Overcoming key challenges to segmentation

Solutions you can use for critical problems

Dr. Steven Struhl

Converge Analytic

Key challenges in segmentation: both operational and analytical

- For successful segmentation, two types of challenges must be addressed and overcome:
  - **Why the study gets done**
    - Linkage of segmentation to broader strategic goals
    - Corporate commitment to segmentation
    - Ways in which strategic goals inform the project
  - **How the study gets done**
    - Variables to include in the study
    - Analytical approaches ensuring that you reach strategic goals
- First we will discuss overcoming the barriers in "why," and then in "what"

Overcoming key problems in "why segment"

Segmentation requires particular care

- All studies should be done with care, but segmentation takes more effort than most
- Care must be take to avoid:
  - Lack of attention to reasons for using the data
  - Key internal stakeholders not fully engaged
  - Predetermined or traditional thinking
  - Lack of a complete analytical plan
  - Mid-stream changes in segmentation goals and/or methods
- Studies must overcome these barriers at the outset to succeed.

Now which of you gentlemen requested the clean glass?

Segmentation requires clear attention to its uses

- Segmentation in its useful marketing definition[1] directs strategy:
  1. Finding groups that will respond differently to:
     - Communications
     - Product positionings
     - Product configurations
  2. Identifying these groups in useful ways
  3. Reaching the groups selectively

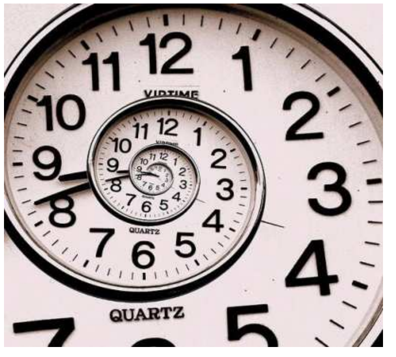
Also fun but not segmentation

- Anything less may be fun but is not segmentation
- The key questions to ask before doing the project:
  - What uses will be made of this segmentation effort?
  - What will get done differently based on application of the findings?

[1] *Segmentation often gets used for any activity in which any groups are located or described. Most are not segments.*

Segmentation: Answerable questions and others

- Sometimes, organizations hope segmentation will answer all questions
  - However, segmentation typically does not address some concerns in detail, including:
    - Market structures
      - In particular, nature and closeness of competitors, substitutability of products
    - Pathways of product use
      - Sequential patterns such as "patient flow" through treatment systems
    - Optimal product configurations
    - Detailed product pricing and price elasticity
- "Just finding out about the market," without plans for using the results, finally wastes everybody's time
  - Management may even get mad about all the time and expense involved



Last minute calls for segments: Have you seen this problem?

- A great deal of trouble may come from “sprinkling some segmen dust” over a study never intended for segmentation
  - This may sound good to some people at the time
    - Still, this almost never gets put to real use
- If there is sudden pressure for more findings than a study was intended to provide, this may signal a problem:
  - Does the study not address its original objectives?
  - Were original objectives not specified correctly?
  - Did somebody want something no study could deliver?
  - Are there new organizational pressures or product problems that we need to understand?

Typically this won’t work

Overcoming lack of stakeholder engagement

- The challenge: Understanding and including key players' needs
- Studies get used only if we first speak with those actually applying the results
  - These usually are senior management people, not researchers
- Another challenge: Finding time to get involvement
  - It takes longer than it may first seem to arrange workshops and/or meetings to define needs and objectives fully
- Engagement also means managing expectations of key players
  - Get these groups in agreement about segmentation and its uses
  - A strong analysis plan is critical

They may not always be progressive but you need them involved

Converge Analytic

Steering around predetermined thinking

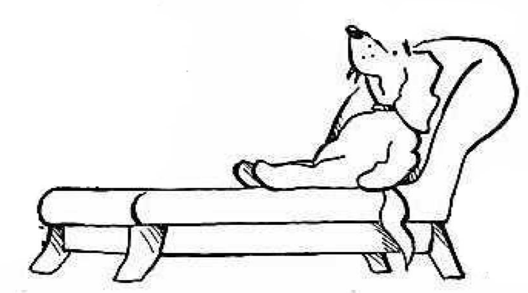
- "A priori" (pre-determined) segments get defined without any original research
- As you would expect, these do not bring true understanding of consumers, patients, doctors, professionals (etc.)
  - In particular, their needs, wants, expectations
- **Some examples,** so you recognize this:
  - "Segments" of doctors by deciles of order volume
    - What you learn (?): doctors in higher deciles order more product
      - How and why this happens—maybe we'll see some other time
  - Segments based on "innovativeness"
    - What you learn(?): "Innovators" are first to try a new product
      - *And*, others try the product later
- "Segments" like these not only give limited insight, but often reproduce some earlier, uninspired work
  - So results are neither sharp nor new

The perils of being not sharp, not new

Converge Analytic

"A priori" thinking (or not) in action (or not)

- We know you won't do this—but this example is too good to pass up
- A fine mix of navel-gazing and not much thinking
    - Segment 1: Mild or moderate depression
    - Segment 2: Severe depression
    - Segment 3: Depression with anxiety
    - Segment 4: Bi-polar
    - Segment 5: Treatment resistant depression



"I ain't nothing but a hound dog."

