level typically hides variation in options. It is the segmentation that allows for a clear picture of differences and the identification of opportunities.

3.7.2.3 Segmenting the Market

From an analytical perspective, market segments are statistical groups of respondents. However, from a marketing perspective they represent groups of customers that would be uniquely attracted to specific product offerings. This action is referred to as "segmenting the market." A key use of analytical market segmentation is to identify groups of customers where this process is feasible.

3.7.2.4 Product and Offering Differentiation

Similarly, it is useful to differentiate the product offering. However, this usually must be targeted at groups of customers rather than the market as a whole. Here again the identification of the appropriate market segment is often critical.

3.7.2.5 Communications

Finally, communications may need to be different in both media and messages for these groups of respondents. There are often differences in the membership of communications segments and those used for developing product offerings.

3.7.3 TYPES OF SEGMENTATION

For analytical purposes, segments are divided into two types: (1) Prior or ad hoc segments that are imposed on the study and (2) Posterior or derived (sometimes referred to as "post hoc") segments that are identified from the research data.

3.7.3.1 Prior (Ad Hoc) Segmentation

Prior segmentation is done usually before the study is undertaken and often determines the sampling strategy. The criteria usually consist of historically accepted bases of segmentation. The criteria used for these types of segmentation are the expected characteristics for analysis.

3.7.3.1.1 Behavior Criteria

Behavioral criteria reflect the known behavior of the respondents. Characteristics such as: (1) products used or purchased, (2) customers and non-customers, (3) frequent users, etc.

3.7.3.1.2 Demographic Criteria

Demographic criteria reflect some external characteristic of the respondent or his activities. These include: (1) age, (2) gender, (3) size of business, (4) position or title, (5) function, etc.

3.7.3.2 Posterior (Post Hoc, Derived) Segmentation

Posterior or Post Hoc segmentation focuses on commonality of responses. This is usually done using statistical clustering or pattern recognition and is based on quantitative research data. The goal is to identify groups of respondents with common characteristics. The key in posterior segmentation is the use of many variables and the resolution of the segments from the commonality in the data. Because we can select different characteristics to use in forming segments, there are many types of posterior segmentation. None are necessarily "correct," nor is there often a single overwhelming form of segmentation that describes a market. The key goal is to identify homogeneous groups of respondents for which a strategy can be directed. The chart below shows the results of this type of segmentation with three clusters.

3.7.3.2.1 Benefit (Preference) Segments

Benefit segments reflect common preferences for features and benefits. These are usually computed based on the stated importance of features and product/supplier characteristics since that data is usually available from customer satisfaction studies. However perceived value (conjoint) and price sensitivity (choice modeling) data are also used for benefit segmentation. Benefit segmentation is particularly useful in the development of new and modified offerings and products. It focuses on the deals and products that would uniquely appeal to these groups of customers. It is a principal tool in positioning studies. Benefit segments are also useful in employee, stockholder and political survey studies. Included in this category is segmentation based on price sensitivity.

3.7.3.2.2 Attitude (Perception) Segments

Attitude segments reflect the attitude of respondents to products, concepts and firms. They are usually computed based on rating scales on the attributes of products. Typically this form of segmentation is used to identify "high probability customers" as well as to structure the market. Attitude segmentation is heavily used in political surveys as well as for commercial studies.

3.7.3.2.3 Demographic (Descriptive) Segments

Demographic and influence segments focus on the key characteristics of the respondent rather than his opinions of products/firms or on the benefits that he will get from them. Demographic segments are based on descriptive characteristics of the respondents. It groups respondents as having common identifying characteristics such as age, location, size of firm, background.

3.7.3.2.4 Influence (Psycho-Sociological, Value) Segments

Influence or sociographic segmentation covers a broad range of criteria including the nature of the decision process, lifestyles, values, hobbies, personality, etc. The goal is to identify common groupings of people that can be appealed to. Usually this is connected with communications strategies and is associated with consumer marketing rather than business-to-business promotion.

3.7.3.2.5 Communications Segments

Communication segments focuses on the preferred channels of communications. Here we seek common vehicles of communications both in terms of where and what is used. This includes reading materials, mass communications, conferences, trade shows, and lately the Internet.

3.7.3.2.6 Decision or Behavioral Segments

Decision or cognitive segmentation consists of identifying groups of respondents with common decision rules. The objective is to group respondents who are expected to behavior in a similar fashion. Methods to do decision segmentation tend to be more heroic than other clustering approaches and therefore are consider to be more speculative.

3.7.3.2.7 Meta-Segments

Meta-segmentation consists of merging a number of segmentation schemes and identifying interactive groups. Typically benefit and attitude segments are used to identify groups of respondents that have both common preferences and common perceptions of the products.

3.7.4 POSTERIOR SEGMENTATION TOOLS

It should be noted that identifying segments is as much of an art form as it is statistics. There are a large number of tools for posterior segmentation. The tools that are identified and discussed here are those that we have found useful. The resulting segmentation structure, however, needs to some extent conform to rational expectations. The results must have some degree of "face" validity; that is, the resulting segments must be "reasonable."

3.7.4.1 Issues of Clustering

When multivariate statistical analysis is used to identify segments, they are referred to as "clusters." The definitions of these clusters are determined by the specific ways in which the data is prepared and the methods used. There are two general types of clusters that can be developed: (1) hard clusters where the respondents are assigned to a specific group and (2) soft clusters where a percentage participation in a cluster is computed.

3.7.4.1.1 Hard Clusters

Traditional posterior segmentation is done based on hard clusters. Each respondent is assigned into a specific cluster and all respondents are assigned to clusters. The characteristics of the clusters are determined by their average attribute values or occurrence of dissatisfaction. Since it defines clearly distinct groups, hard clusters are very useful for positioning and for formulation of marketing strategies.

Problems and Issues:

Underlying this approach is the assumption that such a structure is clear in the sample. The clustering tools will make the assignment whether or not it is inherent to the data. The problems using hard clusters are:

- Due to the forced assignment, hard clusters are an approximation of the market structure.
- The clusters may become ill-defined if poorly fitted respondents are assigned to clusters.
- There is usually a "catch-all" cluster, which is a grouping of non-fits.

- The cluster structure depends on the number of clusters chosen; the type of preparation of the data; and the type of procedures used.

3.7.4.1.2 Soft Clusters

Soft clusters involve assigning a value that relates the respondents to each of the possible clusters. This is often a probability or weight. It is often useful when using soft clustering to think of the cluster points as "archetype" descriptors of the segments. Each respondent is then assigned a value to each of these archetypes based on a "distance or loading." These archetypes can be associated with the mean values of the hard clustering. As such, soft clusters can be computed from any hard cluster structure. Similarly, hard clusters can be assigned from a soft clustering approach with the application of specific rules for assignment. Because, soft clustering provides a flexible view, where a respondent will belong to some extent to several segments, it can provide a means of viewing complex market structures.

The standard method of estimating soft cluster values for hard cluster assignments is based on *Multinomial Logit Regression*, where the probability of belonging to a cluster is estimated from the values of the clustering parameters. This approach can be used irrespective of the way the hard cluster assignments were obtained.

Problems and Issues:

It is implicitly assumed in this approach that the archetypes are meaningful descriptors of the data. These "archetypes" represent underlying characteristics of the market. However, due to the partial assignments, soft clustering is a better approximation of the data structure. The problems using soft clusters parallel those for hard clusters.

- Once again, the cluster structure depends on the number of archetypes chosen; the type of preparation of the data; and the type of procedures used.
- However, there are some overall tests of "goodness of fit"
- It provides a more "fuzzy" structure that is often not actionable.

3.7.4.1.3 The Number of Clusters

While some of the statistical procedures will suggest an optimum number of clusters, in general the selection of the number of clusters is "arbitrary" in that it is chosen by the analyst. Rules help in the selection of the number of clusters to be used based on respondent participation. However, it is still more of an art form than a science in determining the appropriate number of clusters to be use.

There are four methods for identifying the "best" number of clusters to consider based on: (1) the distinction of the average criteria by segment, (2) the grouping of distance among respondents (using the dendrogram generated by hierarchical clustering), (3) tracking the emergence of clusters and (4) improvement in the "*Quality of Segmentation*". The distinction of the average criteria by segment is based on the average values of the parameters used to form them. This is shown in the *Segment Profile* where the standard values of the parameters are compared. The *Segment Profile* is discussed later. The objective is to have a sufficiently large set of segments so that the results are "reasonable" and satisfies strategic needs.

The dendrogram is discussed later. It is used for identifying the number of proper clusters is subjective. It is based on visually identifying appropriate breakpoints in the structure. From these breakpoints estimates of the appropriate number of clusters can be made. This is usually done early in the analysis to provide the initial estimate of the number of clusters.

Tracking the emergence of clusters is based on generating a series of solutions and tracking where the new clusters emerge. Typically there are four processes:

- *Splintering,* gives rise to outliers (very small groups 1 to 3 participants);
- *Parenting,* the splitting of pieces from a single dominate group;
- *Bifurcation,* the division of previously formed clusters;
- *Reformation,* the formation of new clusters with participation for a number of older ones.

In general, the preferred segmentation structure consists of a number (3 or more) of clusters each representing a significant portion of the sample. As such, outliers are usually ignored. Both the *Parenting* and *Bifurcation* processes are the expected method of formation of new clusters. It is useful to think of these processes as based on a hierarchy of selection criteria. New clusters arise by the entry of a new criterion. *Reformation* appears to be different in that new criteria overwhelm the previous sets. This usually results in less distinct cluster structures. Usually *Reformation* is a subsequent process in clustering and therefore, it is usually feasible to terminate the cluster selection before *Reformation* starts.

3.7.4.1.4 The Quality of Segmentation (Clustering)

Statistical clustering processes assign respondents to segments irrespective of the quality of fit. This is just the nature of the process. The key question is how good is the fit or alternatively what is the quality of the assignment. Estimates can be obtained using soft clustering probability values (usually from *Logit Regression*). The average appropriate probability assignment can be used as a measure of the quality. This represents the average of the probability that respondent has been assigned to the correct segment. Here a 99% Quality Value means that the average probability from the Logit regression model for each appropriate assignment is 99%. Alternatively, the measure of mis-assignment can also be used. However, this is directly related to the probability of proper assignment. In either case the Quality of Segmentation captures the distinguishing power of the clustering scheme.

The measure of the Quality of Segmentation can be used to quantify the improvement in the clustering due to the additional of new segments. It should be noted that the Quality of Segmentation should always increase with the number of clusters. This is due to two factors. First, the removal of isolated values and outliers and then the reduction of the probability of random assignments. On average, the probability of a random assignment should be equal to the inverse of the number of clusters ($1/\eta$ where η is the number of clusters). This would mean that there should be a continuous improvement in assignments as the number clusters increases[32]. However, that increase is unlikely to be uniform. At some point, there will be little improvement and therefore, little value for more clusters.

3.7.4.2 Analytical Methods

There are a fairly large number of analytical methods to define clusters. They are traditionally divided into four groups: (1) cluster analysis, (2) discriminate analysis (mainly Q-Factors), (3) regression techniques, and (4) neural-nets. Of these tools, cluster analysis is far and away the most widely used technique for posterior segmentation.

3.7.4.2.1 Hierarchical Clustering

Hierarchical clustering is the most general form of hard cluster analysis and is based on computing distance measures between all respondents[33]. The clusters are defined by the relative distance between groups of respondents. The clusters can be defined by both the internal distances within clusters and the distance between clusters. There is also a large number of ways distance can be defined. This results in a broad portfolio of possible procedures.

However, it is usually used more to "scope-out" the cluster structure than to assign the specific membership into the clusters. A result of hierarchical clustering is the dendrogram that shows the relative distance and connections among the respondents on the selected characteristics. A sample of this type of tree structure is shown below. The closer the respondents are positioned and the shorter the branch the more similar they are.

Problems and Issues:

There are several problems in using hierarchical clustering with survey data.

---

[32] The Quality of Segmentation can be expressed as a ratio of the relative average probability of proper assignment to the maximum range or:

$$[P - 1/\eta]/[1- 1/\eta] = [P\eta -1] /[\eta-1]$$

where P is the average probability of proper assignment. The first derivative in respect to the number of clusters is $[1-P]/[\eta-1]^2$ which is positive for all values of $\eta > 1$ and $P < 1$. This indicates a continuous increasing Quality with the number of clusters. However, the impact of this decreases with the negative second derivative.

[33] Hierarchical clustering is referred to as CLUSTER in both SPSS and as a procedure in SAS, but is referred to as Hierarchical Clustering in SYSTAT.

- Most statistical packages have difficulty doing hierarchical clustering on very large datasets. More than thousand respondents become very time and memory consuming[34]. However, with faster personal computers, hierarchical clustering can be done using several hundred respondents.
- The multitude of methods produces a problem of appropriate selection. While a standard Euclidean point distance using either Complete or Wards methods are usually used, it is not appropriate in some cases.
- Most hierarchical clustering tends to be very sensitive to "outliers" where single entries are identified as separate clusters.
- Hierarchical clustering is not particularly sensitive to data preparation. While cases containing missing data can not be used, it is unnecessary to normalize the data.

3.7.4.2.2 Center Based Clustering (K-Means/ Hill Climbing)

Average clustering based on assigning respondents to clusters based on distances from moving group averages. These iterative methods are usually efficient and can handle large datasets. The K-Means method[35] is probably the most popular of these techniques and is the preferred method for identifying hard clusters.

Problems and Issues:

There are several problems in using K-Means clustering with survey data.

- The number of clusters has to be predetermined.
- There are a number of alternative options in defining the cluster assignment procedures. However, they do not tend to produce significant difference in solutions.
- The technique is very sensitive to data preparation and requires normalized data.
- K-Means clustering can be sensitive to "outliers" where single entries are identified as separate clusters.

3.7.4.2.3 Q-Factors

Factor analysis can be used to group respondents in a similar fashion that it is typically

---

[34] Random sampling can be used to obtain a sub-set for hierarchical clustering with very large datasets.

[35] K-Means clustering is referred to as Quick Cluster in SPSS, FASTCLUS in SAS, and K-Means Clustering in SYSTAT. In addition, an algorithm is available by Sawtooth Software that performance K-Means clustering for several of their procedures.

used to group variables. This is referred to as "Q-Factors analysis." The dataset is transposed and standard factor analysis performed[36]. The resulting factors are viewed as "archetypes" The factor loadings represent respondent weights by archetype. The factor scores define the archetypes in terms of the variables.

There are many appealing characteristics to this type of clustering. The resulting clusters are determined by the data structure. Variations in procedures tend to have very little effect on the results. The resulting soft clusters are a good representation of the market structure. The number of clusters is determined based on the amount of variation that is captured. And finally, there is a natural measure of "goodness of fit."

Problems and Issues:

However, there are a number of serious problems with Q-Factors.

- It can not be done with large data sets. More than 500 cases will produce major problems in most statistical software.
- Soft clusters and archetypes are difficult to explain to clients and
- The procedure is difficult to explain to clients.

3.7.4.2.4 Mixture Regression (Latent Class Regression)

Multilinear and various types of non-linear regression techniques can be used to identify soft clusters. This is generally referred to as the "Mixture Regression"[37] or as "Latent Class Regression"[38]. It is most useful for identifying decision segments, respondents using similar decision or buying rules. The concept is to identify sets of regression models that will optimize some collective criteria. Usually the general criterion is to maximize the variance explained by the predictive (dependent) variable across the models. As with all statistical clustering procedures, regression clustering is basically an assignment problem, where the individual cases or respondents are assigned to a group.

There are two general methods for seeking a solution: (1) multiple staged regression process is used to identify clusters by iterative focusing on outliers[39] and (2) seeking the

---

[36] Usually Varimax rotation is used. However oblique rotation can also be used, but due to intercorrelation among archetypes, this procedure produces problems with assigning contributions.

[37] This is specifically referred to as finite mixture models. However, that term is also used for all types of statistical clustering procedures.

[38] Latent Class Analysis packages allow for automated prediction of the assigned groups using the latent class structure and regression analysis. However, we have found it more useful to do the process directly using iterative regression.

