Multivariate Statistical Analysis in NPD: The Contribution of CHAID for Market Targeting and Customization of a Sustainable Product

Istefani C. de Paula, Marcia Elisa S. Echeveste, Manoel M. Silveira, Carla S. Caten
Rio Grande do Sul Federal University, Porto Alegre, Brazil

Abstract--**In the Services and New Product Development (S-NPD) contexts the understanding of customers' demands leads to the development of projects with higher success chances. Multivariate statistical techniques are valuable tools for identifying and valuing the requirements derived from these demands, mainly in complex contexts as the sustainable NPD in which customers' needs vary considerably. The aim of this paper is to present an approach to identify clusters of consumers using CHAID (Chi-squared Automatic Interaction Detector). Based on the clusters formed it is possible to find different segments and to associate requirements demanded by them, what allows the customization by means of product derivation. The proposed method contributes to the Requirements Management research area comprising four steps, illustrated in the development of an eco-friendly new household cleaning product: (i) collection and organization of product requirements; (ii) identification of priority category requirements; (iii) identification of market segments based on the characteristics valued by the target groups; (iv) examination of associations between product requirements and certain characteristics of the target group. The main contributions provided by this method are the deeper understanding of sustainable market groups and the knowledge of how to offer a choice menu to the customer.**

I. INTRODUCTION

Concerning the development of eco-friendly products, the knowledge about consumer demands and habits is of utmost importance to determine preferences at the moment of purchasing, use and discard phases, and it is also important for guiding the definition of technologies to make the recycling and upcycling of such products viable [8],[11]. Consumers are an essential part of this system, as there is little point in developing eco-friendly products if consumers do not adopt them, buying, using and discarding them appropriately. Therefore, studies have been performed with the aim of understanding green consumer preferences and behavior using marketing and statistical tools like decision trees, direct marketing, conjoint analysis, cluster analysis and others, as it is seen in [5], [8], [9], [12], [20], [22].

Organizations' competitive advantage depends on their manager's surrendering to market segmentation trends, with the purpose of tailoring customer's requirements to each segment. Market segmentation leads to better understanding of consumer needs and of their behavior. A market segment consists of a large group of consumers who have the same preferences [15]. In theory, market segmentation is a tool that supports the business because it helps to position the product in the market, to define promotions, attributes or strategies for clients [2], [15], independent if they are green consumers or not. The final consequence of this analysis is the possibility of adding requirements to the product or service that may favor their acquisitions at the purchase phase.

To determine the market segmentation, the most referenced statistical technique is Cluster Analysis as it is described in the review of [11], as in other authors [4], [7], [24] and reinforced by [14]. Cluster analysis aims at the association of categories to find groups with homogeneous characteristics about a particular criterion of the product/service, depending on their application purpose. Although cluster analysis has been referenced in this area, this paper suggests the use of CHAID method (Chi-square Automatic Indentificator Detector) as an alternative to market segmentation. The basic difference between Cluster Analysis and CHAID methodology is that while the first, groups data by similarity of observations, leaving to the researcher the task of identifying similar attributes variables which form the groups or segments, CHAID seeks relationships between the predictor variables and the categorical outcome measure and presents it in the form of a tree diagram with combinations of features that maximize the differences between the categories. In this study, the application of CHAID enables to identify which features are significantly associated with each group of consumers, (in this case named eco-friendly or regular consumers). This information allows us to address the specific demands of each group, giving insights to product development teams. The team may customize products, fitting the interests of each consumer category. Despite the technique advantage it is not frequently used by product developers, what makes it an opportunity to bring this technique to the eco-friendly product development area. Besides the deeper understanding of sustainable market groups, a final contribution provided by this method to the development team, is the knowledge of how to offer a choice menu to the customer.

Therefore, the aim of this paper is to present the method inserted in the early stages of the Product Development Process, through a practical application, to determine the requirements of a product that are associated with certain segments of consumers. The question underneath our investigation is: how to identify the requirements that are valued by eco-friendly when they acquire, use and dispose products viewing the preservation of the environment? Generally, this method is applied to define market segmentation using socio-demographic characteristics as dependent variables of the model. In this article the proposal is that the independent variables or predictors are the requirements of a product.