- You won't be surprised to learn:
    - Severely depressed people are more depressed than those with mild or moderate depression
    - People who are depressed with anxiety also can be put into segments 1, 2, and 5—and maybe 4
    - People with treatment resistant depression can be put into segments 1 to 4
        - And so on
- Not so apparent: Ways to find and address common needs, wants, beliefs, problems, and so on, in all this

Overcoming problems in “what gets done”

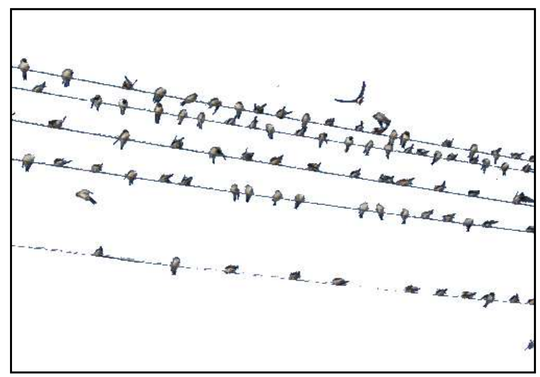
Developing basis question areas

Locating respondents

Converge Analytic

Technical challenges go back to the marketing definition of “segmentation”

- The standard marketing definition of segments also reveals two key challenges in actually performing segmentation, i.e.:
  - Selecting the right basis variables to develop groups that will respond differently
    - Balancing the drive for more information with feasible study goals
  - Reaching the groups selectively
    - Going beyond cross-tabs to develop efficient ways of reaching groups
- Getting these right can lead to a successful outcome
  - Studies that fail analytically most typically fall short in these areas



Locating the right ones can be difficult

Converge Analytic

Find the correct bases for segmentation, then find the audience

- Segmentation studies can only succeed if they zero in on well-defined targeted audiences
  - Getting to that point first requires selecting the right basis variables
  - Even if that happens, a leading cause of study failure is describing groups well, but not being able to reach them selectively
- We definitely can reach our segmentation goals with the right start and sound approaches to locating the most valuable audiences

Possible bases for segmentation: be aware of the options

- Segmentation can serve a wide range of purposes
- Many of lists of basis variables have been developed
  - One of the best, following, was devised over 30 years ago
  - Each type of study described requires many questions
    - Studies succeed by addressing the most important areas and saving the rest for another time
    - Considering the type of study and its proposed uses can help determine whether it is trying for too many objectives

Converge Analytic

Some "preferred bases" for segmentation

For a broader understanding of a market	For studies of new product concepts (and introduction)
• Benefits sought • Needs the product will fill (needs and perceived benefits may not be synonymous) • Product purchase and usage patterns • Brand loyalty and switching patterns	• Reaction to new concepts (measures of intent to to buy, preference over current brand) • Benefits sought • Product usage patterns • Price sensitivity
For studies focusing on product/service positioning	**For studies of pricing decisions**
• Product usage • Product preferences • Benefits sought • Needs the products will fill • Product-, user-, and self-perceptions	• Price sensitivity, by purchase and usage patterns • Product, user and self-images associated with products at different prices • Product usage patterns • Sensitivity to "deals," coupons, etc.

Converge Analytic

Some more "preferred bases" for segmentation

For advertising decisions

- Category usage
- Benefits sought
- Needs;
- Psychographics/"life styles";
- Product-, user-, and self-perceptions
- Responses to creative executions

For distribution decisions

- Store loyalty and patronage;
- Broader shopping patterns
- Benefits sought in store selection
- Sensitivity to deals

- You cannot do these all in one study!
- Even the most thorough single study cannot address more than one or two areas like these
- Picking and choosing can pose real challenges

(Adapted from Wind & Claycamp, 1976)

Basis variables: not too few or too many; behavior is critical

- Either too few basis variables or too much dissimilarity among basis variables can lead to poor results
  - Based on both academic experiments and on many organizations' experiences
  - Therefore, we need to balance the need for inclusiveness with encouragement to put in too many types of questions
- Category-related behavior is a real essential in the basis variables. Not enough emphasis on this can lead to groups that do not respond differently
  - You usually also need opinions, awareness and perceptions
- Variables that help locate people, such as demographics, media habits, etc. can muddy the basis variables
  - They usually are better left outside and put into models that help locate the segment
    - See, e.g., Myers & Tauber, 1976

Basis variables: computational challenges largely resolved

- Newer methods of clustering have resolved long-standing problems with putting all useful variables into the set used as the basis for segmentation
  - Older clustering methods could not handle categorical and binary data along with continuous measures. [1]
- Newer methods that have overcome this problem include:
  - Fuzzy clustering
  - Latent class clustering (e.g., Latent Gold);
  - Two-step cluster (in SPSS);
  - EM clustering.
    - Bayesian clustering, now still being developed, also shows promise.
- Where and why they work best is still are being learned
- If you can use one of these newer methods, technical details will not restrict your options for grouping people

[1] *Semi-academic footnote: Most usually iterative K-means and the hierarchical agglomerative methods, such as "nearest neighbor" or "single linkage," "complete linkage" or "furthest neighbor," "median method," Ward's method, etc.*

Reaching audiences: Going beyond cross-tabs

- Cross-tabs do best as an of “overview” of patterns in the data.
  - They can show where something is happening
  - But they do little to evaluate the extent to which differences matter
- For instance, here we have an automated search of all demographic variables—with a statistic called **adjusted residuals** showing which responses are significantly high and low in
  incidence of the audience
- We can see differences in:
  - Location type
  - Number of children
  - Ages of children
- But how much each matters is not clear
- Fortunately, we can do much better than this. . .