[39] It is usually better to use a non-least squares approach. Least Squares approaches bias results toward reducing large variations and therefore, tends to identify global "best fit" solutions. Logarithmic or Maximum Likelihood procedures emphasize short-range variation and therefore, identify local solutions that are preferred for segmentation. Other "loss" functions can also be used for the same

optimum coefficients directly. The iterative procedure is similar to K-Means but uses a regression procedure rather than mean values and assignments. Mechanically the procedure would start by identifying outliers from an internal collective model and then use the outliers for form a hierarchy of models. These models than used to identify the best assigned cases by variation in the dependent variable and new regressions are then generated.

Problems and Issues:

There are some serious problems using regression clustering, which tends to restrict its use. These problems include:

- Models using regression clustering are not unique and can depend on the method of initiating the process. Final solutions, therefore depends on the way in which the first estimates were obtained. While global optimum solutions may exist, there is no assurance, that the ones found are global.
- Regression clustering can be unstable and therefore, sometime unreliable. That is, in tests with structured data, regression clustering may not reveal the true underlying structure.
- As is the case of other statistical tools, regression clustering can indicate clusters when they may not exist.
- Goodness-of-fit only refers to the sub-population. There is no overall good-of-fit measure. Because only data which agree with the sub-models are used in their determination, goodness-of-fit are by definition very high but do not reflect the overall fit of the collective models.
- Assignment models (based on Logit regression) can indicate weak association. This is due to the nature of regression process itself which is an approximate fit.

3.7.4.2.5 Latent Class Analysis [40]

Latent Class Analysis is a relatively method for identifying clusters using categorical data. It can be thought of as a type of regression clustering procedure but not the same as Latent Class Regression. This method uses predictor/correction (EM type) algorithm combined with Logit regression to estimate assignments into cluster groups. This

---

purpose including truncated approaches such as those using Robust Regression procedures.

[40] Latent Class Analysis is now available in SAS but not in other standard statistical packages at this time. Neither SPSS, nor SYSTAT include the procedures. However, there are a number of special purpose packages that are available, several that are in the public domain including LLCA, LCABIN, LCAP, LEM, and MILLSA. I have used LCAP and found it very easy to use but not compatible with standard statistical packages and works exclusively in the DOS environment. There are also several commercial packages including Latent Gold.

procedure is similar to the EM algorithm used to estimate values of missing data. But in this case, all of the data is missing from the virtual "latent class" variables. Because all of the data is categorical (discrete) standard linear regression would be inappropriate. Logit regression is used instead. This produces soft cluster values as well as assignments. That is, probabilities of being within a cluster are computed. Because of the ability to handle categorical and large amount of missing data, this method of clustering as become popular.

Problems and Issues:

There are some serious problems using Latent Class Analysis, which tends to restrict its appropriate use. These problems include:

- All of the problems and issues seen with Regression Clustering apply equally to Latent Class Analysis.
- Other methods of clustering categorical data (Latent Variable Clustering discussed below) tend to produce more discriminated groups.
- The results of Latent Class Analysis are not unique. Different initial points will produce different assignments. While this may be the case with other Regression Clustering techniques, it is particularly noticeable here. This is a particularly difficult problem. Most of the available packages allow for multiple solutions.

3.7.4.2.6 Latent Variable Clustering

Latent Variable Clustering is an alternative approach to segmentation using categorical data. This method uses factor analysis to identify a number of metric latent variables that captures of variability of the underlying categorical data. The cases are then hierarchically clustered based on these latent variables. The resulting clusters or segments are then analyzed in terms of the original categorical variables to determine the distinction and meaning of each cluster. The process after factor analysis is the same as hierarchical clustering requiring the definition of distance and linkage.

Problems and Issues:

Because these methods are based on hierarchical clustering they carry the same problems and issues. In addition, there are a number of other considerations and difficulties including.

- Because surrogate variables are used for the clustering, the discrimination on the underlying categorical variables may not be as good as using primary variables.
- Typically, categorical clustering involves a fairly large number of variables, this may lead to some variables not being distinct in the resulting clusters.

3.7.4.2.7 Hierarchical Bayesian Clustering[41]

Hierarchical Bayesian procedures can be thought of as a probabilistic version of predictor/corrector analysis[42]. It is based on the use of Bayes' Theorem for correcting a prior estimate with subsequent information. Similar to the other predictor/corrector methods such as the EM algorithm, Hierarchical Bayesian procedures can handle missing data. In fact, its major advantage for clustering is its ability to handle heavy levels of missing data. It is actually able to handle "sparse" data matrices, where most of the data is missing.

The basic clustering procedure is similar to regression clustering in that it is a sequential approach starting with a model for the total data. This is used as the "prior" solution. Subsequent sampled solutions are computed based on this global solution.

Problems and Issues:

Though its ability to handle sparse datasets is attractive, there are some serious problems using this type of clustering, which tends to restrict its use where no other means are available. These problems include:

- It is a heroic method based on variation from a prior solution. While it has been suggested that the eventual solutions always dependent on the prior, there is no way of eliminating its influence.
- It does not generate unique solutions. Multiple applications of the algorithms produce multiple solutions.
- There is no explicit (mechanistic) logic justifying the assignments.

3.7.4.2.8 Neural-nets

Neural-nets are a set of non-statistical procedures for assigning clusters. It consists of a learning procedure in which separation of respondents are made based on some "goodness-of-fit". It is based on an array of computational filters and is design to simulate simple neurological decision processes. The result of the process is the clustering and the ability to predict the cluster of additional data. While there is some reported success using the procedure for non-survey analyses, there is little experience to support its use for this type of application.

---

[41] Hierarchical Bayesian Clustering Analysis is not available on the standard statistical packages at this time. WinBugs is a public domain package is widely available and compatible with Microsoft Windows.

[42] This view of Hierarchical Bayesian analysis was suggested by Professor Eric Bradlow (Wharton) relating it to the EM algorithm.

Problems and Issues:

There are some known problems in using the procedure.

- The process does not produce a clear set of rules that explains the assignment
- In most cases, tests of neutral-net computations did not out-perform statistical techniques.
- It usually requires a very large database for neural-nets to be effective[43].

3.7.4.2.9 Preparing the Data

Multivariate statistical analysis, in general, and cluster analysis, in particular, can be very sensitive to the preparation of the data.

3.7.4.2.9.1 Monotonic Measures

Clustering is simplest if the measures used are monotonic on value. That is, that all measures of goodness are in one direction. Without monotonically valued measures it is not feasible to normalize or standardize the data. The results will be clusters heavily influenced by the general response to all of the attributes rather than concentration on specific attributes.

This often happens with agreement scales when it is desirable to reverse order to reduce potential response bias. In these cases, the selected counter attribute scales need to be reversed before further data preparation and clustering.

3.7.4.2.9.2 Normalization and Standardization

K-means Cluster analysis tends to group respondents based on the “average value” of the attributes first and then by the relative weights of the individual attributes. If the sum of the data differs among respondents, the total will be primarily criteria used in clustering. Under this condition the cluster structure will merely be the range of total attribute scores. On the other hand, with constant sum data, where the totals of the attributes are held the same for all respondents, clustering focuses on the importance of individual attributes. This is usually the desired result.

Normalization of the data involves setting the total of the ratings equal to a constant, usually unity. The spread in the data may vary from respondent to respondent.

Standardization[44] of the data involves fixing both the total value usually set at zero and

---

[43] Most successful applications of neural-nets have been in either repetitive recognition such as for handwriting and fingerprints or with large scale data-mining.

[44] Standard Value = (Value - Mean)/Standard Deviation

the spread or variation of the data, usually set at unity. With standardization, the focus of the clustering is on the specific attributes of interest; variance does not play a role.

Problems and Issues:

While it appears that these approaches are similar, they can produce very different cluster structures. Normalization allows for clustering on variation among attributes while standardization focuses on the main attribute. As a general rule, we have tended to rely on normalization since variation is a major difference among respondents.

3.7.4.2.9.3 Missing Data

Almost all statistical techniques[45] for segmentation require complete data. This is probably the greatest problem in using survey data and greatly restricts the number of variables that are considered. Usually, when segmentation is a key objective of the study, an effort is undertaken to minimize missing data. Beyond the problem of reduced data, is the potential problem of bias in the sample. If feasible, we tend to select the sub-group of variables where missing data is least offensive. For traditional clustering methods (hierarchical and K-Means) all missing data must be corrected. Typically a regressive type missing data method such as the EM algorithm can be used. Note, however, that using missing data correction can introduce bias into the clustering solutions.

3.7.4.3 Scaling of Data

Most of the techniques for identifying clusters are based on quantitative data (interval or ratio scaled). Usually we consider ratings and constant sum scales to be quantitative. However, there are times when both normative and ordinal data must be used. At that time, methods have to be developed that defines a quantitative distance measure. There are several statistical methods including, correspondence analysis which operates on cross tabular data to define Euclidean (quantitative) measures.

3.7.4.4 Meta-segmentation

When there are multiple definitions of segments, it is useful to combine them into a single structure. For example, we may define both benefit and attitude segments. However, we want to look for opportunities where these are consistent or opposing.

With hard clusters identified, it is necessary to form a quantitative distance measure. This can be done using correspondence analysis or one of the normative distance measures provided in the clustering procedures. With soft clusters, the percentage

[45] Latent Class Clustering methods, however, do allow for estimation with missing data. Other methods such as Q factors are based on the covariance matrix that can include missing data. Furthermore, neural-net can also handle partial data. However, the assignment of respondents using these methods can only be done for cases with complete data

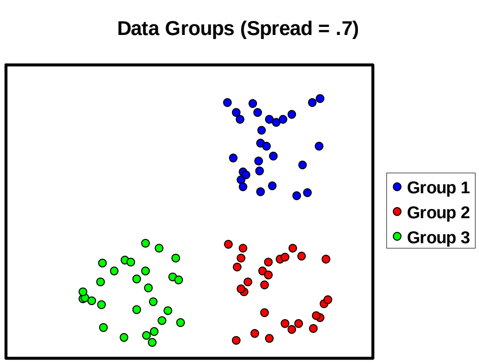
assignment for the respondents can be used as attributes and standard quantitative clustering is then used.

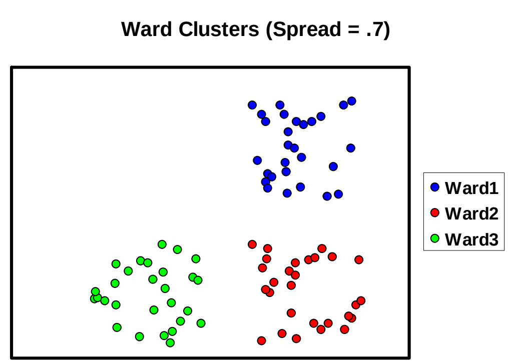
3.7.5 Reliability

Clustering imposes order onto the chaos of data. It is a group of methods that lives up to the adage that "if you torture data enough, it will confess." Using any of the hard clustering tools will result in the assignment of all respondents to some cluster. However, reliability of the results is the key issue; that is "Do the resulting clusters reflect the underlying structure?"

3.7.5.1 Impact of Distinction

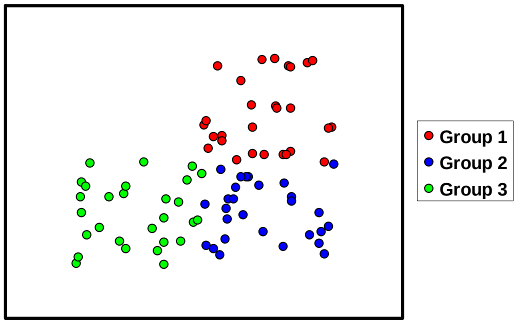
When segments are distinct, any clustering tool should identify the items. Below are the results of clustering synthetic data in which each dataset has different degrees of "distinction" among three segments[46]. The graphs on the left in the figures below indicate the original grouping of data and the figure on the right is the resulting cluster structure. The clustering was done using the hierarchical Ward linkage method on the distance between points. In the first case, the random fluctuation is not large enough to merge the segments; the maximum distance is only 70% of the distance between the center of the clusters. In this case, any clustering technique should assign independent clusters that correspond to the original groups. In the second case, the original groups just touch. Here the clustering just about reproduces the original data with a few exceptions. In the remaining two cases, the data of the groups merge. In the third case, there is some overlap but the groups have some distinction and the clusters appear to correspond to those groups. However, notice that the participation is far different from the original groups. In the last case, the distinction almost disappears. Here the clusters appear to be very different from the original data. This is a characteristic of the clustering process forcing order, whether or not it exists.



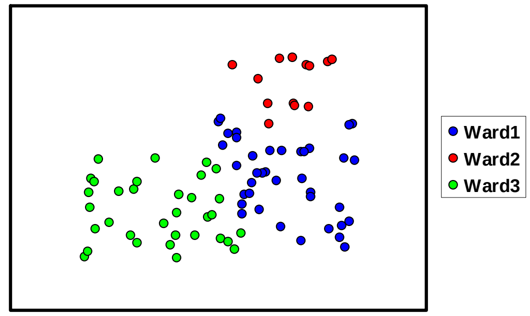


[46] These two dimensional datasets were constructed by introducing different spreads of random fluctuation. The difference in the centroids of each group is one unit. The spread represents the maximum size of the uniform random variation that is added to the mean values.

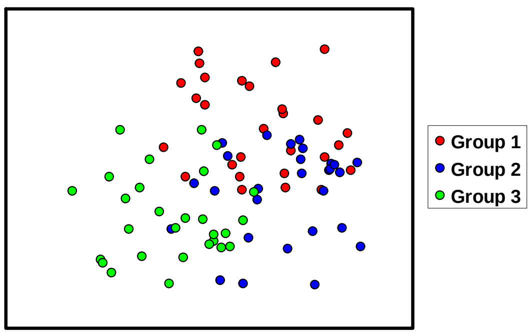
Data Groups (Spread = 1)



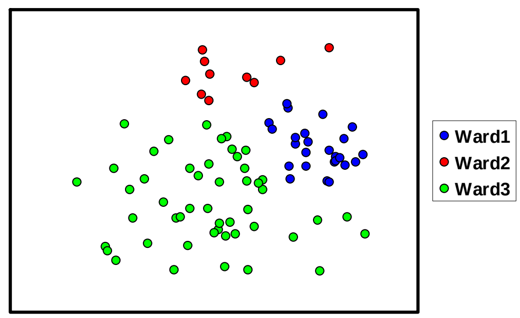
Ward Clusters (Spread = 1)



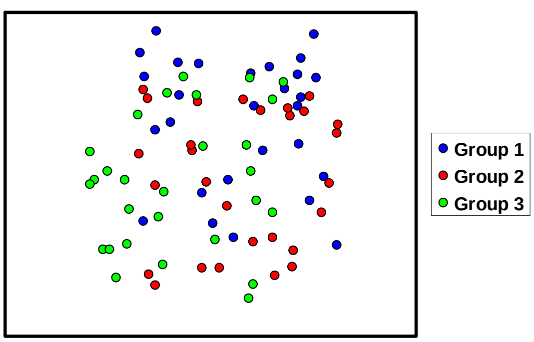
Data Groups (Spread = 2)



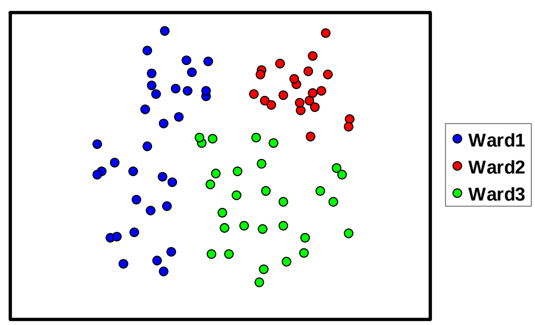
Ward Clusters (Spread = 2)



Data Groups (Spread = 4)



Ward Clusters (Spread = 4)



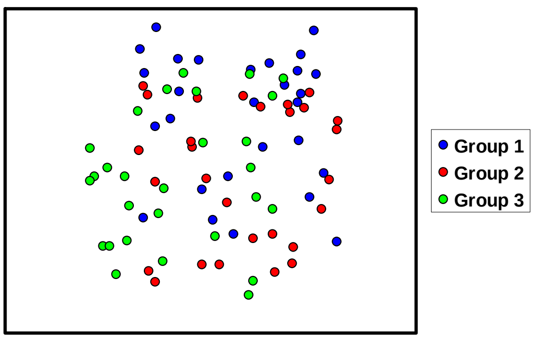
3.7.5.2 Comparison of Methods

There are a number of clustering techniques that are typically used. As previously noted, if the data is distinct, any of the various methods should reproduce the underlying structure. The problem exists in the indistinct case, where the structure is hidden within the noise. To examine this situation the worst synthetic case is reexamined using a number of clustering methods. These included hierarchical clustering (on distance) using Single, Complete, Centroid, Average, and Ward Linkage methods and using the K-means averaging method.