II. CHAID METHODOLOGY

CHAID is a decision tree method for targeting a population of interest. Decision trees are a sequence of partitions of a database so as to maximize differences on a dependent variable [10] and are similar to the dendrograms of cluster analysis, used for representation and graphical interpretation, but are constructed and interpreted in different ways. CHAID tree is generally used when the segmentation is defined in terms of the demographic characteristics or categorical variables with predictive power [17]. Since each segment is associated with a probability of response, these segments can be ordered to select the most promising for some research characteristic [6].

Some classical [19] aspects may be considered relative to the collected data or the origin population from where they come from, such as: (i) it is necessary a large sample to run a CHAID analysis; (ii) most frequently the variable is not directly addressed in the study but it is a binary classification; (iii) the data come from a sample, usually, but not always, collected through an experimental design; (iv) very often the exploratory factors used in the analysis may be correlated; (v) there may be interaction among the effects; (vi) in reality, there are logical properties and a cause and effect relationship among the variables.

CHAID is run by repetitive partitions of the population in subsets, in two or more nodes from the original data [16]. The method is based on Chi-square tests of association and it segments the data set into mutually exclusive subsets that, exhaustively, best describe the response variable [23].

A. The procedure

CHAID methodology operates on a dependent variable categorical, nominal or ordinal scale, and maximizes the significance of the chi-square statistic in each partition, featuring CHAID as a significance test framework [26].

The theoretical CHAID procedure begins with the definition of the dependent variable presenting $d \geq 2$ categories and a c predictor presenting $c \geq 2$. The immediate issue is to reduce the contingency c x d table into a j x d table, with the more significant association resulting from the combination of predictor categories. Conceptually, you must first calculate the $T_{(j)}^{(i)}$ statistical, the Chi-square statistics for the i[th] training method of a table j x d ($2 \leq j \leq c$). Then, if $T_{(j)}^{(*)} = \max T_{(j)}^{(i)}$ is the highest value of the chi-square statistic found for the j x d table, $T_{(j)}^{(*)}$ is chosen as value of greater significance associated [10].

This technique has much in common with other statistical techniques such as discriminant analysis and the dendrogram of cluster analysis [10], since all of them lead to groups creation; despite they have different basis. In CHAID method, variables are called *monotonic* when the dependent variables are ordinal categorical and are called *free* when the dependent variable is described in a nominal scale. The variable is called *floating* when there is some unknown information or *missing* if there is a dismissive value in the analysis [1].

Due to the successive comparison tests applied, in this technique it is calculated a correction factor in the Bonferroni inequality used to obtain an adjusted significance level. The Bonferroni multiplier is the number of possible ways that 'c' categories may give in 'r' categories of the predictor variable. For r = c, B = 1. If the predictor variable is *monotonic*, the Bonferroni multiplier is obtained by the definition of the binomial coefficient, equation (1).

$$B_{monotonic} = \binom{c-1}{r-1} \tag{1}$$

If the variable is *free* type, the multiplier is presented as the equation (2),

$$B_{free} = \sum_{i=0}^{r-1} (-1)^i \frac{(r-i)^c}{i!(r-i)!} \tag{2}$$

If the variable is *floating* type, the multiplier is presented as the equation (3).

$$B_{float} = \binom{c-2}{r-2} + \binom{c-2}{r-1} = \frac{r-1+r(c-r)}{c-1} = B_{monotonic} \tag{3}$$

According to [17], the multiplier is useful in adjusting for multiple testing, and it is based on the group of categories with the highest level of significance. The purpose of this correction is to protect the type I error so it does not exceed a level of significance value ($\propto$) stipulated in the test. The Bonferroni inequality provides a set of methods and significance levels for some groups to ensure that the type I error does not exceed a certain $\propto$ [18]. Then, by inequality Bonferroni, α_0 is the value with which the group of categories should be tested, equation (4), where $N_B(c)$ is the Bonferroni adjustment factor [1].

$$\alpha_0 = \frac{\alpha}{N_B(c)} \tag{4}$$

According to [14] the proposal is searching for a maximum $T_{(j)}^{(*)}$ (chi-square statistic) using the stepwise method, assessing the input of each variable in the model and verifying if their contribution are significant or not, among the predictor variables. The proposal can be summarized in five steps, as follows: (i) for each predictor, make a cross table of predictor categories with the categories of the dependent variable; (ii) find the pairs of categories of predictors (considering only pairs determined by different types of predictors) for which 2 x d is less significant. If this significance is not a critical high value, unite the two categories, and repeat this step; (iii) for each category made up from the original three or more categories, find the most significant binary partition to which the mixture of the categories can be solved. If the significance is higher than a critical value, implement the division and repeat step (ii); (iv)

calculate the significance of each considered predictor and isolate the most significant of all. If the significance is higher than a threshold value, subdivide the data in accordance with the number of categories of the next predictor. This step requires a significance test of the reduced contingency table; (v) for each partition of the data that has not been tested yet, return to step (i).