Who They Are	Name	Variable/Level Label	Adj.Res.
	Q79RE	Type of location	
		Suburbs	2.84
	Q80	Number of children at home	
		1	-2.33
		2	-2.29
		5	4.15
	Q2IRE	Ages of children	
		18-24	-2.66

Beyond cross-tabs to locate audiences: Classification trees

- Classification tree methods have proven their value with extensive use by firms in direct marketing, in particular, to help identify groups likely to hold best prospects
- These methods include CHAID, CART, C&RT, Quest, and many others less well known.
  - Earlier methods (e.g., AID) go back to about 1982, and received some criticism for statistical inadequacy—all long solved
- Tree-based methods can lead to valuable output like the *gains chart* (following).
  - These move us well beyond cross-tabs
  - They can in fact effectively guide tactics for selectively reaching desired audiences

Not our type of CART

Classification trees: Showing how we zero in on a segment

- A fictionalized example: the CalmX study.
  - CalmX is a drug to give parents who think their children need medication to improve their behavior.
- Trees enter the analysis after we have described a target segment
  - Nearly 20% are in this segment, called The Worry Warts
  - No surprise—they worry a lot, and not coincidentally hope that CalmX will, well, calm them
- Now that we have characterized this segment, we will use classification tree analysis to see if we can locate them efficiently
  - This step can be make-or-break
    - If the analysis does not clearly direct us to the target, an alternative segmentation scheme may need to be tried

Clear guidance on directions is important

Converge Analytic

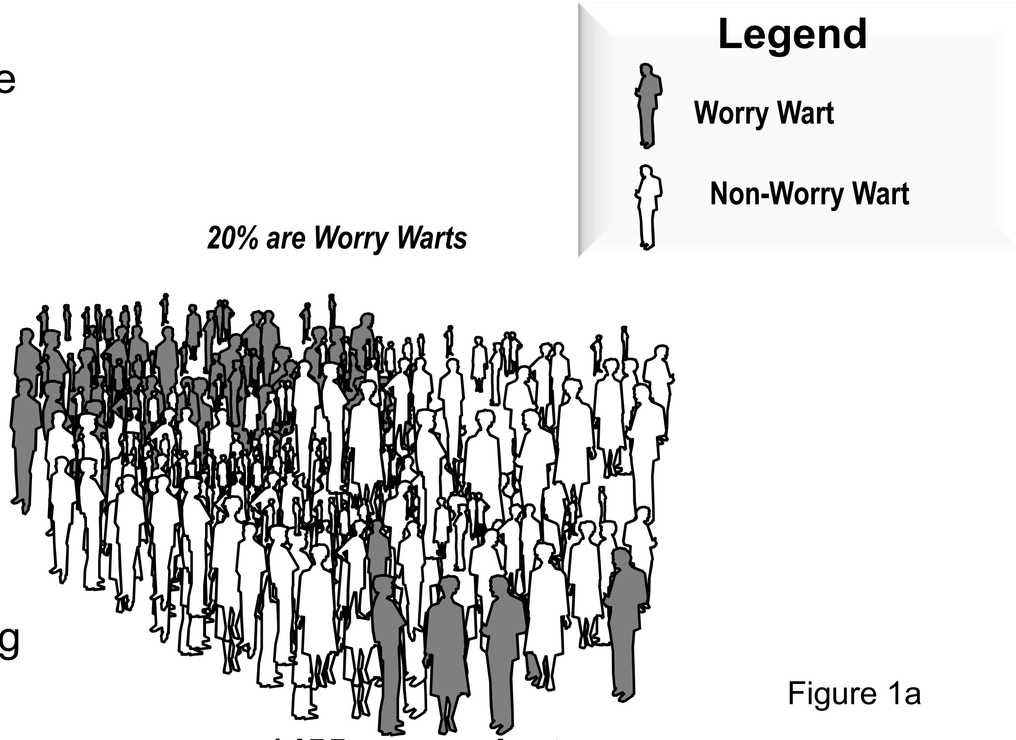
Classification trees work by splitting the sample many times

CHAID seeks the best way to split the sample into 2 to 15 smaller groups[1] as *different as possible* in terms of prevalence or incidence of *Worry Warts*.

We set the procedure to examine 46 demographic items

We will evaluate many possible ways to split the sample, looking for both strong significance and usefulness.

We have 1455 respondents in the survey, of whom nearly 20% are identified as being members of the *Worry Warts* segment



1455 respondents
(men, women, and very small humanoids)

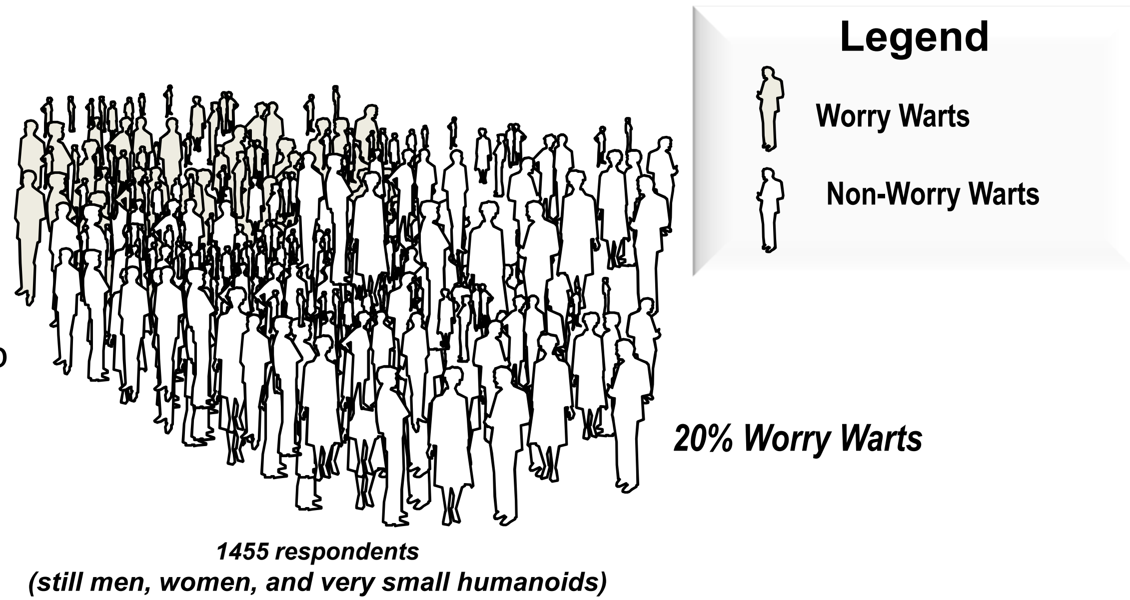
Figure 1a

[1] *CART C&RT, QUEST and related methods are restricted to two-way splits, and have some other small differences*