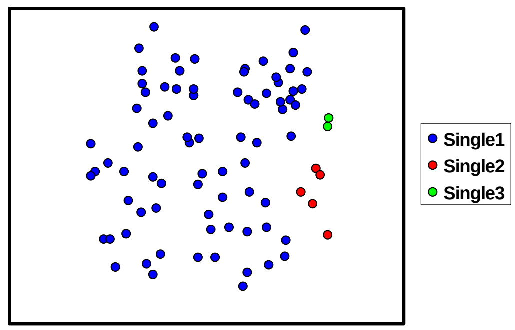
Typically the Ward Linkage and the K-means methods are preferred, as will be discussed later. The first chart below on the left is the same data group chart shown above and is for reference. The Single linkage method involves the simplest view of the data. Notice that this method did not show significant distinction within the dataset. It is probably the most realistic interpretation of the data itself without the imposition of order. Unfortunately, it neither reproduces the underlying structure nor meets the needs for identifying clusters.

The next chart down on the left shows the results of Ward Linkage and the one on the right is the results of K-Mean clustering. Notice that these are very similar[47]. The last three charts show other hierarchical clustering methods. The Centroid and the Average Linkage methods are similar and produce similar results but are very different from the Ward method. The Complete Linkage methods produced another very different result.

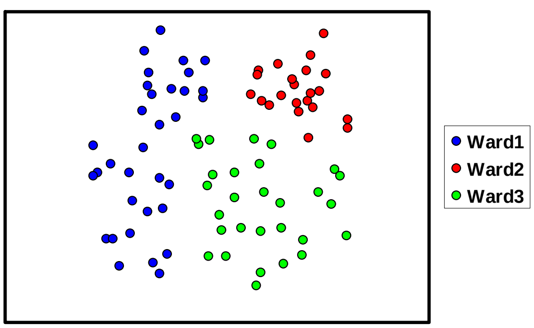
Data Groups (Spread = 4)



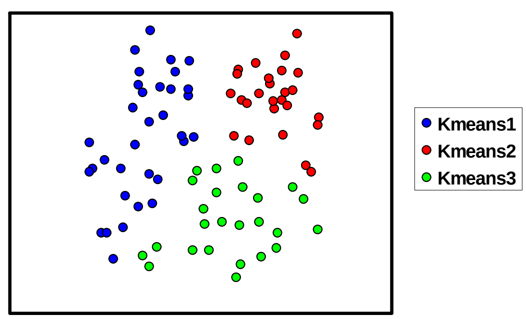
Single Clusters (Spread = 4)



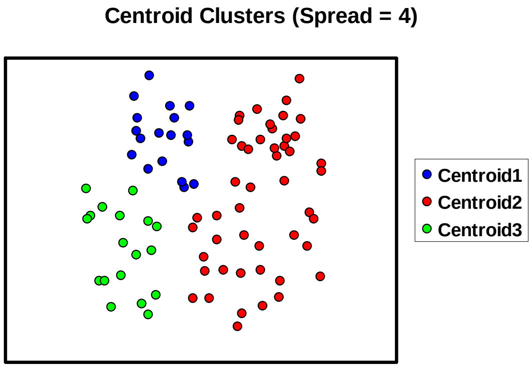
Ward Clusters (Spread = 4)

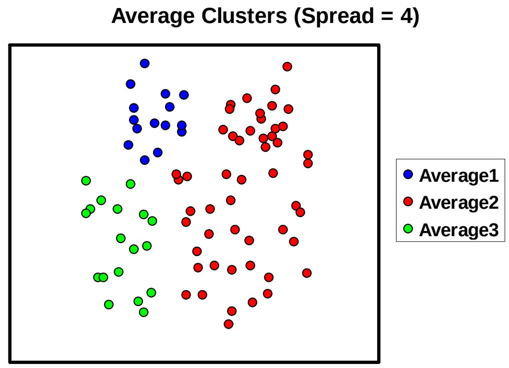


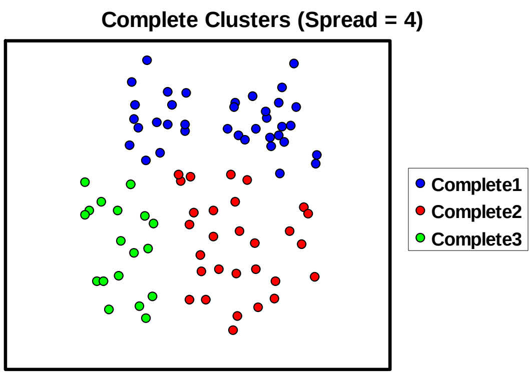
Kmeans Clusters (Spread = 4)



[47] It should be noted that these methods are extremely different and while it should be surprising that the results are similar, it is a consistent finding.







It should be noted, however, that had we used four clusters, the results from Ward, K-Means, Centroid, Average, and Complete Linkage would have likely been more similar. However, they would not reflect the underlying structure that was used to construct the dataset. Below we should a comparison between these clustering results on the underlying data. Notice that none of the clustering procedures produce a good agreement between the assigned clusters and the original groups. At best, only 54% of the points were properly assigned. However, the mean values of the clusters using K-means and Wards Linkage methods were in pretty good agreement with the original groups. The other methods produce wildly different cluster structures than that originally generated. For this reason we tend to use K-Means and Ward linkage in general segmentation applications. However, it should be noted that the Single Linkage method is probably the best view of the data structure.

	Against Original Data Model					
	complete	average	centroid	single	kmeans	ward
Data Assignments	**46%**	**50%**	52%	68%	70%	70%
Averages	71%	56%	55%	62%	**16%**	**13%**
Size of Clusters	28%	27%	51%	124%	**13%**	**10%**

3.7.6 Preferred Methods

Posterior benefit and attribute segmentation (that we are considering here) are typically based on cluster analyses of individual importance and perception data. The key problems in this type of segmentation are determining the number of segments and the best method of their definition. The goal of the clustering process is to identify distinct groups of respondents with common but unique characteristics.

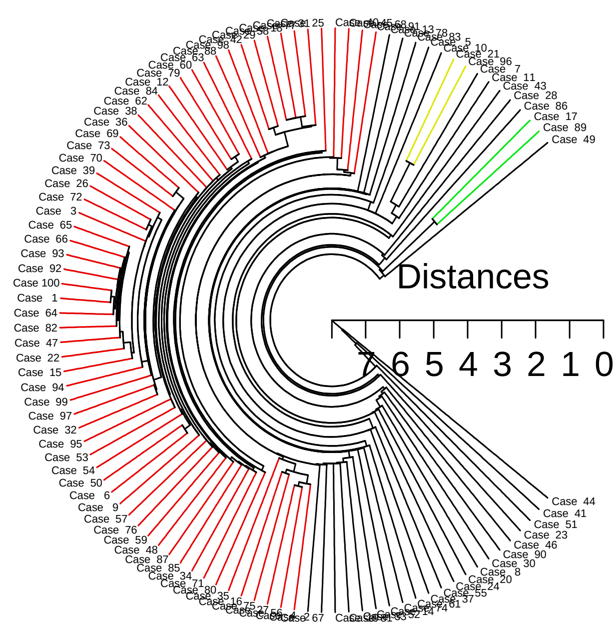
3.7.6.1 Data Normalization

As previously noted clustering tools are very sensitive to total values of individual responses. It is critical that all of the importance and attribute rating results sum to a constant value. However, we try to retain the impact of the variation in the spread of the data that standardization would remove. We, therefore, **normalize** the data.

3.7.6.2 Hierarchical Cluster Overview

Hierarchical clustering displays metric distances between respondents and tends to indicate the number of appropriate clusters. However, this form of clustering is based on the similarity of data not on common characteristics. The chart below shows the dendrogram tree graph indicating the relative similarity among respondents[48]. Hierarchical clustering is used as a first step in the clustering process.

Cluster Tree



[48] This chart shows the "polar" version of the tree. The standard form is a large linear plot. However, I find this form much more useful.

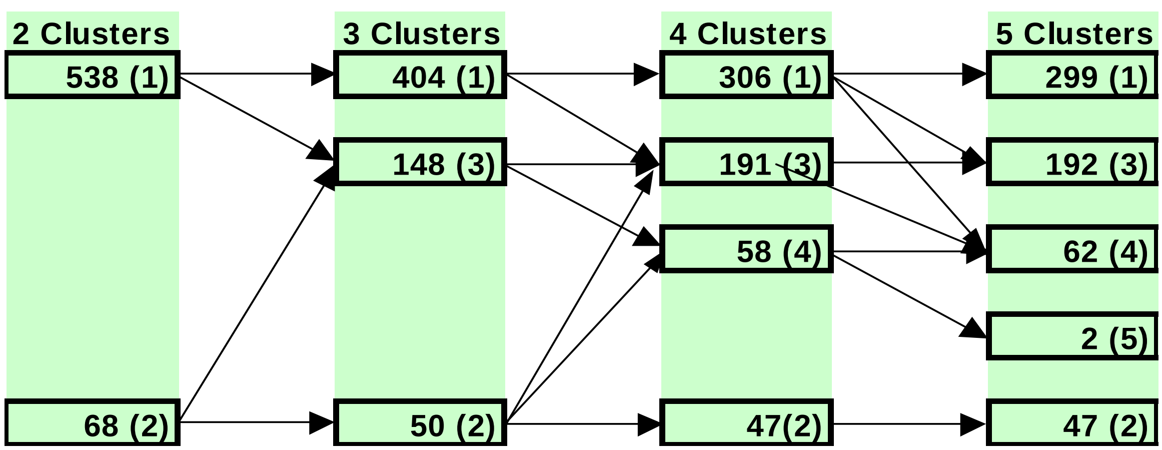
3.7.6.3 Cluster Identification

Hierarchical clustering is the traditional clustering method of choice for small and moderate sample sizes. Typically three to five segments are used. There are a large number of options using these methods. Typical several kinds of distance measures between points and between clusters (linkage) can be defined[49]. Position maps, linkage maps and profile analyses are used to determine the best solution.

K-Means clustering is a common "hill-climbing" approach to clustering based on identify distinctive groups with large numbers of respondents. Similar to hierarchical cluster, these procedures require setting the number of desired clusters as well as distance measures. In this case, we typically generated groups that have from two to six clusters. Tree structures, position maps and profile analyses are used to determine the best solutions.

3.7.6.4 Tree Structure

Cluster trees show the results of tabulation of the cluster generation process and help to identify the "break-up" of the clustering structure. K-Mean clustering leads to a bifurcation of the respondents where new clusters arise from other large clusters. When clustering generates outliers and where new clusters arise from mixes of many previous ones, clustering has probably gone too far. In this case, we viewed groups containing three or four clusters as the best solutions.



3.7.6.5 Conditions for Segmentation

We view the K-Means clustering as a cascaded process starting with two clusters and increasing the number generated. Older clusters are split to form new ones. In general, we are looking for the condition where new clusters arise mainly from a single older cluster. This represents a bifurcation process where new classifications arise. This can

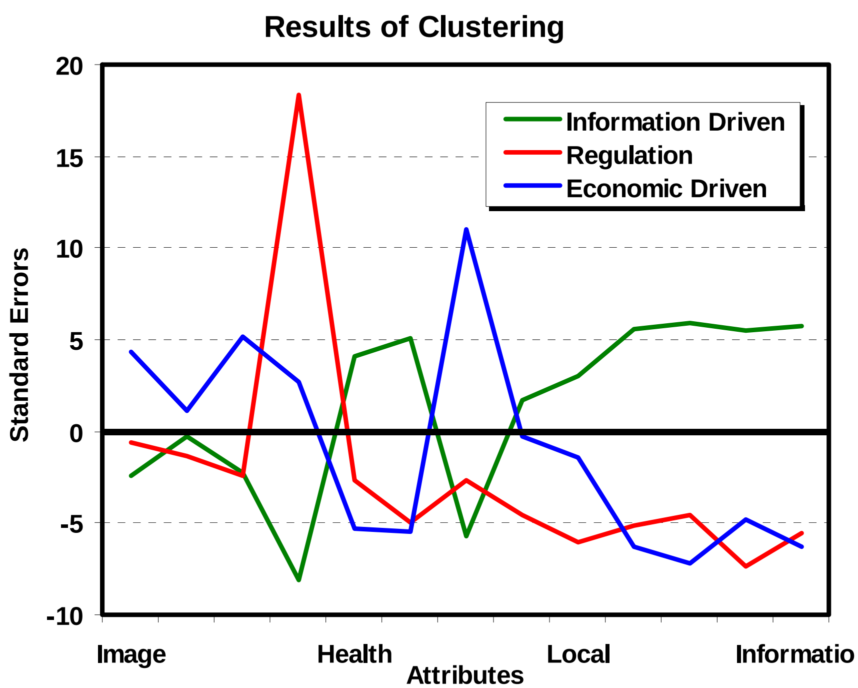
[49] We have found the best results to be using Euclidean distances and Wards or Complete Linkage gives the best results.

be seen between the 2 and 3 cluster solutions, above. In this case, we used the 3-cluster solution as the largest number still showing an orderly process.

At some point, outlier clusters develop. That can be seen with the 5-cluster solution where only two cases are included in cluster 5. This may or may not indicate that we have gone too far. Some outliers can be considered artifacts of the surveying process, particularly when the show up early in the cascade. However, when they show up late as in this case, it is indicating that small differences are generating the clusters and therefore, too many clusters are being used.

3.7.6.6 Profile Analysis

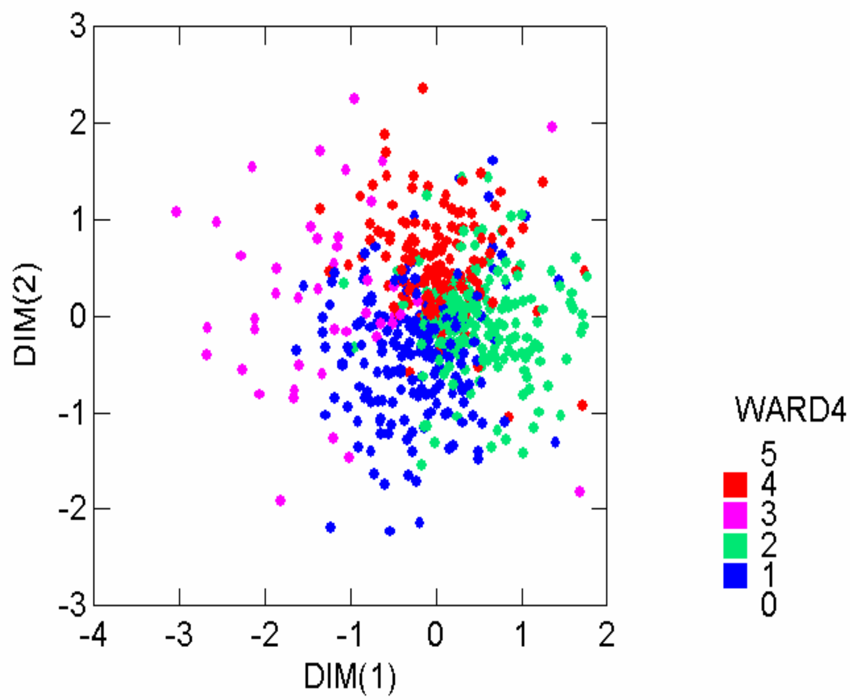
We use **standardized average deviations** to identify those segments that are distinct rather than just a separation of levels. This involved standardizing average characteristic values against the total response using standard errors based on the cluster sizes. In this case, we found that the fourth cluster was merely a higher level of a previous cluster. As such, we used the three cluster solution as the best description of the sample.