B. Considerations about the sample

For CHAID application purposes it is necessary relatively large samples to generate more reliable results and to form subgroups that enable the sequence analysis. For samples sizes considered small it can be run the Fisher's exact test [3] as an alternative to the chi-square test.

C. Advantages and Limitations

The results obtained using CHAID are presented graphically and are easy to interpret and reading [13]. An important consideration of the CHAID results is that it can be used to generate single probability scores of individuals in the sample belonging to a particular node. As combinations of predictor variables define the segments or the response of interest, new cases can be classified for certain segment by the values of these variables, so the odds for new cases can be estimated. Furthermore, this method can make estimation for all the considered population or only part of it [6].

The disadvantages of the method are that the independent variables (predictors) are considered sequentially and not simultaneously, moreover CHAID does not guarantee a single optimal solution [21]. Diepen and Franses [6] indicate two main problems regarding the method: i) instability of CHAID tree - when the tree can adjust a set of data in an acceptable manner, if the original table data undergoes change, a completely new different tree is created; ii) over-fitting occurs when the variance between the average value generated by an estimator and the observed values is very large.

D. Validation

Three criteria may be considered for validation of the model adopted: (i) graphic evaluation, which is the representation of the cumulative gain (gain chart). This graph is characterized by displaying an arc shape about a straight diagonal. The axis of abscissa of the graph varies from 0 (zero) to 100%; (ii) estimated risk indicates the risk associated with misclassification of reference category of the dependent variable; (iii) percentage of correct classification that the model gives to the category used as reference. Both for (ii) and (iii), the acceptable values are part of the decisions to be taken by researchers considering the category used as a reference in their search.

III. METHODOLOGICAL PROCEDURES

The following items relate to the CHAID application guidelines are summarized in eight steps. As computational resource, it was used the statistical package [SPSS 18®].

(i) **Definition of the research problem** – The problem has to be determined so as to reveal the variables to be modeled. The problem is contextualized and specialists in the research team identify the variables involved, their level of measurement and the usefulness of the information generated by the variable in the solution of the given problem. The research problem is presented in introduction and section 4.

(ii) **Sample characterization -** The sample size must be large enough to ensure the application of the statistical chi-square test. This is an important requirement because this test is applied to verify the association between variables arranged in crosstabs, originated from the nodules subdivisions for each upcoming variable inserted in the model. Sample size are calculated like chi-square tests, the expected frequency should be $np > 0.5$ for each cell. In this study, students of a special course at UNIPAZ-SUL/Porto Alegre city, directed to eco-oriented people, were the interviewed group. These students were selected due to their green consumption orientation and because a pilot market segment would be of interest, in the case of development of a product. The UNIPAZ population in Brazil is of thousands of students, if other states units are considered, and they are a potential market for an eco-product. Other consumers, considered regular were selected and interviewed in supermarkets from Porto Alegr city. More details from the sample are described in section 4.2.

(iii) **Determination of the dependent variable –** The determination of the dependent variable for the model is a direct consequence of the determination of the research problem: “how to identify the requirements that are valued by eco-friendly consumers when they acquire, use and dispose products viewing the preservation of the environment?” Therefore, the dependent variable is the “consumption value” that differentiates individuals in two categories: “regular consumer” (0) versus “eco-friendly consumers” (1).

(iv) **Determination of predictor variables** – The determination of predictive variables for the CHAID model corresponds to the selection of the independent variables in case this is a regression analysis problem. The only condition for choosing these variables is that they are categorical. In this study the independent variables are: store (or shop site), product certification, brand, convenience of preparation and product efficiency.

(v) **Descriptive evaluation of the variables –** The first study conducted on the sample data is considered to hold a descriptive analysis of the variables, univariate analysis, to avoid inconsistencies or to detect some aspect that is relevant to the search. Frequency tables can bring initial information on the proportions and absolute frequencies of the categories

in each variable. An example of descriptive analysis is a bar chart that serves to make certain frequency comparisons between characteristics of interest.