The first split of the sample reveals a strong contrast

The procedure finds that the strongest contrast in incidence of Worry Warts is based on the type of community in which they live. A higher incidence of this segment lives in the *suburbs* than in either *cities or rural areas*. They are 130% as prevalent in the suburbs as other places (that is, 22%/17%= 130%).

Note that the method is smart enough to combine cities and rural areas into one group—we did not need to instruct it. Cross-tabs would not find this strong difference by looking at the groups side-by-side.



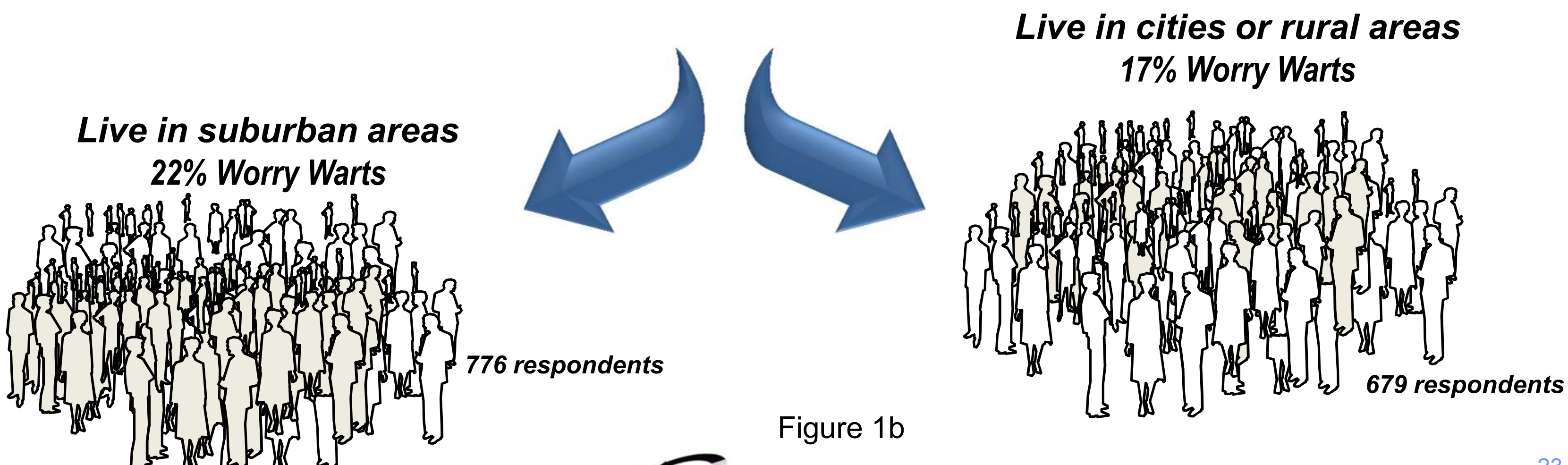


Figure 1b

Classification trees: Beyond the first split, power grows

In the first split, by finding the best way to divide the sample into contrasting groups, classification trees already were "smarter than" crosstabs.

Now, the trees show their real power by going on to split the groups just formed.

Starting with the suburban group should lead to at least one subgroup still higher in incidence of Worry Warts.

If this most significant variable does not lead to an analysis that is useful once the whole tree is constructed, we may go back and try another. We know that six other variables could be put in its place and make nearly as significant a split—differing at most at the 0.00001 (99.999% certainty) level.

Let's see what happens when we split the group living in *suburban areas*. . .