3.7.6.7 Position Maps

Position maps are used to display the distribution of respondents within and between clusters. Two types of maps are used: (1) dimension compression map and (2) reduced dimension map. Dimension compression is done by either using FACTOR analysis to Multiple Dimensional Scaling (MDS) to present the multiple variable data as a flat projection. In all cases, the compression produces an imperfect representation of the data. Each method collapses the data based on optimizing an objective function. For

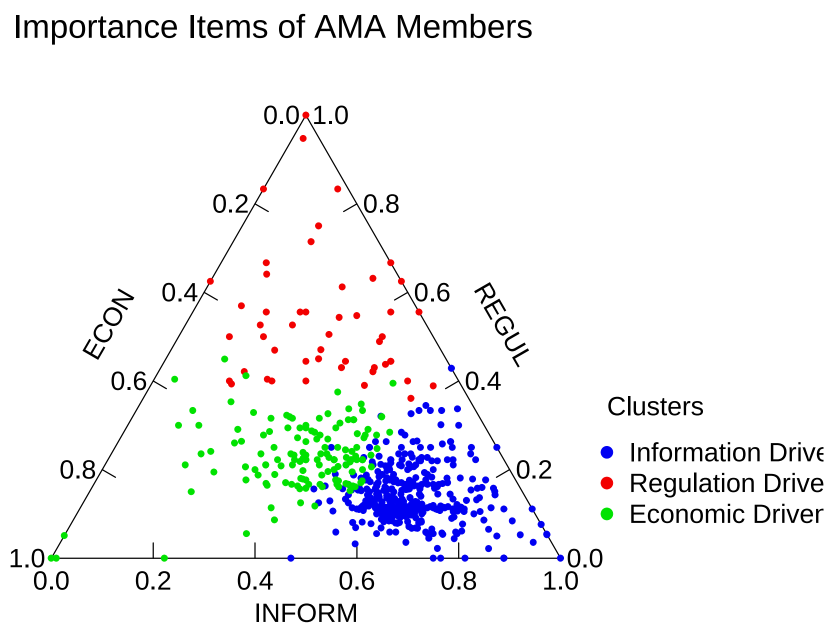
Factor analysis, that objective is to maximize the variance explained. For Multiple Dimensional Scaling, it is to minimize the error in the total distance between points. Since clustering (particularly) hierarchical clustering is based on distance, we recommend using MDS for constructing these maps[50]. The chart below shows the typical cluster position map using MDS. The apparent mixing of clusters is both due to the compression of the dimensions and the inaccuracy of the clustering process.



Reduced dimension maps involve combining variables based on the profile analysis. Principle variables for each cluster is computed and calculated in terms of a percent. This is then plotted for the three cluster solutions as triangular scatter diagrams. Four cluster solutions can also be plotted as a three dimensional triangular scatter plot. Mixing of the clusters in this case is due to the blending of cluster groups. If the clusters are not clearly identifiable it might be more useful to consider a soft clustering structure where each respondent is considered to be partial within multiple clusters.

---

[50] It should be noted, however, that MDS is limited to relatively small sample size in the same range has hierarchical clustering. For large databases, we usually use Factor Maps with K-Means clustering.



Importance Items of AMA Members
0.0
1.0
0.2
0.8
0.4
0.6
0.6
0.4
0.8
0.2
1.0
0.0
0.0
0.2
0.4
0.6
0.8
1.0
ECON
REGUL
INFORM
Clusters
Information Drive
Regulation Drive
Economic Driver

3.8 STATISTICAL MODELING

Statistical modeling is exploring the relationship between criteria and attributes. We are looking for relationships between and among the variables.

3.8.1 MARKET MODELING

The simplest market models are linear relationships between market behavior or criteria and their "predictors." These models are used to describe "good" or satisfied customers in terms of their opinions and characteristics. Typically we are interested only in the description or the "predictive power" of the model. These are "effective" models in that they describe the results but may not describe the process. The attempt at describing the process is referred to as "Key Driver Analysis" and is covered in the next section.

3.8.1.1 Regression Modeling

The statistical methods of fitting data to a model are referred to as regression techniques. These cover a broad range of procedures focusing on optimizing some "loss criteria" in the effort to fit the model parameters to the data.

3.8.1.1.1 Linear Models

The most widely used method is referred to as Multilinear (Gaussian or OLS) Regression where the single dependent criterion is fit to a linear combination of predictors (independent) variables. The process acts to maximize the variance explained (R-Square)[51]. The process assigns a coefficient to each of the "independent" variables[52]. While we usually use only metric data, categorical data can also be used by converting it to a series of zero and one valued variables (dummy variable regression).

3.8.1.1.1.1 Stepwise Regression

Not all of the variables have "significant" impact on the criteria. A series of methods referred to as "Stepwise" regression allow for the inclusion of only the significant variables[53]. This results usually in a far simple model with few parameters but with a reduced R-Square value.

---

[51] Because of changes in the number of parameters and data (degrees of freedom) the R-Square may overstate the fit of the model to the data. These valued can be modified producing an Adjusted R-Square takes these factors into account.

[52] Simple multilinear regression analysis can be done in *Microsoft Excel* as well as the standard statistical packages such as *SPSS, SAS* or *SYSTAT*.

[53]Stepwise regression is available on the standard statistical packages. The methods include forward and backward approaches as well as changes in the criteria for inclusion.

3.8.1.1.1.2 Rationalization

In many cases, we can assume that the coefficients in the market model have specific signs, they should be either positive or negative. If this is the case, variables with inappropriate signed coefficient are often dropped. This “irrationality” can be caused by intercorrelation or the existence of excluded variables[54]. The results of dropping significant variables, however, are both a reduction in the fit of the criteria and potentially missing some important issues.

3.8.1.1.2 Factor (Latent Variable) Regression Models

Intercorrelation among variables can make the coefficients unreliable. It may be, therefore, useful to first factor analyze the predictive (manifest) variables to identify “latent” ones, which are independent linear combinations of the original variables. Interpreting these latent variables is an “art form” but often produces insight into the actual drivers in the market. This process is used extensively with Key Driver analysis, discussed later.

3.8.1.1.3 Non-linear Models

Because of the noise in the data as well as the lack of theoretic bases for describing the mechanisms of market behavior most of the models we use are linear. Occasionally interactive and non-linear models are constructed. Many of these can be rearranged to be in a linear form for fitting using conventional regression. However, in some cases, the fit is done using non-linear curve fitting techniques[55]. This ability can greatly enhances the modeling capability.

3.8.1.2 Discriminate Analysis Modeling

Discriminate analysis involves creating models that splits the population by some criteria. While fairly complex criteria can be used, typically, this involves a single categorical variable. Traditional methods of discrimination such as in the formation of Discriminate Trees, develops models that either predict positive or negative behavior. This can confound the interpretation of the results. An alternative is using “Stochastic” regression.

3.8.1.2.1 Stochastic (Logit) Regression

Stochastic regression, of which Logit Regression is the most widely used, is designed to estimate the likelihood of an event. They are a form of non-linear regression where a

[54] Excluding price in a purchase prediction model can produce this type of effect.

[55] It should be noted, that introducing loss functions other than least squares requires non-linear regression procedures even though the model may be linear.

non-linear function[56] of the linear combination of predictor variables is fit against a discrete (zero or one) valued dependent variable. This allows for categorical dependent variables. Since the interpretation of these models is straight forward is has become the preferred method of modeling categorical data and for discriminate analysis.

3.8.1.3 Hierarchical Bayesian Analysis

Hierarchical Bayesian Analysis has been previously mentioned in regards to an alternative method for identifying market segments where there is significant missing data. As noted then, the Hierarchical Bayesian approach is based on computing a variation from a base case situation. Traditionally, however, that base case is a prior model. The objective of Bayesian analysis is to seek a modified model based on the prior experience and the new data. This technique focuses on the distribution of the parameters of the models and how they change from the prior estimates. Hierarchical Bayesian analysis is particularly useful with periodic data where improved apparent precision is feasible.

3.8.2 KEY DRIVER (PATH) ANALYSIS

The objective of Key Driver Analysis is to model that relates overall criteria, which may be overall satisfaction or share, with the perceptions and characteristics of the offering or organizational attributes. This type of modeling goes beyond the formation of the statistical relationships but seeks to clarify the major drivers and what affects them. The goal is insight into the relationships. We seek what is driving the overall behavior and attitudes. However, once again, it must be noted that the analysis is only on the variation of responses. We are trying to identify the key discriminators that explain the variation in criteria. Attributing the key discriminators to causes of the variation in criteria is a non-statistical leap of faith. We are assuming that those factors that drive discrimination also drive the decision process.

At this point we have to recognize that this goal exceeds the capability of statistics. Statistics, as noted above, focuses on relationships in the data. It is concerned with "statistical significant relationships". On the other hand, we are interested in the importance of attributes and in simplicity of the model.

3.8.3 CAUSATION AND CORRELATION

It almost goes without saying that, statistical model only reveals relationships among data, but not the cause of those relationships. There is no way of determining from the data alone the "correct" model or the "correct" interpretation. The statistical techniques

[56] The non-linear function is the inverse a standard cumulative probability distribution such as the normal distribution. This type of regression using the normal distribution is referred to as Probit Regression which is fairly complex due to the nature of that distribution. A simpler, logistic function, is often used with has the same characteristics of the normal. Curve fitting using this function is referred to as Logit Regression.

used here are only based on the correlations among variables. They do not rely on the fundamental nature of the data.

3.8.4 DATA SCALING

Typically the attribute evaluation data available can be considered, for the purposes of modeling, interval scaled. This allows a broad range of multivariate techniques to be applied[57].

3.8.5 MISSING DATA

While some regression procedures can handle missing data, typically only respondents with full datasets are used. This allows for unambiguous measures of goodness-of-fit[58].

3.8.6 UNDERLYING ASSUMPTIONS

In order to seek out the "best" model, we will utilize a number of assumptions including:

3.8.6.1 Goodness of Fit

All other things being equal, the statistical goodness-of-fit is the appropriate measure of quality of predictability and is used as the primary criteria for modeling. There are two measures of goodness-of-fit that are use for this type of analysis: (1) R-Squared and (2) Multiple R.

3.8.6.1.1 R-Square

The R-Squared is the variance explained by the model. It is a natural measure of fit for ordinary least squares curve fitting procedures including multi-linear regression. This is used as the standard for comparing models.

3.8.6.1.2 Multiple R

The R-Squared understates the variability that is captured by the model. The Multiple-R is the square root of the R-Squared and is roughly equal to the standard deviation explained. It is a better description of the total influence of the attributes on the criteria. As such, it is used to compute the partial influences of the attributes and latent variables.

3.8.6.1.3 Adjusted R-Square

---

[57] In this chapter we are concerned only with attribute ratings, attitude evaluation, and importance scales for general modeling. However, discrete criteria are also discussed. The general topic of the analysis of nominal/categorical and tabular data is not covered here.

[58] Some regression procedures as well as the traditional Path Analysis tools (Ramon and Listrel) use the correlation matrix rather than the data directly. However, the R-Squared measure requires an effective sample size. This makes the goodness-of-fit ambiguous.

As the number of variables approaches the number of respondent cases, the fit becomes increasing good and the R-Square will indicate a perfect fit. This is a result of the nature of the regression process. The adjusted R-Square takes this loss in "degrees of freedom" into account and gives a more realistic estimate of fit.

3.8.6.2 Simplicity

The simpler the model the better! This is Occam's Razor. It is the belief that nature is fundamentally simple. A few things dominate.

3.8.6.3 Topologies

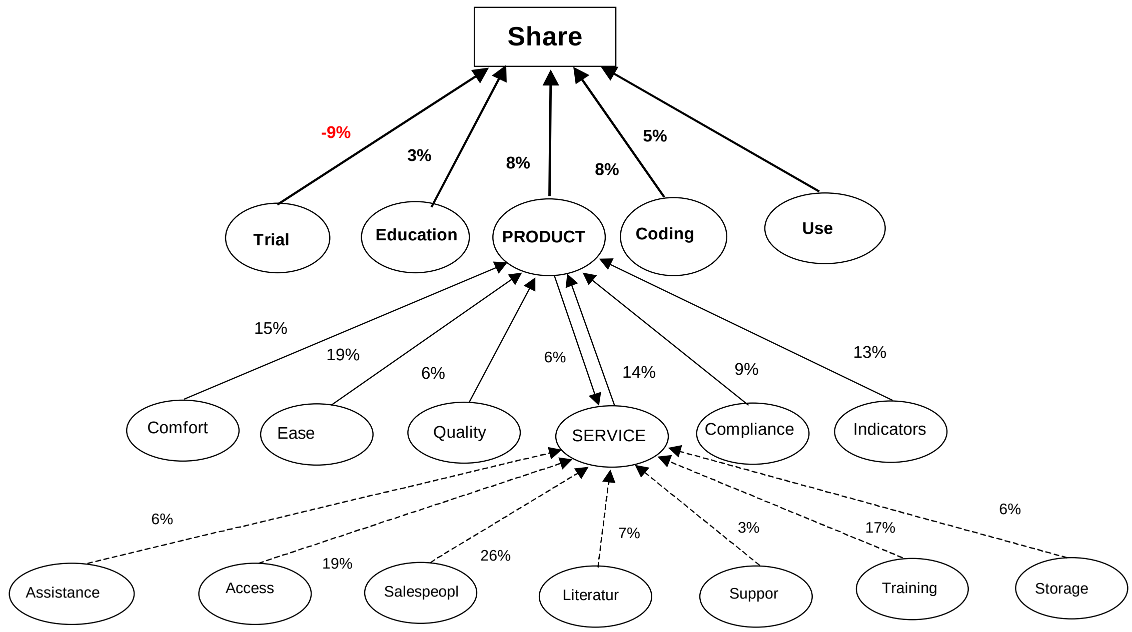
We assume that the variables are interrelated and that the model takes on particular forms.

3.8.6.3.1 Latent Variables

Some groups of attributes are highly intercorrelated and seem to act as a single variable. This may be due to either a "halo" effect or may be derived from a more basic attribute. These underlying attributes are referred to as "latent variables."

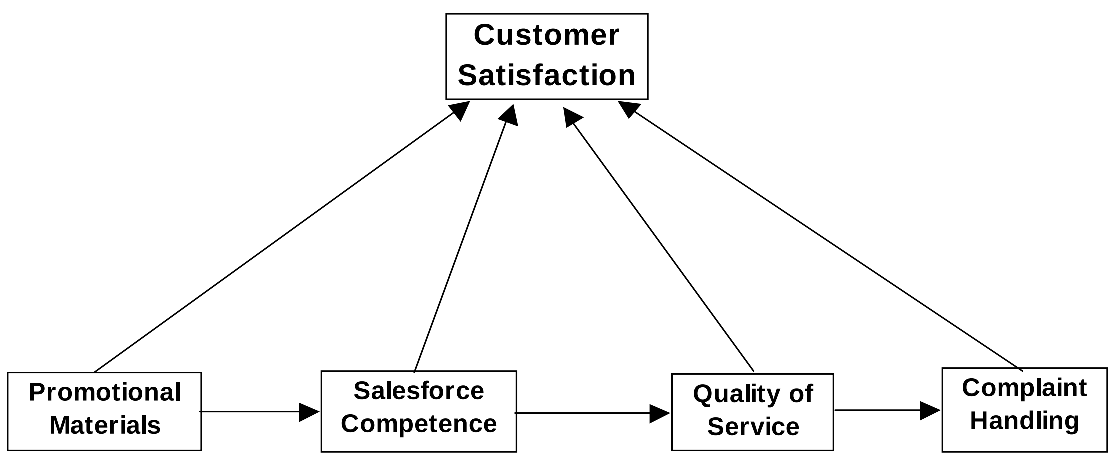
3.8.6.3.2 Hierarchical Structured

There are two types of structures or topologies that we typically expect. The hierarchical structure is ranking of attributes where one set variables is subordinated to others. This is shown on the figure below. This shows three levels of attributes. Note that this can be either explicit attributes or latent variables. In this case, the lowest level attributes only influence share through two levels of other attributes. This is the standard topology that we use for developing key driver models.