(vi) **CHAID graphical representation** - A major advantage of the method is that the result can be interpreted through a graphical representation of CHAID. The decision tree of this method starts with a single initial module that allows a measurable value of a dependent variable. This value contained in the first node is split into other nodes, which in turn are also subdivided in accordance with some characteristic associated. The reading tree is bottom up, starting at the last node subdivided called 'end node' and following their precursors nodes until reaching the initial node.

(vii) **Evaluation of CHAID table –** In addition to the graphical representation of the results of this method CHAID decision tree can be presented in table form. In the CHAID table, the columns indicate the frequency ratios of categories for each node. It is also listed in the table the percentage of the reference category and the other categories in relation to the subdivision being held and for the total sample.

(viii) **The gain graphical representation**. Three criteria may be considered for validation of the model adopted: (i) graphic evaluation, which is the representation of the cumulative gain (gain chart); (ii) estimated risk indicates the risk associated with misclassification of reference category of the dependent variable; (iii) percentage of correct classification that the model gives the category used as reference. The acceptable values for (ii) and (iii) are defined by researchers taking into account the category used as reference.

The eight steps described and the proposed analysis will be detailed concomitantly with the following sample application.

IV. CHAID IN THE DEVELOPMENT OF AN ECO-FRIENDLY CLEANING PRODUCT

The development of an eco-friendly cleaning product was chosen as the research unit for application of the method proposed in section 3. Regarding the problem of reconciling market expansion to the new consumer needs, most companies have adhered to new concepts of production and product development. In a systemic view, the concern in product development is meeting the requirements of the new paradigm of sustainability. A sustainable endeavor must meet four basic requirements: being environmentally friendly, economically viable, socially just and culturally accepted [25]. In this paper the expression eco-friendly will be adopted for simplification purposes.

Apart from the definition, consumers are an essential part of this system, as there is little point in developing eco-friendly products if consumers do not adopt them, buying, using and discarding them appropriately. Thus, it is important to identify the characteristics facing the consumer with respect to the acquisition, use and disposal of products, although it is known that to change consumption habits is not an easy task for product developers or governments.

A. Definition of the Research Problem and requirements

Considering that eco-friendly consumers value special characteristics in products, these product characteristics could also motivate regular consumers to acquire green products. Therefore, the research problem is: *how to identify the requirements that are valued by eco-friendly when they acquire, use and dispose products viewing the preservation of the environment?* The aim is to contribute in the understanding of what product requirements would attract a regular consumer and retain eco-friendly consumers, using CHAID. Thus, companies might focus their attention on the requirements most valued by eco-friendly consumers or define best strategies to win the other consumers, offering products with more attractive features to the market and also meeting sustainability issues.

For investigation purposes it was necessary to define requirements to be used in CHAID methodology. Five requirements of a cleaning product were included in the investigation: 'Product Brand', 'Convenience', 'Certification', 'Store (or shop site)' and 'Efficiency'. The requirements were defined by applying the Factor Analysis statistics technique on a set of 37 product requirements, in a prior study. More details of this Factor Analysis can be found in [8]. The selected requirements were those with highest factor loadings that were negotiable for the product (it means that they were neither mandatory nor normative and could be adjusted to meet a specific market segment). 'Certification' and 'Store' requirements were entered into the model due to the interest of the researchers in their influence.

Each requirement was divided into two levels: (+) represents the presence and (-) the absence of the attribute, which are the categories that summarize their main characteristics. For instance, the 'Product Brand' requirement was divided into (+) Well-known brand and (-) Brand not known; 'Convenience' was split into (+) Ready for use and (-) Requires preparation, etc. From the combination of these levels it was developed eight scenarios or profiles of the cleaning products as demonstrated in Table 1. The scenarios were presented to respondents who ordered them according to their preferences.

To each scenario, it was estimated the monetary value based on the comparative market price of the product. This estimation does not necessarily represent the actual market value, but the basis used to relativize the consumer's choice, indicating the price that they were willing to pay for the product represented in a given scenario.