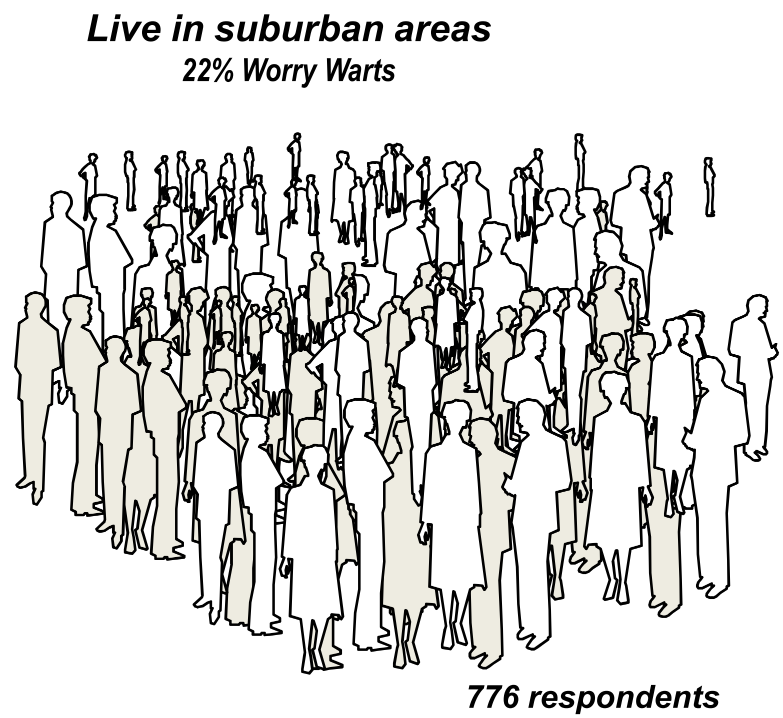


Figure 2

Converge Analytic

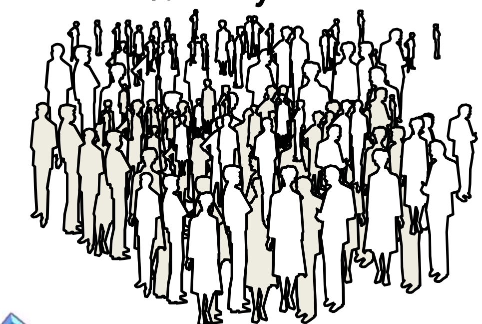
Classification trees: The next split shows far stronger contrasts

Among this group of suburbanites, the procedure found 3 groups differing strongly in incidence of *Worry Warts*, based on how many children are in the household.

This time, splitting into 3 subgroups produces the maximum difference. Large suburban families have the highest incidence. ** Now the highest incidence is 2.8 times the lowest (28% vs. 10%)—a very strong contrast.

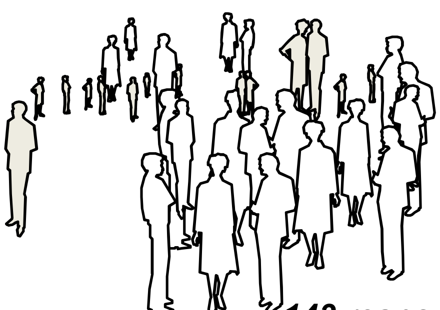
Live in suburban areas
22% Worry Warts

776 respondents



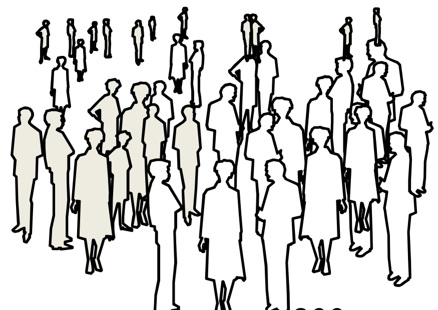
Note: the procedure has tested all possible ways to divide this group of 776 people. For this variable alone, there are 30 possible ways to split the sample into 2 to 5 groups

Kids at home: 1 to 3
10% Worry Warts



143 respondents

Kids at home: 4
19% Worry Warts



200 respondents

Kids at home: 5 +
28% Worry Warts

433 respondents

**This is not a surprise

Figure 3

Classification trees: Strategies for continuing

- Since all the subgroups formed in the last step have well over 100 people, we could then choose one of them and see if we want to split it further

- We could, for instance, go to the group of 443 people who live in the suburbs and have 5+ kids

- This group already has the highest incidence of Worry Warts.

- Splitting it again should lead to at least one subgroup with a very high incidence of Worry Warts.

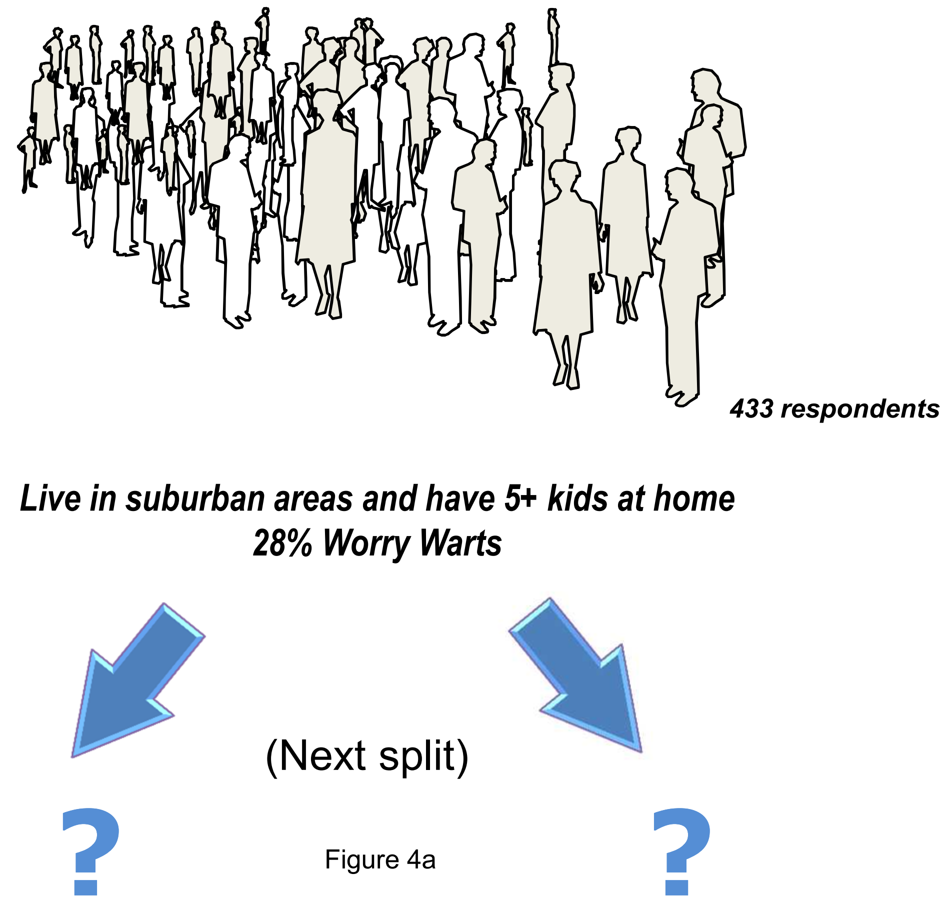


Figure 4a

Alternative strategies are possible for growing the tree

- We can explore different strategies to continue splitting. For instance, we instead could go back to the group of 679 urban and rural dwellers we found in the first split, and see how to divide them further.