3.8.6.3.3 Cascade Structure

The hierarchical structure is primarily a linear process where the major effect is driving toward changes in attitudes. However, the other common structure, the cascade, has a strong interactive component. While several attributes influence attitudes, they interact together. In the structure below, we see a typical case where satisfaction with each attribute depends on the satisfaction with the proceeding one. While this is an interesting structure, it is not typical of most key driver models. However, occasionally it does appear.



3.8.6.3.4 Recursions

The special tools for path analysis allow for complex interaction including what are referred to as recursions. These involve multiple paths that an attribute may directly take to influence the attitude. These models tend to be unstable and are usually not considered for this type for market key driver analysis[59].

3.8.6.4 Dominate Paths

Because of intercorrelation, there is often multiple indirect paths that allow attributes to effect attitude. We assume that the model that indicates the largest (or dominate) importance for the attribute (or latent variable) is best. For example, given two models with similar "goodness-of-fit", one with many direct variables, and one with only a few, the one with fewer variables is always considered better. This is even the case when the introduction of the other variables is "statistically significant."

3.8.7 TRADITIONAL TOOLS

The following are tools that can be used to develop Key Driver Models (Path Models). Not all are used for every case and some only rarely.

3.8.7.1 Ramon and Listrel

Ramon and Listrel are the traditional linear path analysis tools. They are a broad family of tools designed to test models. They are considered "confirmatory" tools in that they do not determine the best model, but only the best coefficients for the model. Most of the unique capabilities of these tools, recursive modeling and handling missing data, are not needed for these applications. Furthermore, neither technique allows for use of non-linear procedures such as Logit regression that are often required for key driver analysis[60]. It should be noted that path analysis procedures use the correlation matrix rather than the actual data for its computation.

3.8.7.2 Correlation Analysis

The mapping of inter-correlations, using Multiple Dimensional Scaling (MDS), is used as an exploratory tool to examine potential relationships. The closer attributes and criteria are positioned the greater their association.

3.8.7.3 Factor Analysis

Factor Analysis is used to identify latent variables that underlie the attributes. This is, at least, a three-stage process:

---

[59] There are additional problems using recursive models including that most are not solvable. Recursive modeling requires simultaneous solutions of regression models that leads to their instability.

[60] This eliminates the problems of missing data since "pairwise" computation of correlations can be used. However, it produces an ambiguous R-Square, which requires an effective number of respondents.

(1) Non-rotated Factor Analysis of the data set to identify the number of potential latent variables;

(2) Rotated (Varimax) Factor Analysis of the dataset specifying at least as mean Factors as the number of potential latent variables, and

(3) Factor Analysis of the each latent variable-set to compute effective factor scores representing the latent variable values.

3.8.7.4 Canonical Correlation

Canonical Correlation is used to determine if it is feasible to merge criteria variables. The objective is to obtain the best linear (additive) combination criteria variables to give a single descriptive model.

This process is similar to the use of Factor Analysis but deals with both criteria and attributes. The process produces a series of models describing linear combinations of the criteria variables ("dependent") by linear combinations of describing variables ("independent"). These solutions are orthogonal to each other similar to the independent latent variable factors using traditional Factor Analysis. As in Factor Analysis the solutions are in decreasing amounts of captured variation.

3.8.7.5 Regression Analysis

Regression analysis is the primary statistical modeling tool for Key Driver Analysis. Usually multilinear regression is used to estimate the impact of attributes on criteria with quantitative data. When discrete criterion (dependent variable) is used, such as with customers versus non-customers, stochastic regression (Logit) is appropriate[61]. Stepwise Regression is used to determine important contributing variables in the models[62].

3.8.8 AN EXPLORATORY APPROACH

Key Driver Analysis is a form of "path analysis" also referred to as structure and casual modeling. We seek the simplest (non-recursive) regression structure that describes the overall satisfaction or likelihood to purchase. This involves identifying latent (intercorrelated) variables as well as identifying topology of the drivers. The following are the steps followed:

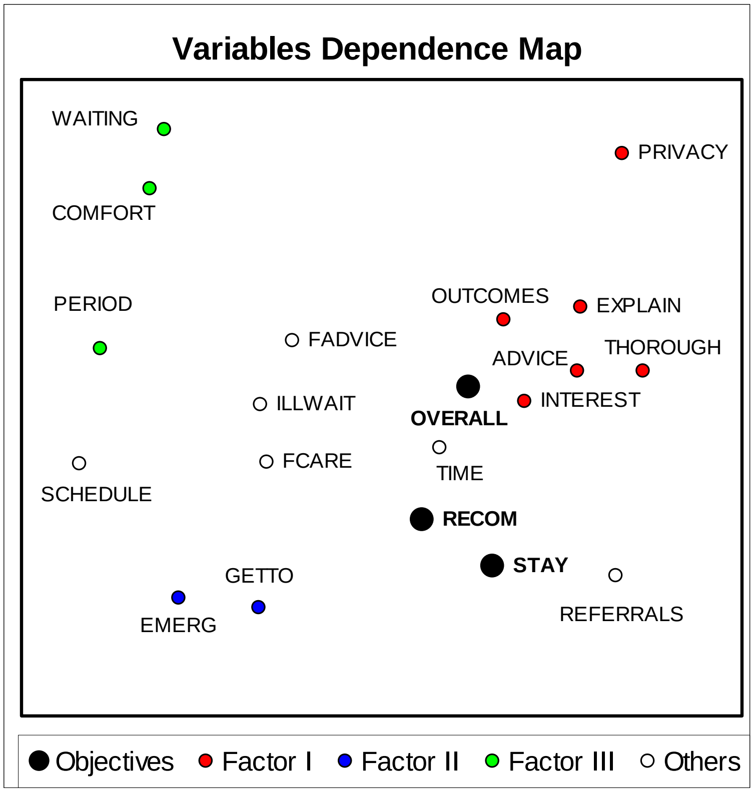
3.8.8.1 Correlation Maps

Develop the correlation map based on the multiple dimensional scaling of the correlation

[61] Logit Regression as well as other non-linear types are not available with the standard path analysis tools such as Listrel.

[62] There are several options in using Stepwise procedures. They produce different models and it is often necessary to use several to understand the feasible model variations.

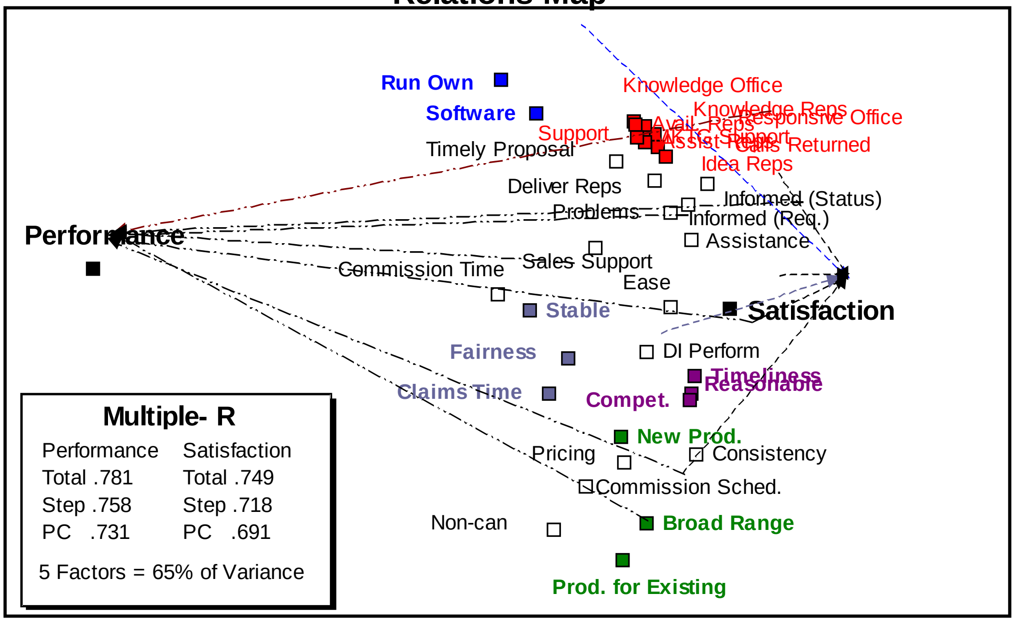
coefficients. This map displays the interrelationships among variables.



3.8.8.2 Multilinear Regression

Identify the highest degree of fit using multilinear regression.

Relations Map



3.8.8.3 Merge Dependent Variables by Canonical Correlation

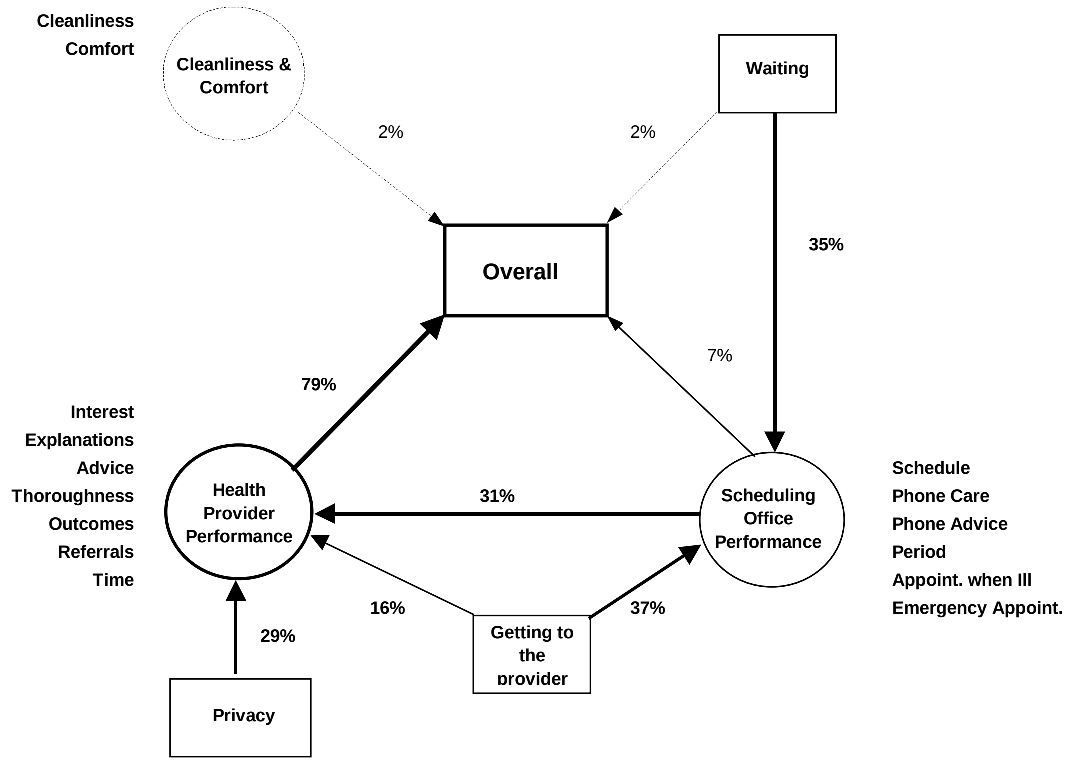
Identify the need to used separate or independent models using canonical correlation. This technique merges factor analysis and regression to develop the "best" merged linear model that predicts linear combinations of the dependent variables by orthogonal sets of "independent variables."

3.8.8.4 Identify Latent Variables by Factor Analysis

Identify latent variables through factor analysis (Varimax rotated factors) -- We used factor analysis to identify the key intercorrelated characteristics. We then factor analyzed the subsets of the key characteristics to define the latent variables.

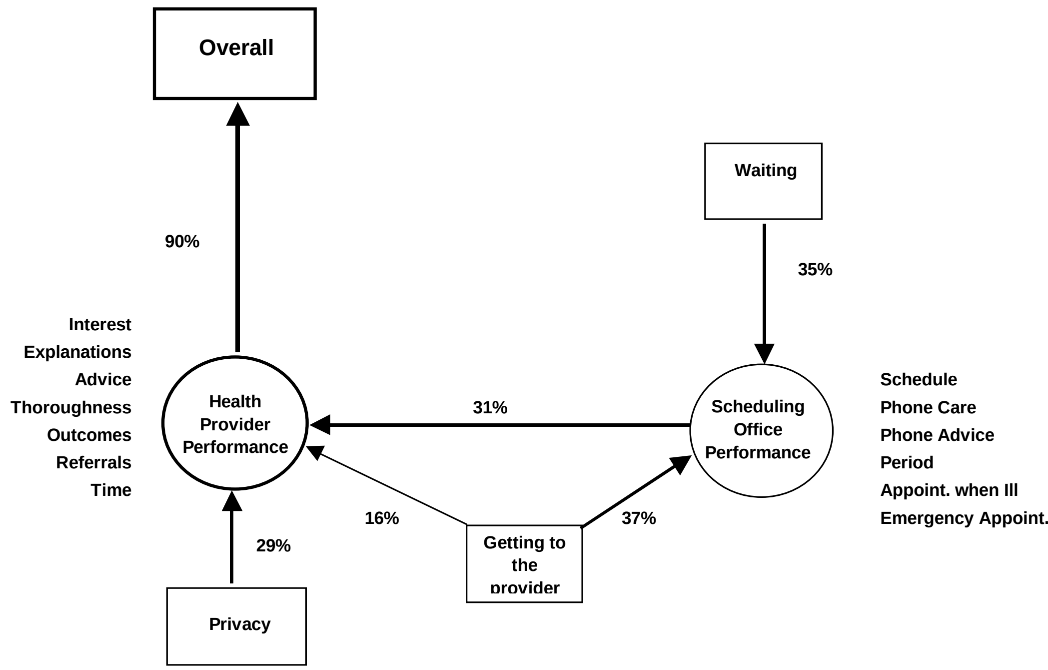
3.8.8.5 Construct First Order Models with Stepwise Regression

Key drivers and first order models are identified using stepwise multilinear regression (backward and forward) of the remaining characteristics and latent variables.



3.8.8.6 Reduce the Model

Models are reduced by testing the impact of removing the less dependent variables in the hierarchical or cascaded structure. This is a heuristic process of testing the impact of removing variables. Due to intercorrelation, the impact of a variable may enter directly or indirectly through key drivers. In general, maintaining the variable gives improved prediction but with increased complexity. Judgment has to be applied to determine if the gain in fit is worth the loss of simplicity in structure.



3.8.8.7 Compute Key Driver Influences

The actual attribute influence, percentage explained, is computed based on the standardized regression coefficients and the multiple R values. The standardize regression coefficients are first normalized and used as weights against the multiple R. For negative coefficients, the absolute value is used to compute percent variation explained.

3.8.9 NON-LINEAR BENEFITS

Estimates of attribute value are used to direct the allocation of resources for improving the market position and determining the marketing mix. These estimates focus on the value of attribute categories rather than specific proposals. They are typically single, "linear" estimates that are used in quadrant analysis or in the estimation of overall utility of products and brands as previously mentioned. These estimates are obtained either directly from respondents in the form of stated importance or the perceived value of proposed new features, or as aggregated market value through linear regression of product attribute performance estimates in the form of key driver analyses.

However, these estimates can be misleading in that they focus on the value of attributes across products and levels and do not take into consideration any effect of the changing marginal utility of those attributes. This non-linearity in value can effect both the attitude toward existing features as well as the potential for improvements.