TABLE 1 – SCENARIOS CREATED FOR PREFERENCE INVESTIGATION CONCERNING SUSTAINABLE CLEANING PRODUCTS

Scenario	Product Brand	Convenience	Certification	Store	Efficiency	Price
S1	Well-known brand	Requires preparation	With green seal	Real store	Efficiency 30% lower	$7,85
S2	Well-known brand	Ready for use	Without green seal	Virtual store	Standard efficiency	$3,80
S3	Well-known brand	Ready for use	With green seal	Virtual store	Efficiency 30% lower	$7,15
S4	Brand not known	Requires preparation	With green seal	Virtual store	Standard efficiency	$6,30
S5	Brand not known	Ready for use	With green seal	Real store	Standard efficiency	$8,00
S6	Brand not known	Requires preparation	Without green seal	Virtual store	Efficiency 30% lower	$2,65
S7	Well-known brand	Requires preparation	Without green seal	Real store	Standard efficiency	$4,50
S8	Brand not known	Ready for use	Without green seal	Real store	Efficiency 30% lower	$4,35

B. Sample characterization

Sampling was carried out from 4th October 2010 to 30th October 2010. The population studied consisted of students from the UNIPAZ-SUL institution, in Porto Alegre city, South Brazil, who receive a trans disciplinary and holistic education, thus being characterized as an eco-friendly group. Consumers without declared environmental concerns form the regular consumer group, interviewed in supermarkets and in public places. The survey included a sample of 102 elements, being 44 classified as eco-friendly and 58 regular consumers. The sample size was considered satisfactory for the purposes of this research, since this is an exploratory study.

C. Determination of the model variables – dependent and predictor

The dependent variable in this study is called 'segment' and has two categories defined as: eco-friendly and regular consumers. The category of interest or reference in this study is considered the eco-friendly consumers. The predictor variables in this study are the product requirements (product brand, convenience, certification, store and efficiency).

D. Descriptive evaluation of the variables

As a first result it is presented a graph of scenario frequencies according to the preference of the respondents. Fig. 1 demonstrates that scenarios S5, S4 and S3 presented higher frequency of choice with 55.60%, 20.00% and 15.60% respectively. Analyzing the preferred scenarios by respondents and unfolding the requirements levels that define them, it is observed that, in general, consumers have no preference for the attribute or requirement 'well-known brand' when considering cleaning products.

It can be observed through this descriptive analysis, that there was willingness to pay a higher value for the product when it presented attributes or requirements like: 'a reliable green label', 'ready for use', ' efficiency equal to regular cleaning products offered in the market' and if the 'purchase of it is made in physical store'.