- By the end of the analysis, we in fact will split both this group and the group in the suburbs with large families.

- We generally keep on splitting until we reach a lower limit on group size (that we set), or until we run out of significant predictors.

Live in cities or rural areas
17% Worry Warts

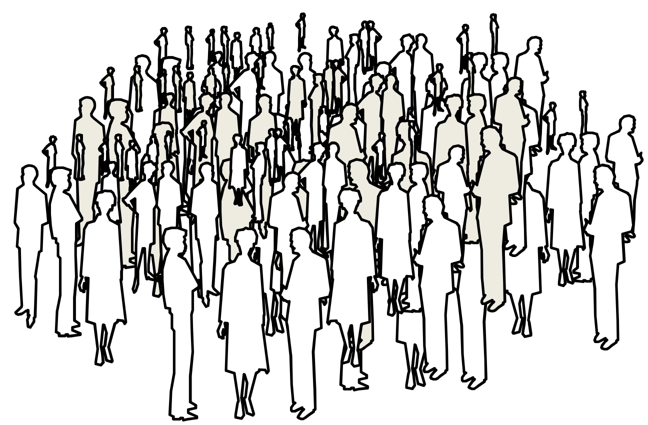


Figure 4b

679 respondents

(Next split)

? ?

Classification trees: Solution appears in a tree diagram

This *tree diagram* shows what happened after we split the sample using the most useful significant variables, and with no group smaller than 50 (or about 3% of the total).

Those in the group with highest prevalence (29.3% –in Group 3) are *5.5 times as likely* as those in the group lowest in prevalence (5.3% - Group 5) to be in the Worry Warts group.

What we did not show:
Getting to this point involved rejecting one "best" predictor because the tree could not go to another level (split again) after that variable, but could split with the next significant predictor tried at that spot instead.