3.8.9.1 Attributes, Features and Benefits

It is important to distinguish between offering attributes, features (or levels of attributes) and benefits. Attributes are categories of possible features. The features of the products could be viewed in this context as levels of the attributes. These attributes include both explicit and cognitive items such as size or function, or they may be implicit items including brand and image. Attributes for the most part are characteristics as viewed from the sellers' and manufacturers' perspectives. They compose the products as they are offered. Benefits on the other hand are from the buyers' and users'. They compose the value to the user. To some extent, features transcend these two groups. They may be perceived both as characteristics of the product by the seller and as benefits to the user.

As we discuss non-linear value, we have to understand that we are working really on the user perspective. We need to focus on benefits and possibly on features. Unfortunately most of the data available is on the attributes. As we will discuss later, we use latent variable analysis partially to define benefits in terms of attributes as a means of connecting the two.

3.8.9.2 Suggested Alternative Approaches

The "Kano" method attempts to capture this non-linearity in regards to current performance and future improvements by measuring the response of potential buyers to changes in the existing situation. They define, "Must Have's" as items that have high negative impact if lost, but small or no impact if improved. Unfortunately, this definition of "Must Have's" does not reflect actual buyer behavior where improvements of demanded features can also be positive[63]. The Kano procedure while interesting tends to be expensive and does not distinguish well among attributes.

Dapeng Cui[64] argued for the use of non-linear models in order to define the potential of the attributes. As a means of obtaining non-linear functions he suggested the use of the power-law or multiplicative model (what is referred to in this paper as the Cobb Douglas function). That is that the total value of an offering is the product of quantity of each attribute raised to a power:

$$\text{Value} = B_0 \bullet X_1^{A_1} \bullet X_2^{A_2} \bullet \bullet X_n^{A_n}$$

This can be put into a linear regression form by applying logarithms as:

$$\text{Ln(Value)} = \text{Ln}(B_0) + A_1\text{Ln}(X_1) + A_2\text{Ln}(X_2) + + A_n\text{Ln}(X_n)$$

---

[63] Typically we use a separate profiling exercise on specific propose features to seek out the "Must Have's." Even combined with attribute ratings and some perceived value exercises, it is simpler than the traditional Kano exercise and far more effective.

[64] Dapeng Cui, "Uncovering What Drives Product Performance: An Ipsos-Insight Approach to Driver Analysis", Promotional white paper by Ipsos-Insight (2004)

However, the use of this multiplicative model for market value is problematic. The author is using it mainly to capture the asymmetric effect but then confounds the exponents with importance. Since the ratings typically have only a narrow range, this regression model should not give significantly different results than using linear regression[65].

An additive non-linear value function would be more general:

$$\textbf{Value} = B_0 + B_1X_1^{A_1} + B_2X_2^{A_2} + + B_nX_n^{A_n}$$

There are two problems with this form: (1) the increased number of parameters and (2) the inability to use linear regression with its additional tools and variations. The increased number of parameters could be a problem with small datasets. As we will discuss later, the identification of latent variables and the use of linear stepwise regression reduces the number of variables being analyzed and therefore reduces the number of parameters. Furthermore, the database that is usually used for this type of analysis is sufficiently large to handle the problem.

The need to use non-linear regression is due to the combination of exponential and linear parameters in the model. Non-linear regression is available using standard statistical packages (*SPSS*, *SAS*, or *Systat*). But unfortunately, they often don't allow for this open form. However, the models can be fit using *Solver* in *Microsoft Excel*. This has been our preferred method because it allows for the use of a range of "loss" functions including "least squares[66]" as well as imposing only positive values of the parameters. An example of the results is shown below:

$$\textbf{Value} = -4.93 + 1.12\,\textbf{Suppler}^{0.53} + 0.63\,\textbf{Value}^{1.11} + 0.07\,\textbf{Effective}^{1.63}$$

An alternative approach is to transform the variables prior to the fitting process. This would allow the use of linear regression. The process requires seeking an exponent value that is appropriate for each variable. The traditional Box Cox method of doing this is to seek the exponent that gives the best correlation of value. The transformation that is used is:

---

[65] One reason for Cui suggesting this model is the potential for the use of "Ridge Regression". Ridge Regression can reduce the problems of handle colinearity. Unfortunately use of Ridge Regression is very dubious. It is as much of an art form as a structured procedure. It is only appropriate for very modest intercorrelation, in the order of 20 to 30 percent. While it is an interesting procedure, recent statistical studies have indicated that there is little negative impact with correlations in this range with standard (least squares) linear regression anyway. The real problem is high levels of intercorrelation, which in the case of attributes is normally interpret as resulting from underlying latent variables which drive the behavioral process.

[66] Least Squares fitting also acts to maximize the "R-Squared" measures of fit. Other loss functions such as Maximum Likelihood could also be used but typically produces similar results.

$$Y_i = (X_i^{\lambda i} - 1)/\lambda_i$$

Unfortunately the fitted exponent using this procedure is not in good agreement with the general non-linear regression results. As such, this procedure is not recommended.

3.8.9.3 Latent Variables

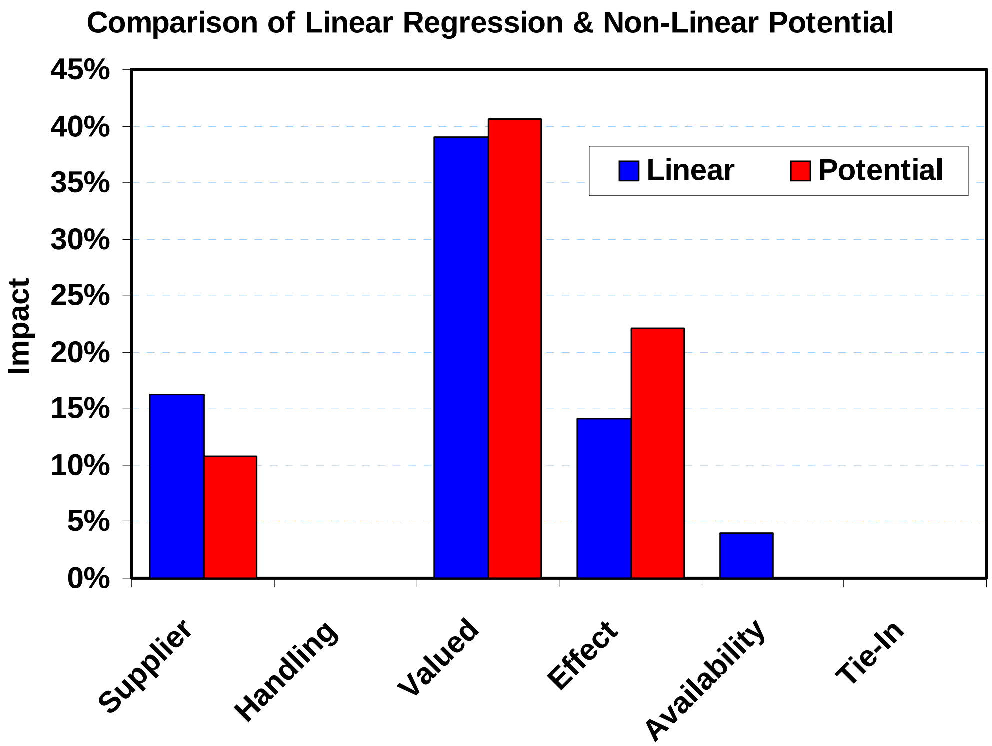
As previously mentioned, latent variables represent, what is believed to be the underlying drivers in the market. They are combinations of the attributes into groups that are highly intercorrelated or clustered together. We use factor and cluster analyses to identify them. Ideally, the resulting latent variables would be independent of each other, orthogonal. However, this result of factor analysis results in mixed combinations of all variables. In order to get more meaningful set of latent variable a sub-set is generated. These include only the specific variables of interest. The resulting latent variables, however, may be intercorrelated but to a far lower extent than the original attributes. This process usually reduces the number of variables from 20 to 40 to less than nine[67].

3.8.9.4 Importance and Potential

There are several measures of importance and potential that are useful for planning purposes. The linear model can be considered a sub-set of the non-linear form with unity exponents. As with key driver analysis, the coefficients are related to the importance of each variable. In terms of linear importance, these measures will vary depending on the level of the attribute. The linear importance would represent a typical value over the range of data available. The importance at the current level of performance would be the measure of interest looking at marginal improvements or declines in performance. For the most part, this measure of non-linear importance should be similar to the results obtained using the linear measure.

The potential is the improvement in value that can be expected with the maximum increase in performance. It is depends on the current performance, the maximum reasonably achievable and the conversion of that performance to value. With the linear model, this last factor is determined solely by the linear importance. However, some variables show decreasing marginal utility. That is the marginal improvements decrease as the improvements proceed. Benefits from suppliers tend to be of this form. Others may actually increase greater than unity indicating acceleration in value with increased performance.

[67] A traditional rule of decision making is that decision-makers can handle no more than 7 ± 2 variables in the process. This leads to the rule of 5 to 9 variables. However, this is also much larger than we typically find necessary. Typically stepwise linear regression is used to reduce the number to a minimum that captures the meaningful variation.



3.8.9.5 Interaction

So far we have dealt with models that excluded interaction among the variables as they drive value. Interaction here is not necessarily intercorrelation. Variables may be independent of each other in a correlation sense but still interact in ways the generate value. Because the number of parameters increases rapidly when including interactions, we typically only analyze them in pairs. This may be done using Response Surface analysis. This is based on general quadratic relationships:

$$\textbf{Partial Value} = B_o + B_1X_1 + B_2X_2 + B_3X_1^2 + B_4X_2^2 + B_5X_1X_2$$

This then plotted as a surface to reveal the nature of the interactions. However, this is typically not done unless important interactions are suspected.

3.9 MARKET SIMULATORS

The market simulators and models using attribution evaluation data are designed to estimate the impact of changes in customer/employee perception on their attitudes. The "What-If" question focuses on the impact of marketing and operational programs on the market attitude and potential market share. The inputs into the simulator are expected changes in perception and the output is the resulting market changes. The purpose of these simulators is to provide insight into the importance of changing perceptions.

The simulators are generally given as personal computer decision support systems. Today, these are most often produced as user friendly *Microsoft Excel* spreadsheets.

3.9.1 LIMITATIONS

The simulators are based on current rating and importance of attributes of the existing products. The simulators are trying to predict the impact of changes that have not been measured. This produces three fundamental issues.

3.9.1.1 Extending Beyond the Limits

The simulator is based on relationships confined to the range of the data collected. However, the simulator is intended to predict conditions beyond the scope of the data. We are asking what happens if we make our customers extremely happy while the competitors do not. The relationships that are derived from the data do not have these conditions. The market simulations are therefore, extrapolations of the data[68].

3.9.1.2 Modeling Variation, Not Action

Many of the simulators are based on regression models. Regression estimates what is the impact on criteria variation due to variability in other variables. However, these are not causal relationships. They are descriptive of the present situation[69]. In the simulators, however, we assume that the relationships are predictive of attitude changes and are ,therefore, causal.

[68] This is not the case for either pricing or perceived value market simulators. In these cases the customers' response has been measured for these new conditions.

[69] The regression models for the pricing and perceived value market simulators are by design predictive of decisions.

3.9.1.3 Making the Perception Change Happen

As previously noted, these simulators are design to predict the impact of a change in perception. These changes are assumed to be exogenous; they "just" happen. There is no connection within the simulator between programs and these changes. There are viewed as goals. The simulators therefore, test the impact of goals in changing perception on attitudes.

3.9.2 BUILDING THE MARKET SIMULATOR

The market simulators are developed by combining the predicted action of each of the respondents. This produces a two stage process: (1) evaluating the decision process of the respondents and (2) combining them to get an estimate of the total market behavior.

3.9.3 RESPONDENT DECISION MODEL

The purpose of value modeling is to obtain estimates of the impact of changes in perception on an individual's attitude. It is the process of merging the information that the respondent has given us to get a composite reaction. The models are divided into three classes: (1) statistical or regression models, (2) utility modeling, and (3) hybrid modeling. Statistical and hybrid modeling require an overall measure of "goodness" that is referred to as the criteria. It might be the likelihood of purchase, an overall satisfaction or a measure of loyalty. The goal here is to predict a change in this criterion with changes in perception. Value modeling focuses on changes in an overall measure of value with changes in perception.

3.9.3.1 Multilinear Regression Model

The simplest statistical model is the result of multilinear regression. We assume that sensitivity of the respondent will be the same as that captured by variability of the sample. That is a change in a criterion will follow the regression model from the data. This is expressed as:

$$\textbf{Individual}_k \textbf{ Value for Product}_i = \sum_{j=1}^{N} (\textbf{Rating}_{jik}) \bullet (\textbf{Regression Coefficient}_j)$$

where the results are for the k^{th} respondent's overall opinion of the i^{th} product based on ratings on all attributes. The regression coefficients are based on a multilinear regression predicting the criteria based on ratings of attributes.

Problems and Issues:

While this is a fairly straight forward approach, there are a number of severe problems.

These include the assumptions that:

- The ratings of the attributes by the respondents are independent.
- The individual will behave in a manner consistent with population.
- Modeling the variation in responses captures the potential reaction to changes beyond the scope of the captured data.

It should be noted that the Goodness-of-Fit (R-Squared) of these models can be very poor.

It should be noted that the measures of Goodness-of-Fit (R-Square) is strongly dependent on the number of data points and variable used. As previously noted, if the number of variables approaches that of the number of points the "adjusted R-Square" must be used.

3.9.3.2 Latent Variable Regression Model

The problem of independence of attributes can be severe. The Path models often show high degrees of interdependence making the simple multilinear regression model inappropriate. An alternative is to use a latent variable (principle component or factor) regression model. This uses factor analysis to identify high intercorrelated variables to form relatively independent factors. A regression model is then formed using these factors for the criteria. The resulting model is of the form:

$$\textbf{Individual}_{\mathbf{K}}\ \textbf{Value for Product}_{\mathbf{i}} = \sum_{l=1}^{N} (\textbf{Latent Variable Factor Score}_{\mathbf{lik}}) \bullet (\textbf{Regression Coefficient}_{\mathbf{l}})$$

Problems and Issues:

While this form does not have the intercorrelation problem of the simple multilinear regression approach it is far less useful. The resulting model does not indicate the impact of changes in attributes but rather in groups of attributes[70]. In addition, we still have most of the original problems and another:

- The individual will behave in a manner consistent with population.
- Modeling the variation in responses captures the potential reaction to changes beyond the scope of the captured data.

---

[70] Some investigators use the factor loadings to extend the latent variable regression to predict the impact of variables. Unfortunately, these tend to produce an approximation of the simple multilinear regression model.

- Goodness-of-Fit (R-Squared) will be poorer than the simple multilinear regression model that itself may not be particularly good.

3.9.3.3 Utility Models

The purpose of utility modeling is to obtain a composite value for both the perception and the preference (importance) information. We assume a simple linear combination of these attributes[71]. That is we assume that partial value for a respondent of an attribute is the product of its rating and its stated importance.

$$\textbf{Individual}_{\mathbf{k}}\ \textbf{Value for Product}_{\mathbf{i}} = \sum_{j=1}^{N} (\textbf{Rating}_{\mathbf{jik}}) \bullet (\textbf{Importance}_{\mathbf{jk}})$$

It should be noted that, unlike the statistical models, this is on an individual basis. Only the respondent's individual data is used in the computation. We do not assume that the respondent will react like the market.

Problems and Issues:

Unfortunately there are some severe problems including:

- The resulting value is not ratio (or interval) scaled. It is best described as an ordinal scale measure of preference.
- It assumes that the stated importance really dictates actual behavior.

3.9.3.3.1 Price Utility Scaling

The utility has been scaled against price if a number of standard products are available. This is done based on curve fitting the computed utilities of existing products against their market prices. This allows for price comparisons using attribute rating data[72].

Problems and Issues:

This method is not recommended because of a number very significant problems including:

- As noted above, utilities are ordinal scaled. Yet the scaling is appropriate only for ratio scaled data.
- Prices are based on what is offered by suppliers rather than what is desired.

[71] There are any number of people who have claims to the model's origin and it is referred to as Fishbein model.

[72] This is sometimes referred to as part of a value mapping process.

This makes separate demand from the market offerings.