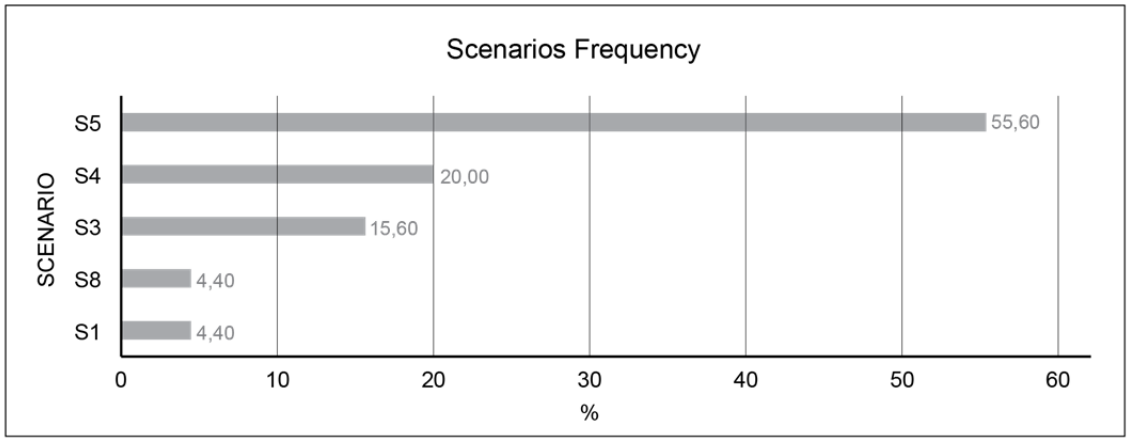

Fig. 1 – Scenarios preferred by the consumers

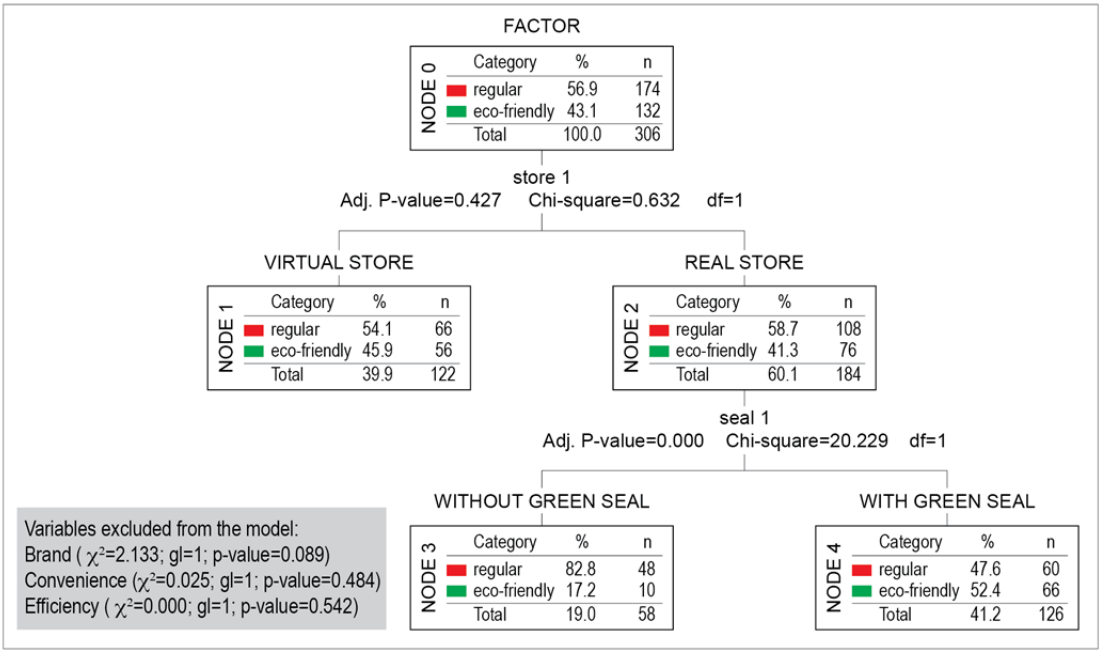

Fig. 2 – Representation of the CHAID diagram

E. CHAID Graphical Representation

The graphical representation of the CHAID resembles the branches of a tree. Each end node (without subdivision) is a differentiated group or segment of the study population and, considering the tree, it determines the requirements that a given segment most value in the product. The zero node provides a summary of the entire sample in relation to the dependent variable. From the categories of this node, the predictor variables are tested by cross tables following the CHAID methodology. The CHAID diagram is shown in Fig. 2.

Accordingly to CHAID diagram, it was observed that 'Store' and 'Certification' are the most relevant requirements for the product. The independent variable, 'Store' has been included in the model because it is a variable of interest of the researchers for the development of future work. Thus, Fig. 2 shows that the fourth node is associated with the declared eco-friendly respondents and these, in turn, show a clear preference for a product that presents a certificate of guarantee, an ecological character and still prefer to purchase these cleaning products in physical stores.

F. Evaluation of CHAID table

A table summarizing the results can represent the tree diagram (see Table 2). This table shows the relevant information available in CHAID diagram. Based on the decision tree the table shows the higher frequency category and its percentage for each node, indicating the dependent category in each node.

The preponderant dependent category (eco-friendly/regular) is the one that presents more than 50% of the frequency between the two categories of the dependent variable in the resulting node. The fourth node, for example, shows the eco-friendly consumer as the preponderant dependent category. For this node, the eco-friendly column (N) denotes a frequency of 66 cases representing 52.4% of the total nodule frequency. The fourth node has in total 126 cases representing 41.2% of the total number of the cases analyzed. The interpretation for the fourth node is that eco-friendly consumers have a preference for buying sustainable products with a certification label and that the product is purchased in physical store (see Fig. 2).

TABLE 2 – TABLE OF THE TREE CHAID - I

	Regular		Eco-friendly		Total			
Node	N	Percentage	N	Percentage	N	Percentage	Category	Previous nodes
0	174	56,9%	132	43,1%	306	100,0%	Regular	
1	66	54,1%	56	45,9%	122	39,9%	Regular	0
2	108	58,7%	76	41,3%	184	60,1%	Regular	0
3	48	82,8%	10	17,2%	58	19,0%	Regular	2
4	60	47,6%	66	52,4%	126	41,2%	Eco-friendly	2

TABLE 3 – CLASSIFICATION TABLE

	Predicted		
Observed	Regular	Eco- friendly	Correct percentage
Regular	114	60	65,5%
Eco-friendly	66	66	50,0%
Total percentage	58,8%	41,2%	58,8%

In the classification table (table 3), the lines correspond to the categories observed by respondents and the columns represent the dependent categories (predicted) using CHAID model. The model showed approximately 58.8% of overall accuracy to classify correctly the respondents in relation to their status as being: eco-friendly or regular.