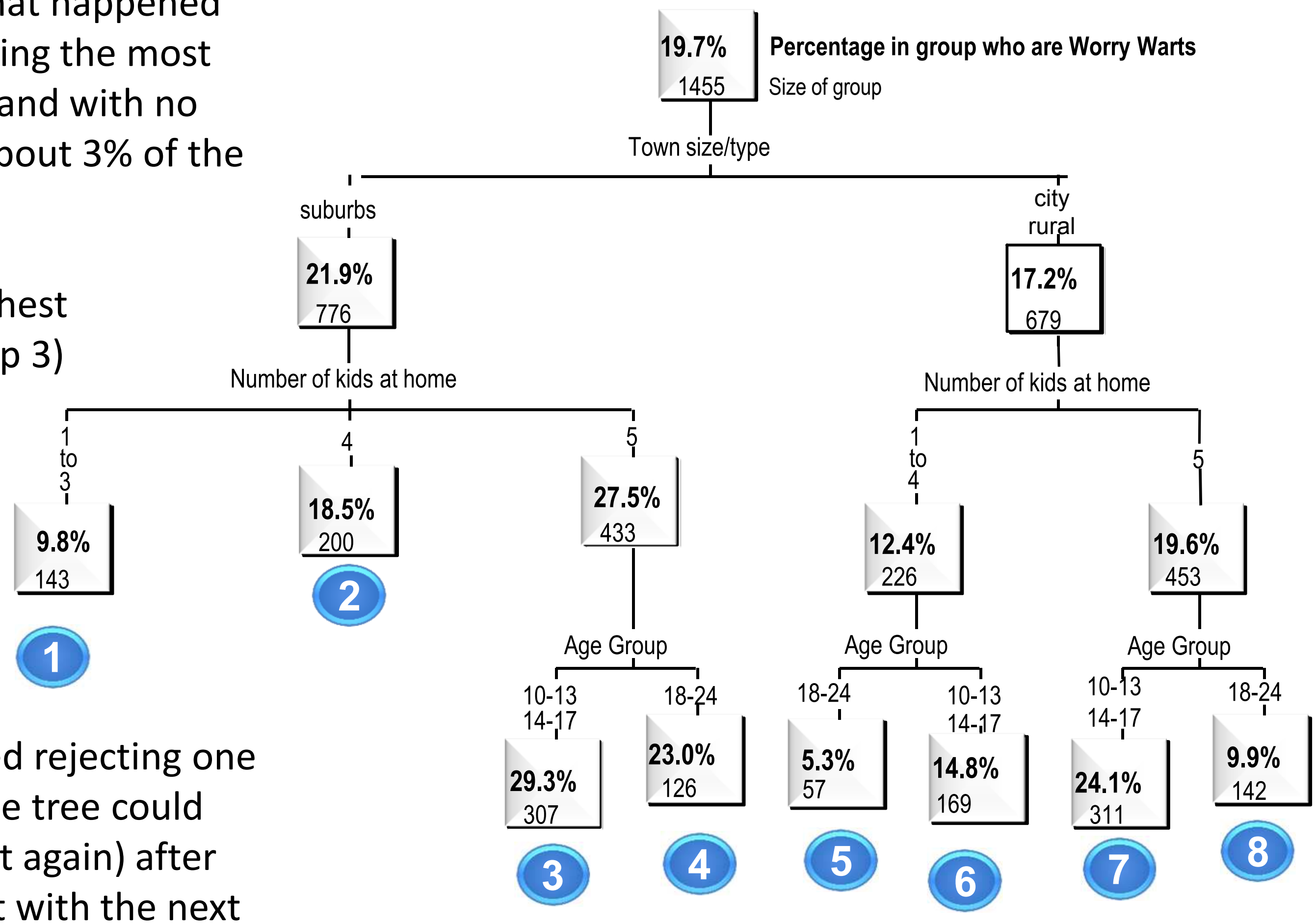


Figure 5

Classification trees: Deciding about the tree

- We can split and re-split the sample, as long as the procedure finds significant variables (predictors) that define the subgroups formed
  - We need to set lower limits on how small groups can become
  - We also may decide to limit the complexity of the model, e.g., by stopping after three or four **tiers** or **levels** of splits
  - You also may decide to “trim the tree” because splits, while significant, do not add to correctly identifying groups
- Programs also can allow tree growth on autopilot.
  - Here tree growth was guided by the analyst
  - However, some programs do not even offer this type of interactivity
  - This is not recommended, because classification trees cannot **look ahead**
    - An initial split that looks most promising may lead to poor results further down the tree
    - So far, only a person can tell, by trying various variables at different points in the analysis
- Trees grown automatically may look highly definitive, but really only suggest possibilities until tested

Converge Analytic

Classification tree output: beyond the tree to gains

- As informative as the tree is in figure 5, it is like a basic roadmap, showing how the tree grew. We can add other valuable details
- A gains chart puts the groups in order, from highest incidence to lowest and shows how the groups are described
- It also provides valuable statistics that show how the groups compare to the overall average
- It typically gives two types of figures:
  - Describing a group, and
  - Cumulative, or the average across all groups down to a certain point
    - For instance, if the first three groups are all the same size and incidences of the target segment are 50%, 40% and 28%, this is how incidences in the groups and the cumulative figures compare

Group percentage	Cumulative percentage	
50%	50%	
40%	45%	That is (50%+40%)/2
28%	41%	That is (50%+40%+28%)/3

- The figures also have index values. If overall incidence is 10%, the first group has an index of 500, or (50/10)*100, the second of 400, and the third of 280.

Classification trees: Gains chart shows this method's power

- The first group (highlighted) has incidence of the target segment is almost 1.5 times the average, and it has 21% of the sample.
  - Efficiency with targeting this group would be 1.5/0.21 (1.5 times the incidence, and expending effort on only 21%) or **700% of the efficiency** of not having this model.
- The top 2 groups together would lead to efficiency of 1.35/0.48 or **320% the efficiency** of not having this model.
- Another page would continue the chart, showing the groups below average in incidence—or those to avoid.

Top of a gains chart: Groups above average in incidence of Worry Warts

Group Characteristics	Group number in the tree diagram	Group size	Group as a % of the total	Incidence of Target Segment	Lift or leverage: Index (100= average)	Groups as a cumulative % of the total	Cumulative Incidence	Cumulative lift or leverage: Index (100= average)
Age Group: 10 -13 and 14-17 AND Number of kids at home: 5 AND Town size/type: suburbs	3	307	21%	29.3%	148.7	21%	29.3%	148.7
Age Group: 10 -13 and 14-17 AND Number of kids at home: 5 AND Town size/type: city or rural	7	311	21%	24.1%	122.3	42%	26.7%	135.4
Age Group: 18-24 AND Number of kids at home: 5 AND Town size/type: suburbs	4	126	9%	23.0%	116.8	51%	26.1%	132.3
OVERALL STATISTICS		**1455**	**100%**	**19.7%**	**100.0**			

More classification tree output: rules for easy classification

- Classification trees also generate rules that define the groups found at the ends of "branches" of the tree.
- These rules are simple "if-then" statements, much easier to understand than equations.
  - These give every person an expected likelihood of use or expected level of use.
  - Here are two rules from the tree diagram we showed:
    - Rule_1: IF Town size/type = suburbs AND number of kids = 1, 2 or 3 THEN percentage Worry Warts = 9.8%.
    - Rule_2: IF Town size/type = suburbs AND number of kids = 4 THEN percentage Worry Warts = 18.5%.
- These rules also go easily into database programs, to give all people in a database **scores,** or expected likelihoods of belonging to the target group.

More about classification trees in action

- Missing data does not interfere with classification tree analysis.
  - "Missing" is simply treated as another kind of response.
  - More flexible classification tree analysis programs can put missing values with other responses where they seem to belong (based on the dependent variable in the analysis) or always hold them to one side, or exclude them.
- Classification trees can handle categorical data with huge numbers of categories.
  - Over 5,000 if the values are to be kept in sequential blocks (ordinal data);
  - Over 2,000 if the values can combine in any sequence.
    - You could, for instance, put every ZIP code in the US into the analysis as a predictor variable

Final example: Trees work amazingly well with many categories

- Here we are seeing significant differences based on many nominal-level responses. Nothing else does this as well. This method can be invaluable in profiling segments.