3.9.3.4 Hybrid Model

The fundamental problem of the latent variable model is that there is no unique statistical method of assign values of the attributes to the latent variables. The latent variables are combinations of intercorrelated attribute ratings. Because of the intercorrelation, there is no unique way to define their weights. The hybrid model uses importance weights to define the latent variables. In this way, the result is a model for a criterion based on the original attribute ratings. It should be noted that this redefines the latent variables and therefore, it is necessary to recompute the resulting regression model. This is the preferred method of value model for criteria if there are high intercorrelations.

The advantages of this procedure are:

- The model captures the impact of changes in attribute ratings on changes in attitude and
- The model contains the individuals' variability in importance.

Problems and Issues:

However, there are still some problems using this approach.

- Modeling the variation in responses predicts the potential reaction to changes beyond the scope of the captured data.
- Goodness-of-Fit (R-Squared) is even poorer than the latent variable regression model.

3.9.4 MARKET MODELS

The purpose of the market model in the simulator is to merge the individual predictions to produce an estimate for the total market. There are two types of predictions: (1) changes in overall attitudes and (2) market share.

3.9.4.1 Weighting Responses

In both cases, a major issue is that of weighting the results or using unweighted values. Typically the choice is based on the variation in the importance or size of the respondents and the order and randomness of the sample. If sampling has been highly stratified, weighting is likely to be required. Alternatively, if defined segment analysis is needed, the respondents from each segment will be segregated and modeled separately. Under this condition it is usually not necessary to weight the results[73].

---

[73]In any case, weighting is not preferred since it reduces the effective sample size and decreases effective

3.9.4.2 Changes in Overall Attitudes

The overall attitudes that are of interest usually consist of measures of overall satisfaction, loyalty and intent to purchase. Most of these overall attitudes are captured in the survey data and predicted on an individual basis.

3.9.4.2.1 Single Variable Criteria

In many cases we are interested in predicting the change in a single defined variable such as overall satisfaction, interest or intent to purchase. Typically we compute either an average value or more preferred the percent dissatisfied. The fractional measure is usually preferred since the underlying predictors are usually ordinal in nature and average values are not particularly insightful.

3.9.4.2.2 Derived Criteria

A more complicated problem exists when we are looking for some derived measure of attitude such as loyalty for which there is no one single criteria that has been measured. In this case, we seek a combination of criteria, a latent criterion, which is associated with the issue of interest. This requires returning to the respondent decision modeling. Fortunately, we normally explore this issue during the key driver analysis. Canonical Correlation facilitates exploring for these aggregated models. The major problem is determining cut-off points, thresholds, for evaluating "loyal" from "disloyal" respondents. The cut-off points often can be estimated by relating the derived criteria levels with purchases or likelihood to change[74].

3.9.4.3 Market Share

Estimating market share is a more complex problem. It is usually assumed that market share will be a function of the relative value computed for the product. In this situation we need comparative values for each of the competing products. In general, each individual will consider only a small number of such products (his consideration set). However, it is uncertain how relative value translates into purchases. There are at least three methods that are used: (1) Winner-Take-All, (2) Distributed Assignment, and (3) Scaled Value Share.

3.9.4.3.1 Calibration and Validation

Unlike attitude or purchase intent, market share is generally computed independent of overall market data. Therefore, the estimates of share often are very different from the values used by management. In order to provide "face validity" of the simulator, the

---

precision.

[74] Discriminate analysis and Logit regression can be used to determine the relationship between customers and the derived discrete criteria values.

model is often “calibrated” by introducing adjustable parameters that are set in such a way that initial predicted share agrees with expectations. Unfortunately, this also changes the sensitivity of the model and affects its reliability.

It is critical not to confuse calibration with validation. Agreement of the calibrated model with expectations does not imply validity. It is a forced fit. Agreement of the non-calibrated model with data would imply stronger confidence in its reliability. It should be noted, however, that true predictive validation can only be determined by comparing non-trivial predictions from the model with actual market behavior.

3.9.4.3.2 Winner-Takes-All

The simplest, and usually the most reliable approach to merge respondent predicted behavior is to assign all the sales to the product with the highest computed utility. This is clearly appropriate in cases with single purchases. However, it is more problematic when the buyer can consider multiple purchases or when the respondent is viewed as representing a group of potential customers.

Problems and Issues:

Though it tends to be the preferred method, the Winner-Takes-All scheme has a few problems.

- It does not take into consideration very close values. This can be a problem with cases where the products are perceived as almost identical.
- Small samples can produce unrealistically large changes in shares with only minor changes in perceptions.
- The only adjustable parameters are weighing factors. This makes this type of model difficult to calibrate.
- Equal product values can be problematic. Usually some rule is imposed to handle the problem.

3.9.4.3.3 Order Distributed Statistical Model

An alternative approach is to assign a share depending on the rank order of the values. Typically, we use a known linear rank order statistical distribution (referred to as the broken stick rule) for this purpose. This is often used for industrial purchases when multiple brand purchases are expected. For example, physicians will tend to use more than one product within a category. It is also used with segmentation, where respondents are considered to represent a group of customers.

Problems and Issues:

However, there are some key problems in the use of this method.

- There is no adjustment on share for close value estimates between products. Near misses are the same as large differences. The underlying assumption is that shares follow position not value level.
- Similar to the "Winner-Takes-All" case unless adjustable parameters are introduce there is no mechanism for calibration other than through weighing factors.
- Equal product values also produce problems. As in the case of "Winner-Takes-All," special rules have be introduce to handle this situation.

3.9.4.3.4 Scaled Value Share

Though not recommended, there is another approached used to estimate share based on the relative utilities. This method scales share[75] using the form:

$$\text{Market Share}_{ki} = \exp[\alpha \bullet \text{Value}_{ki}]/\exp[\alpha \bullet (\sum_{j=1}^{N} \text{Value}_{kj})/N]$$

This is the estimate for the market share for the ith product by the kth respondent. It is the ratio of the exponent[76] of the scaled value divided by the exponent of the scaled average value. The scaling factor, α, is adjusted to the best agreement with the actual market shares given the present utilities[77].

Problems and Issues:

Value share methods are filled with problems and therefore not recommended. However, they do allow for calibration.

- The model is basically arbitrary. Any form can be used which means that predictions are also somewhat arbitrary.
- Simple value share predictions are known to be poor estimators.

---

[75] There are two other methods mentioned in the literature to scale this data neither of which are satisfactory. The simple value share approach assumes that market share will follow value share. Unfortunately, the value is at best an interval scale number and therefore does not have a natural zero for forming a measure of share. Value share is therefore arbitrary. An alternative approach is using power-law model using a scaling factor as an exponential. Unfortunately here again the interval nature of the scale produces inherent problems in comparing results among respondents.

[76] The exponential form allows for the computation of a meaningful ratio scaled value from the interval scaled data. Any number added to the numerator and denominator will cancel out.

[77] The form can be extended to use more scaling factors. As many as 1 minus the number of competitors can be used. This would improve calibration of the model; however, it may not improve its predictability.

- The calibration factor often exceeds 5 using standardized product utility. This is extremely high and tends to make the share very sensitive on utility changes.

3.9.4.3.5 Uncaptured Variables

It must be noted, however, that there are other factors not included in the model that will influence share. Typically pricing, feature changes and availability are often not included in the attributes. As such, one must expect the simulation based only on changes of attribute ratings to be indicative of the market but not predictive of actual market share.

3.9.5 DEVELOPING THE SIMULATOR

The following are the major steps in developing the market simulator with attribute data.

3.9.5.1 Missing Data

Typically there can be a significant amount of inappropriately missing data[78]. This involves data missing for products within the respondent's considerations set. Either only compete datasets are used or the missing data is substituted by some standard methods of regression or using average values[79].

3.9.5.2 Respondent Level Decision Modeling

The key to the simulator is the ability to predict expected respondent attitude changes with changes in perception. This is an extrapolation process. We assume targeted changes in perception and compute the effect.

3.9.5.2.1 Constraining Variable Change

The perceptions are generally scaled with in a given range such as from 1 to 10. If the change in variable exceeds these limits, it's unacceptable. The new value can not be greater than the upper limit or less than the lower. As such, the range of change has to be constrained. The overall effect is that even though a normative change is put into the simulator, the actual average change will be somewhat less than that indicated change.

3.9.5.2.2 Capping on the Predicted Range

---

[78] The need to “correct” for missing data depends on the severity of the problem. Projects intended to produce simulators usually require respondents to complete the questionnaire and therefore have minimum missing data. Typically missing data in these cases are less than 25%.

[79] Statistical methods for handling missing data such as the EM algorithm are inappropriate here. These methods are design to retain the variability in the dataset. For simulators it is more important that the estimate of the missing data is proper than retaining the statistical characteristics of the original data.

Similarly, the resulting attitude ranges are often limited. The resulting prediction must be constrained within those bounds. For example, if overall satisfaction is scaled from 1 to 10, the resulting prediction for each respondent must be within that range even if the underlying model predicts higher or lower values. This requires capping the range.

3.9.5.3 Merging Results

To create the market simulation, the individual predictions are merged to give overall values for the market or/and selected market segments. As previously mentioned, attitude values are usually computed as percentage satisfied.

3.9.5.4 Precision Estimates

It is often useful to compute the precision of the changed estimates to determine significance. Though significance is less critical than the magnitude of the change itself it is often required. Since percentage or share measures are being used, we can assume that the estimates are binomial distributed[80].

Using *EXCEL* the following formula is used to compute the error bound:

Error Bound = BETAINV(Level,[P • N],[N - (P • N)+1],0,N)/N

where **P** is the share, **N** is the sample size, and **Level** is the confidence level[81].

3.9.6 THE DECISION SUPPORT SYSTEM

The purpose of the decision support system is to provide a friendly environment to encourage the management team to explore possible scenarios and to gain insight into the impact of various factors on market behavior. Typically users want assistance in setting priorities for improving operations.

Until recently decision support systems of this type were constructed either using a standard procedural computer language[82] to produce command user interfaces or using object oriented language such as Visual Basic or Visual C designed to provide a Windows interface. These were fairly expensive undertaking and limited to proprietary systems[83]. However, the advent of high speed personal computers and the expansion *of Microsoft EXCEL* Spreadsheet package have enabled the use of a standard open system

---

[80] This is straight forward for unweighted data and using "Winner-Takes-All" share estimates. It is more complex for weighed values or using one of the other share estimating procedures. However, even if other methods are used we typically use the standard binomial computation for simplicity.

[81] This formula uses the beta distribution that is a continuous version of the binomial and allows inversion.

[82] This included *BASIC*, *FORTRAN*, or "*C*".

[83] These were provided by marketing research and consulting firms as unique products.

for building these tools. The result is far more affordable systems that the user can modify to expand it use.

3.9.6.1 The User Interface

Microsoft EXCEL allows the modification of the decision support system in a number of ways to provide a unique “look and feel.” However, the basic elements tend to be invariant. The spreadsheet provides two areas for: (1) entering desired changes in the perception of product/firm attributes, and (2) reporting the resulting changes in overall attitude, market and utility share.

The “change entry” section contains a list of the attributes, the present average values, and a place to put in the changes. In some cases, changes are allowed for competing products and brands as well as the target product. Similarly the reporting section contains, both the computed results and initial values for comparison. These are shown for the total market and by segment. However, the size of the screen limits the effective presentation of the results. Typically, if multiple segment results are computed, they can be presented on other sheets so that the user can “drill down” to see the details. More sophisticated systems allow “hot” or “live” customizing of the reports. The user can redefine ad hoc segments based on criteria and select variables to be viewed and even models that will be used.[84].

Prior to the availability of high performance personal computers, computation speed had been a problem for these systems. However, at present, systems with as many as two to three thousand respondents are feasible.

[84] However, this is typically not required for most applications. The specific uses of the simulator are usually understood and wide flexibility not required.

Employee Statisfaction

Question	Difference	Mean
1 Work Area	0	2.65
2 Help People	0	2.79
3 Help Resources	0	2.79
4 Work Distr.	0	2.90
5 Time	0	2.84
6 Job Desc.	0	3.37
7 Empower	0	3.33
8 Difference	0	3.31
9 Reward	0	2.87
10 Potential	0	2.95
11 Advancement	0	2.65
12 Technology	0	2.71
13 Training	0	2.67
14 Dept Team	0	3.08
15 Dept Comm	0	2.90
16 Unit Team	0	2.85
17 Unit Comm	0	2.74
18 Sup FeedBack	0	2.92
19 Sup Expect	0	3.00
20 Sup Needs	0	3.26
21 Sup Avail	0	3.15
22 Sup Recog.	0	3.08
23 Mgt Access	0	2.77
24 Mgt Comm	0	2.91
25 Mgt Concern	0	3.10
26 Mgt Action	0	2.61
27 Salary	0	2.30
28 Flex-Time	0	3.53
29 Vacation	0	3.29
30 Retire	0	3.44
31 Benefits	0	3.89

TOTAL OF ALL RESPONDENTS

	% Satisfied		Average Score	
Performance	Value	Standard	Value	Standard
Recommend	65.9%	65.9%	3.70	3.70
Working	84.0%	84.0%	4.18	4.18
Again	69.5%	69.5%	3.84	3.84
Available	77.4%	77.4%	3.99	3.99
Skills	93.1%	93.1%	4.44	4.44
Used	52.4%	52.4%	3.41	3.41
Satisfaction	32.7%	32.7%	3.02	3.02
Utility			2.93	2.93

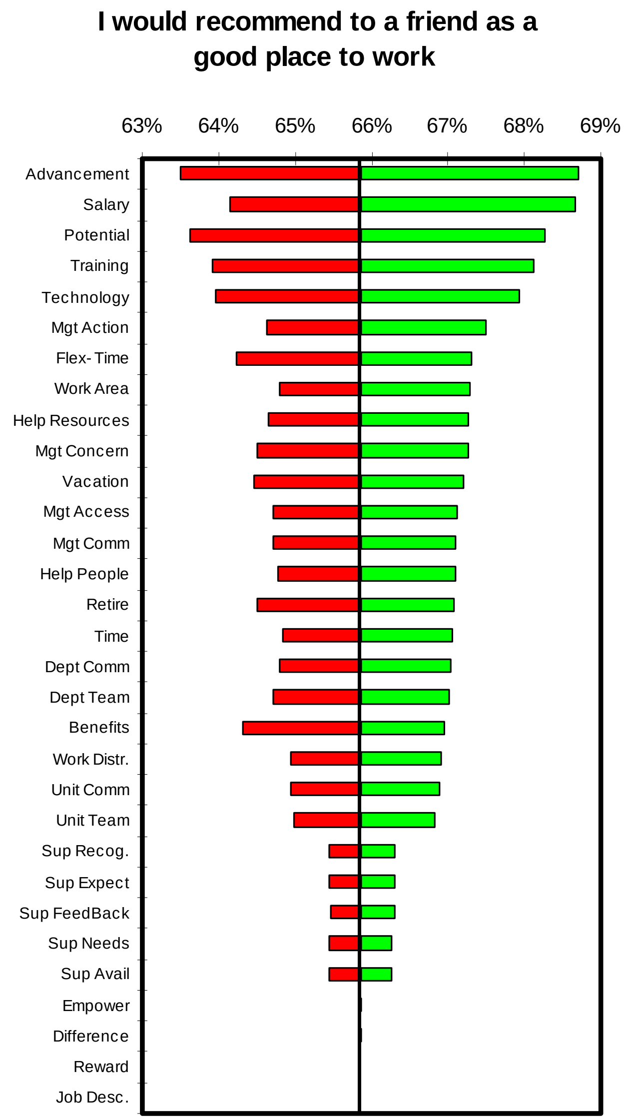
Exclude Overall Satisfaction Effect

☐

3.9.6.2 Sensitivity

As previously noted, the simulators are typically developed to assess the impact of changes in perception on specific attitudes. As such, it is often useful to generate a sensitivity plot as shown below. This plot shows changes in attitude with one unit positive and negative change in the perception of the attributes. An analysis spreadsheet is often setup to automatically compute this plot given any change in competitive values[85].