The value 0.412 (41,2% Table 3) indicates a risk that the criteria used to characterize the groups of eco-friendly consumers and regular may not have been sufficiently discriminatory to detect this distinction. In relation to the classification table it is possible to ponder: for consumers considered eco-friendly it is correctly predicted approximately 50.00% of the cases.

This leads to the assumption that the criteria used to define the eco-friendly and regular segments were not sufficiently discriminatory. Respondents are distinguished by their spending habits and about their attitude toward the environment, other characteristics not considered in the study would affect the results. Thus, in many cases consumers' behavior end up being so similar that the differences are not captured by the tests.

G. The gain graphical representation

Fig. 3 shows the cumulative gain graph for the model. The gain of the vertical axis of the graph is the cumulative percentage of profits gained from the table column while the horizontal axis represents the cumulative percentage of cases of orderly sample. So, for instance, considering as reference the *x*-axis value 50, the corresponding value is found in the gain curve is about 50%. It is assumed that the proportion of the reference category did not increase from the initial node.

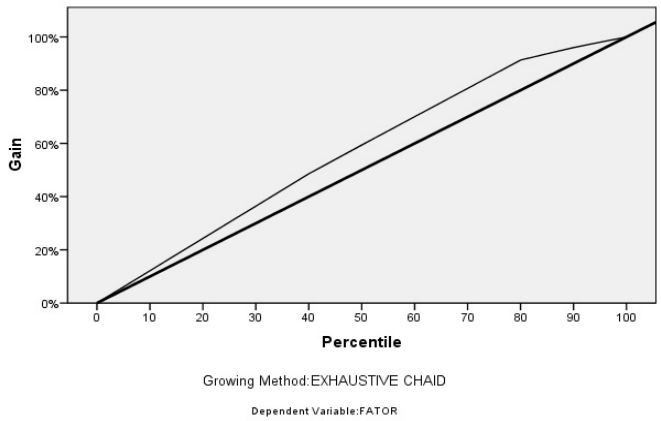

Fig. 3 – Gain Graphical Representation

In a visual gain graphic interpretation, the more the curve presents a convex aspect in relation to the straight reference, the more the predictive value is associated with the model. In case, the curve does not present an intense convex aspect, this suggests the need of reconsideration of the sample used.

V. FINAL THOUGHTS

The aim of this paper was to present theoretically the CHAID method inserted in the early stages of the Product Development Process and to accomplish a practical application of it. The example considered the association of the requirements of a green cleaning product with certain customer segments. It is possible to highlight three main contributions: (i) the method was discussed and applied in an objective way, hence product developers may use it as a different option to aid product development decision-making analysis; (ii) based on the case of development of a cleaner product it is easier to be adapted to other products, in the early stages of development; (iii) some consumer profile characteristics were assessed considering the purchase, use and disposal of eco-friendly cleaning products, though the sample is very limited and represent a specific Brazilian sample.

The results were considered satisfactory because it was possible to identify groups and specific product requirements that analyzed by the researcher can help them to determine in what groups to concentrate their development efforts. Moreover, the use of product requirements as predictor variables in CHAID adds one more option analysis to detect consumer preferences. The product development team can make decisions based on these analyzes as: to emphasize product certification to retain eco-friendly consumers and to motivate regular consumers to purchase green products; independent of the consumer orientation give preference to selling their products in physical stores.

Other benefits of CHAID method results in this study are: companies that develop green cleaning products could apply special programs for their products viewing to increased consumer's acceptance; the product development team may perform thorough survey on sustainability requirements that can be added to their products; companies developing the product can trace specific sales strategies for defined segments, highlighting aspects of the product to retain the segment that uses a given product or conquer another new segment. All observations are limited to the sample analyzed. Despite the proposition of using statistical tools in understanding consumers, the power of learning by iterative buying behavior and the effect of word of mouth in influencing other consumers cannot be neglected. The project team has to maintain a broad range of strategies in creating competitive products.

Besides the increase of samples in the continuity of this exploratory investigation, it is suggested for future work to add more requirements to the product and more levels to these requirements, so that the trade-off analysis will present

a more discriminatory character between eco-friendly and regular consumers.