Legend

Average consumption	***avg=5.78***
# of records	n=6238

Country Where Consumed

Significance = 0.000000000000003

CHINA (110) SINGA (20) KUWAIT (9) HUNGAR (8) GHANA (5) CZECH (4) IRAN (4) ST.LUCI (3) ALGERIA (3) JORDAN (3) ZAMBI (2) KENYA (2) SUDAN (2) CONGO (2) TUNISIA (1) ARUBA (1) NEPAL (1) GAMBIA (1) BULG (1) U.VOLTA (1)	AUSTRAL (256) PORTUG (75) GREEC (31) YUGO (13) EGYPT (9) BANGL (3) W.GER (3) ANGOLA (1) LIBERI (1)	BRAZIL (456) SPAIN (157) N.ZEAL (79) INDONES (48) INDIA (46) MOROCC (27) S.AFR (11) BOLIVI (11) PAKIST (7) ANTIGU (2)	JAPAN (1488) TURK (75) POLAND (45) URUGUAY (10) ZIMBAB (3) SRI LAN (2) BELIZE (1) LESOTH (1) HONDUR (1)	W.BER (355) ITALY (352) TAIWAN (108) BELGIU (104) S.KOREA (97) ARGENT (96) SWEDE (86) FINLAN (74) PHILIPP (40) ISRAEL (39) NORWAY (28) HONG K (16) MALAY (12) LEBAN (6) IRELAN (6) USSR (3) ZAIRE (2) THAI (2) N.ANTIL (2) LIBYA (1)	FRANCE (213) SWITZ (154) AUSTRIA (120) NETHER (47) DENMAR (42) PERU (41) ICELAN (39) JAMAIC (12) CHILE (7) NIGERIA (3) GRENAD (3)	UK (419) TRINID (68) COLUMB (36) DOM REP (23) GUATE (10) ECUADO (8) PANAMA (8) GUYANA (6) SAUDI (4) ROMANI (2) MALI (2) BAHRAIN (1) GABON (1) CAYMAN (1)	U.S. (212) MEXICO (147) VENEZ (65) COSTA (14) BARBAD (11) CANADA (10) BAHAMA (6) SENEGAL (5) GUINEA (4) IVORY (4) LIECH (3) HAITI (3) NICARA (3) EL SAL (2) BERMUD (2) MALTA (1) MONTSER (1) N.YEMEN (1)
2.3	***2.91***	***3.67***	***4.25***	***5.09***	***6.47***	***9.37***	***14.74***
183	392	844	1626	1429	681	589	494

Note: The figure below the variable (country) is the significance of the difference, or better than 99.99999999999% certainty of a difference.

Converge Analytic

What to do with this? Questions? Comments?

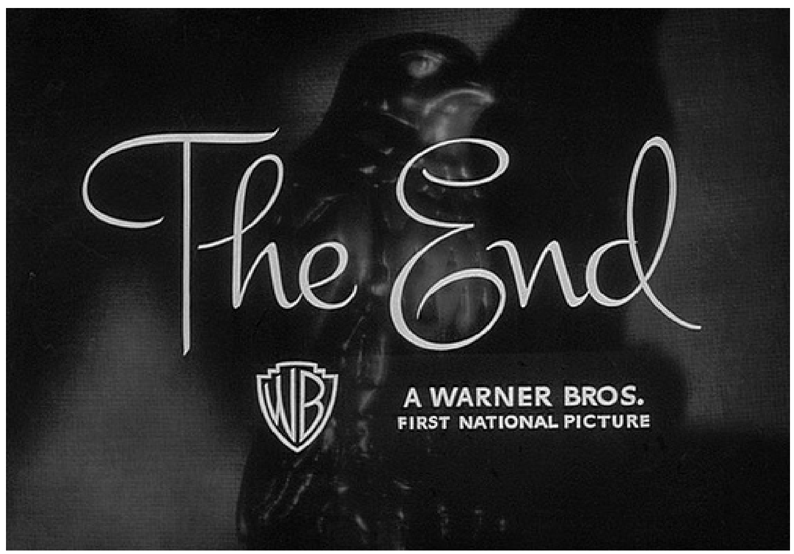
Examples of segmentation output appear in “Sample Segmentation Output”

For more, contact: Dr. Steven Struhl

smstruhl@convergeanalytic.com

smstruhl@gmail.com

847-624-2268



Some key references: classics on the basic topics

- Aldenderfer, M. S. & Blashfield, R. K. (1988). Cluster Analysis. (Sage: Newbury Park, CA.).
- Everitt, B. (1982). Cluster Analysis. (2nd Ed.) (New York: Halsted).
- Green, P. E., Tull, D. S. & Albaum, G. (1978). Research for Marketing Decisions. (Englewood Cliffs, N.J.: Prentice-Hall).
- Kaufmann, L. & Rousseeuw, P.J. (1990). Finding Groups in Data. (New York: John Wiley and Sons).
- Myers, J. H. & Tauber, E. (1977). Market Structure Analysis. (Chicago: American Marketing Association).
- Punji, G. and Stewart, D. R. (1983). "Cluster Analysis in Marketing Research: Review and Suggestions for Application." Journal of Marketing Research, XX (May).
- Schaffer, C. M & Green, P.E. (1996). "An Empirical Comparison of Variable Standardization Methods in Cluster Analysis." Multivariate Behavior Research 31 (2).
- Schaffer, C.M. & Green, P.E. (1998). "Cluster-based Market Segmentation: Some Further Comparisons of Alternative Approaches." Journal of the Market Research Society, 40 (2) (April).
- Smith, W. R. (1956). "Product Differentiation and Market Segmentation as Alternative Marketing Strategies." Journal of Marketing VI (July).
- Sonquist, J. A., & Dunkelburg, W. C. (1977). Survey and Opinion Research. (Englewood Cliffs, New Jersey: Prentice-Hall).
- Stewart, D. W. (1981). " The Application and Misapplication of Factor Analysis in Marketing Research." Journal of Marketing Research, XVIII, (February).
- Tull, D. S. and Hawkins, D. I. (1980). Marketing Research. (2nd ed.). (Englewood Cliffs, N.J.: Macmillan).
- Wilkie, W. L. (1971). "Market Segmentation Research: A Conceptual Analysis," Paper No. 324, Institute for Research in the Behavioral, Economic and Management Sciences. (Purdue University).
- Wind, Y. (1978). "Issues and Advances in Segmentation Research." Journal of Marketing Research, XV, (August).
- Wind, Y. and Claycamp, H. (1976). "Planning Product Line Strategy: A Matrix Approach." Journal of Marketing, 40 (January).