85 The trick in setting this up is using a DATA TABLE to substitute all values for the changes in perceptions for each component.

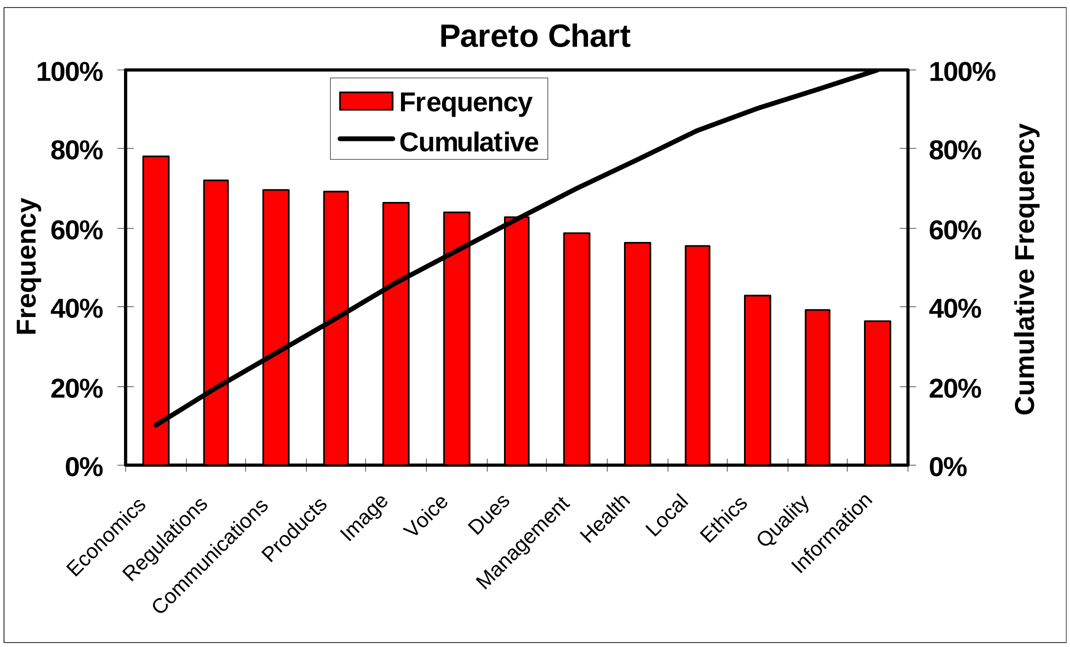

I would recommend to a friend as a good place to work
63%
64%
65%
66%
67%
68%
69%
Advancement
Salary
Potential
Training
Technology
Mgt Action
Flex- Time
Work Area
Help Resources
Mgt Concern
Vacation
Mgt Access
Mgt Comm
Help People
Retire
Time
Dept Comm
Dept Team
Benefits
Work Distr.
Unit Comm
Unit Team
Sup Recog.
Sup Expect
Sup FeedBack
Sup Needs
Sup Avail
Empower
Difference
Reward
Job Desc.

3.10 QUALITY DEPLOYMENT

Total Quality Management (TQM) is a philosophy determined to drive operations so that reasonable customers' "expectations are met or exceeded." Attribute evaluation studies are often undertaken to obtain insight for these programs. One of the approaches is Quality Function Deployment (QFD) which involves relating outcomes and customers' satisfaction with operational performance. Here too, attribute studies are primary sources of information. The major question being asked is "What should we be working on?"[86]

3.10.1 Pareto Charts

Pareto Charts are a standard TQM tool for identifying areas of concern. They are a bar-line diagram indicating the frequency of "faults" and the cumulative percentage of total faults they represent. The charts are usually prepared by segment to show the most dissatisfied attributes. Analysis consists of determining the key attributes that contribute to the majority of dissatisfaction.



3.10.1.1 Steps

The steps in preparing these charts are:

(1) Criteria for determining dissatisfaction must be defined. On a ten point scale this is usually a rating of less than 7.

[86] Attribute evaluation studies may not be adequate to fully address these issues. Perceived value studies (measuring the perceived importance of changes in performance) may also need to be undertaken.

(2) Dissatisfaction occurrence rates for the attributes are computed for the attributes.

(3) The attributes are ranked in order of the highest to the lowest occurrence rates.

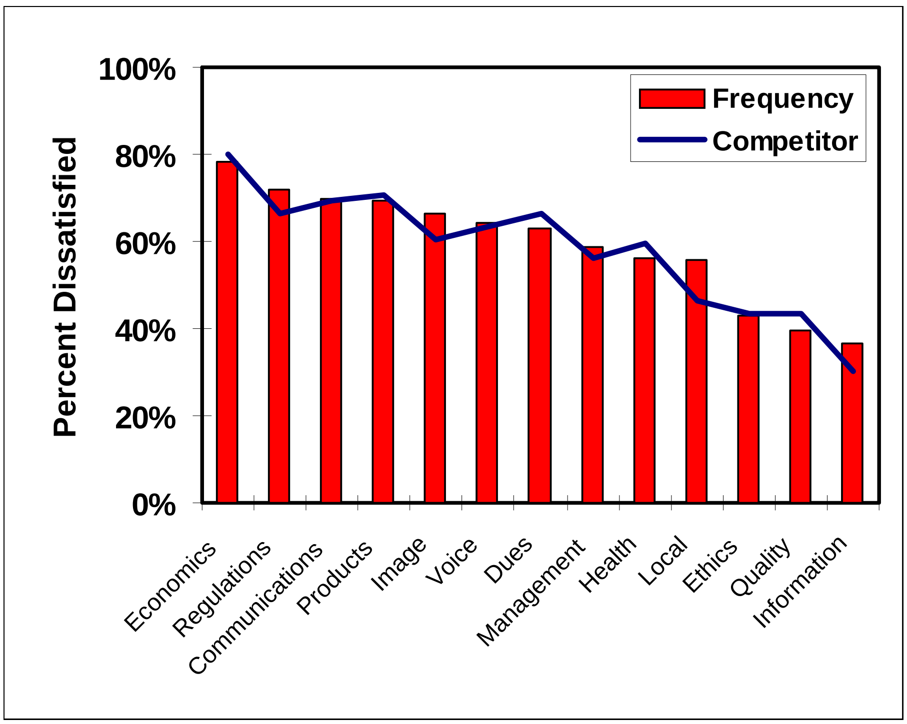
(4) The occurrence rates are normalized and the cumulative frequencies computed.

3.10.1.2 Error Bounds

The error bounds are often computed around the bars. These are computed using the binomial distribution[87].

3.10.2 COMPETITIVE CHARTS

Competitive Charts are similar to Pareto charts but show the competitive values as lines and bars for the performance of the target product. Importance scales (if rating are used) are sometimes included.



3.10.3 IMPORTANCE

The Pareto Charts focus totally on attribute ratings. However, importance can also be a critical factor. There are at least two types of importance measures used:

[87] The formula for computing the error bounds is discussed in the Market Simulator section of this chapter.

3.10.3.1 Stated Importance

Stated Importance is the self explicated values from the respondents. These are "top of mind" estimates of influence. The respondents are asked to: (1) rate the importance of attributes, (2) provide the percent influence of the attributes, or (3) note the most attributes. Based on those responses, a normalized or standardized measure of stated importance is computed.

3.10.3.2 Derived Importance

The derived importance is the relationship between the ratings of the attributes and some overall measure of satisfaction or intent to purchase. Typically the correlation coefficients are used as the measure of derived (or implied) importance[88]. The derived importance is also usually standardized.

3.10.3.3 House of Quality Diagrams

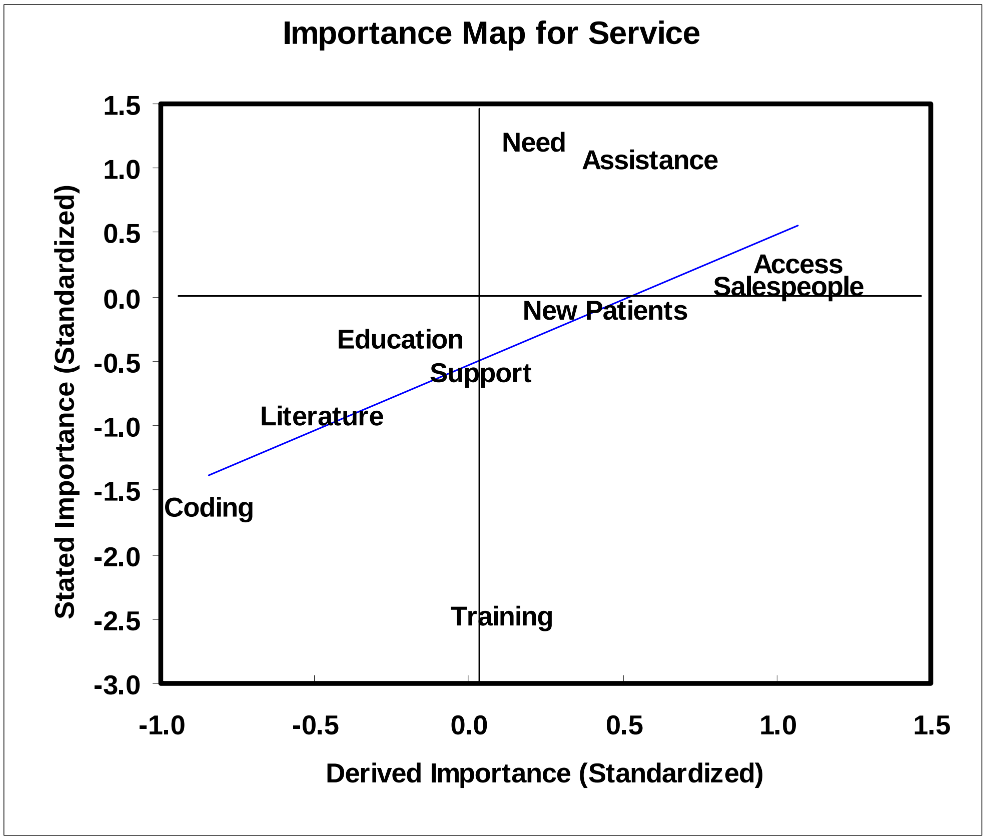
Derived importance is used with multiple criteria to construct an interaction matrix that is part of the larger "House of Quality" Diagram. This matrix relates the importance of attributes on key decision measures.

3.10.3.4 Importance Maps

Maps comparing stated and derived importance (see below) are used to indicate discrepancies. It is not unexpected that there should be differences. Stated importance focuses on the cognitive values and may exclude attributes that are "assumed to be" satisfactory. Derived importance, on the other hand, captures differences in the offerings as well as in perceptions.

---

[88] The regression coefficients would be the appropriate statistic if the attributes were independent. Unfortunately, there are usually high degrees of intercorrelation among the attributes and as such the regression coefficients are unreliable. See the section on Key Driver Analysis for information on regression modeling.



3.10.3.5 Interpretation

Typically these maps are analyzed based on four area quadrants.

(1 and 2) Those items that have either both high or both low importance scores represent consistent evaluations and normally consist of most items.

(3) Those with high derived importance but low stated importance represent "latent" attributes which while not fully recognized greatly influence decisions. These are the attributes where the respondent may assume satisfactory performance.

(4) Those with low derived but high stated importance are thought to be emerging attributes. These are attributes that have top-of-mind awareness but have not yet entered the decision process.

The regression is shown in the chart to indicate the high degree of agreement between stated and derived importance.

3.10.3.6 Error Bounds

Error bounds around the points can be computed though typically this is not done unless there is a small sample size[89].

3.10.4 Quadrant Maps

While traditional Quadrant Map displays only two measures of the attributes, we tend to use three: (1) performance, (2) importance, and (3) competitive advantage.

3.10.4.1 Performances

Product performance is measured either as the dissatisfaction occurrence or the average score. To be consistent with the standard Quality procedures, we prefer the dissatisfaction occurrence rate.

3.10.4.2 Importance

Both stated and derived importance can be used. Typically, we use the stated importance, unless it is not available (not measured in the survey). If stated importance has been collected as a constant sum, we use the average value otherwise it is standardized.

3.10.4.3 Change

In some cases, we find that the stated importance is not particularly meaningful in that all attributes are viewed as important. Under this condition, and where historical data exists, we use the change in performance as a key measure.

3.10.4.4 Competitive Advantage

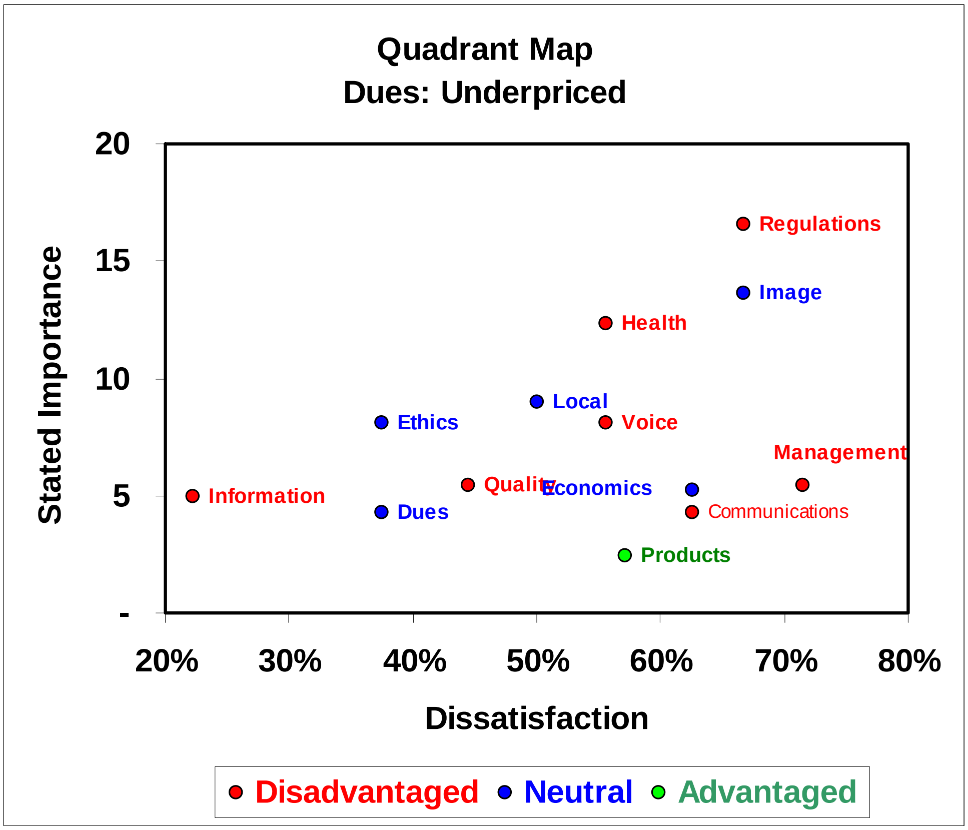
Finally, we consider the relative performance of the attribute compared to competition. If an overall Quadrant Map is being prepared, competitive advantage is taken as the average difference between the performance of the product and its best competitor by respondent and by attribute. If the Quadrant Map represents a comparison with a specific competitor, advantage is computed based on that competitor.

3.10.4.5 Mapping

The resulting maps are similar to that below. These are also done by segment or against a specific competitor. Axes are usually imposed at the median or average score levels creating the four quadrants that give this graph is name. Alternatively three part divisions are also used, set at ± one standard deviation around the mean. This allows for identification of major differences from those that are probably statistically the same.

---

[89] The error bound around the derived importance can be significantly larger than around the stated importance and is based on the uncertainty of the correlation coefficient.



3.10.4.6 Interpretation

Interpretation depends on the focus of the analysis. From a Quality perspective those items that are important, for which the respondents are dissatisfied and for which the competition is doing better are clearly critical. These are the top items. The next level would be those that have two criteria of interest (those that are: (1) important and competitively disadvantaged, (2) with dissatisfied respondents and competitively dissatisfied, and (3) important with dissatisfied respondents. It should be noted, that for communication research purposes satisfaction and competitive advantage are critical and different attributes will be focused upon.