REFERENCES

[1] Biggs, D., De Ville, B. and Suen, E. 1991. "A method of choosing multiway partitions for classification and decision trees". *Journal of Applied Statistics*. 18, 1 (Jan. 1991), 49–62.

[2] Chen, J.S. 2003. "Market segmentation by tourists' sentiments". *Annals of Tourism Research*. 30, 1 (Jan. 2003), 178–193.

[3] Conover, W.J. 1971. *Practical nonparametric statistics*. Wiley.

[4] Crawford, C.M. and Di Benedetto, C.A. 2003. *New products management*. McGraw-Hill Irwin.

[5] Desai, K.K. and Trivedi, M. 2014. "Do consumer perceptions matter in measuring choice variety and variety seeking?" *Journal of Business Research*. 67, 1 (Jan. 2014), 2786–2792.

[6] Van Diepen, M. and Franses, P.H. 2006. "Evaluating chi-squared automatic interaction detection". *Information Systems*. 31, 8 (Dec. 2006), 814–831.

[7] Dolnicar, S. 2002. "A review of data-driven market segmentation in tourism". *Faculty of Commerce - Papers*. (2002).

[8] Echeveste, M.E.S., Paula, I.C., Silveira, M.M. and Marx, A.M. 2013. "Multivariate analysis applied in the initial phase of the development of a new eco-friendly product". *Technology Management in the IT-Driven Services (PICMET), 2013 Proceedings of PICMET '13:* (San Jose, CA, USA, Jul. 2013), 1594–1603.

[9] Elliott, R. 2013. "The taste for green: The possibilities and dynamics of status differentiation through "green" consumption". *Poetics*. 41, 3 (Jun. 2013), 294–322.

[10] Hair, J.F. 1995. *Multivariate data analysis with readings*. Prentice Hall.

[11] Hauschild, M., Jeswiet, J. and Alting, L. 2005. "From Life Cycle Assessment to Sustainable Production: Status and Perspectives". *CIRP Annals - Manufacturing Technology*. 54, 2 (2005), 1–21.

[12] Haws, K.L., Winterich, K.P. and Naylor, R.W. 2014. "Seeing the world through GREEN-tinted glasses: Green consumption values and responses to environmentally friendly products". *Journal of Consumer Psychology*. 24, 3 (Jul. 2014), 336–354.

[13] Hoare, R. 2004. "Using CHAID for Classification Problems". *Proceedings* (Wellington, New Zealand, 2004).

[14] Kass, G.V. 1980. "An Exploratory Technique for Investigating Large Quantities of Categorical Data". *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 29, 2 (Jan. 1980), 119–127.

[15] Kotler, P. and Keller, K.L. 2006. *Marketing management*. Pearson.

[16] Linoff, G. and Berry, M.J.A. 1997. *Data mining techniques: for marketing, sales, and customer relationship management*. Wiley.

[17] Magidson, J. 1993. *SPSS for Windows TM CHAID TM release 6.0*. SPSS.

[18] Miller, R.G. 1981. *Simultaneous statistical inference*. Springer-Verlag.

[19] Morgan, J.N. and Sonquist, J.A. 1963. "Problems in the Analysis of Survey Data, and a Proposal". *Journal of the American Statistical Association*. 58, 302 (Jun. 1963), 415–434.

[20] Olson, E.L. 2012. "It's not easy being green: the effects of attribute tradeoffs on green product preference and choice". *Journal of the Academy of Marketing Science*. 41, 2 (May 2012), 171–184.

[21] Perreault, W.D., Jr. and Barksdale, H.C., Jr. 1980. "A Model-Free Approach for Analysis of Complex Contingency Data in Survey Research". *Journal of Marketing Research*. 17, 4 (Nov. 1980), 503–515.

[22] Sonnenberg, N.C., Erasmus, A.C. and Schreuder, A. 2014. "Consumers' preferences for eco-friendly appliances in an emerging market context". *International Journal of Consumer Studies*. 38, 5 (2014), 559–569.

[23] Ture, M., Tokatli, F. and Kurt, I. 2009. "Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients". *Expert Systems with Applications*. 36, 2, Part 1 (Mar. 2009), 2017–2026.

[24] Urban, G.L. and Hauser, J.R. 1980. *Design and marketing of new products*. Prentice-Hall.

[25] World Commission on Environment and Development 1987. *Our common future*. Oxford University Press.

[26] SPSS Statistics Base 17.0 User's Guide - SPSS-Statistcs-Base-Users-Guide-17.0.pdf.