AVANCES
EN CIENCIAS E INGENIERÍAS
ARTÍCULO/ARTICLE SECCIÓN/SECTION C

Beyond main effects assumption in Conjoint Analysis: Comparison of Conjoint Value Analysis vs. Choice-based Conjoint. Statistical approach and construction of designs applied to New Product Development

Andrés Paredes[1], Alvaro Enríquez[1], Danny Navarrete[1*]

[1] *Universidad San Francisco de Quito, Colegio de Ciencias e Ingeniería -El Politécnico- Ingeniería Industrial. Calle Diego de Robles y Vía Interoceánica, Campus Cumbayá, Quito, Ecuador.*
[*] *Autor principal/Corresponding author, e-mail: dnavarrete@usfq.edu.ec*

Editado por/Edited by: Cesar Zambrano, Ph.D.
Recibido/Received: 2015/07/21. Aceptado/Accepted: 2015/09/27.
Publicado en línea/Published on Web: 2015/12/30. Impreso/Printed: 2015/12/30.

Abstract

The assumption of only main effects in Conjoint Analysis methods has created a debate whether to focus or not on the impact of interactions in determining the most prefered combination of attributes of a product. In this research a comparison of Conjoint Value Analysis CVA and Choice-Based Conjoint CBC surveys were undertaken to contrast them through utility scores, importance values of attributes and goodness-of-fit using ready to drink beverages as the subject. The main effects assumption in the CVA composition rule was compared to the interaction terms in the CBC one. Two scenarios were developed; the first one considered inner characteristics of the subject and a sample size of 250 respondents. The second one considered the presentation characteristics of the subject and a sample size of 150 respondents. The two higher total utility scores were obtained in the CBC using an interactive composition rule. In Scenario 1 a higher goodness-of-fit was found in the CBC, including significant interactions, in contrast with Scenario 2, where no interactions were found, and CVA had a higher goodness-of-fit.

Keywords. Conjoint Analysis, New Product Design, Fractional Factorial Design, Hierarchical Bayes Estimation, Design of Experiments.

Más Allá de la Suposición de Efectos Principales en Análisis Conjunto: Comparación Entre Análisis Conjunto Tradicional Vs. Análisis Conjunto Basado en Elección. Enfoque Estadístico y Construcción de Diseños Aplicado al Desarrollo de Nuevos Productos

Resumen

El supuesto de considerar solo efectos principales en el Analisis Conjunto ha creado un debate si enfocarse o no en el impacto de las interacciónes para determinar la combinación con mayor preferencia en los atributos del producto. La comparación se realizó entre las encuestas del Análisis Conjunto Tradicional CVA y el Análisis Conjunto Basado en la Elección CBC para contrastarlas a través de los valores de utilidad, valores de importancia de los atributos y bondad de ajuste en ambas metodologías, usando una bebida lista para tomar como sujeto de prueba. La suposición de efectos principales en la regla de com-posición del CVA fue comparada con la inclusión de términos de interacción significativos en el CBC. Se desarrollaron dos escena-rios; en el primero de consideró características internas del sujeto de prueba y se utilizó un tamaño de muestra de 250 encuestados. En el segundo escenario consideró características de presentación del sujeto de prueba y un tamaño de muestra de 150 encuestados. Los dos valores de utilidad más altos se obtuvieron del CBC, usando una regla de composición con interacciones, acabó considerando a la Cerveza, en cambio en el CVA este nivel reporto una utilidad negativa. En el Escenario 1 se encontró una bondad de ajus-te más alta para el CBC, incluyendo interacciones significativas, en contraste con el Esce-nario 2, donde no se encontraron interacciones significativas y en ese caso el CVA tuvo una bondad de ajuste mayor.

Palabras Clave. Análisis Conjunto, Diseño de nuevos productos, Diseño factorial fraccionado, Estimación Jerárquica Bayes, Diseño de experimentos.

Introduction

Conducting business decisions is critical, and the use of resources has to be efficient. Some examples of these decisions are: product/service design, product line and portfolio optimization, capacity planning, customer support management, as well as volume and mix flexibility decisions [1].

To respond to those critical decisions efficiently and effectively, CA has been used as a technique that allows researchers to translate and predict customers' needs and

ISSN 1390-5384

expectations into product characteristics [2–4]. In this way, industries can use this technique to launch successful products/services in a competitive market [5]. In this manner, conjoint analysis has allowed different areas like statistics, probability, and experimental design as well as econometric modeling to predict more realistic behaviors of the marketplace [6].

CA has been around for about 40 years, and since then a lot of improvements have come in hand with technology that has eased the calculations and made possible to test with more representative samples and improve the estimation of the data collected [7, 8].

There are essentially four methodologies of CA mentioned in detail in Rao [9], briefly explained next. The first one is CVA, and it is the traditional conjoint methodology that uses Rb stated preferences or ranking combinations [1, 9].

The second one is CBC; this CA uses stated Cb to select a combination from a set presented; the ways preferences are obtained are partly deterministic and random [1].

The third methodology is ACA in which, first, a self-explicated elicitation task considers attribute importance values and desirability levels using ranking and subsequent rating in order to tailor partial profiles for each respondent; these profiles are followed by a paired presentation of those choices in a graded comparison scale.

The fourth CA methodology is called self-explicated, where respondents are asked to evaluate the desirability of each level of all the attributes as well as the relative importance values assigned to them.

Rao [9] also mentioned that there are two different kinds of models in which the four methodologies are classified: decompositional and compositional. The first three methodologies mentioned earlier are decompositional because the data are decomposed to obtain partial utility scores of each attribute level of a combination; this is in contrast with the compositional approach, where utility scores are composed from the data obtained of each of the attribute levels.

Karniouchina [8] expressed that from more than 150 publications in top journals about CA, only 5 studies have compared the elicitation methods used before in CVA and CBC as Rb and Cb, and no concrete results were obtained, but they concluded that CBC performed better at the individual-level.

There are some benefits from using Rb or Cb elicitation methods. For example, Rb simplifies market decisions, whereas Cb is easier for respondents to adapt and give information about their preference, but in terms of design, Rb is easier than Cb [1, 2].

Now, CBC is becoming more important due to the realism in the way respondents make trade-offs, simulating what happens in the market place; this is in contrast to CVA where respondents can either rate or rank a specific combination [9, 11]. Also, choosing a combination is a simple and more natural task that everyone can understand [10].

One of the main reasons that CVA was first used was the utility estimation through OLS for Rb or monotone regression for ranking that permitted individual utility estimation, unlike CBC, where the Multinomial Logit Model (MNL) only permitted to estimate aggregate utilities. However, recent powerful estimation methods like HB have allowed to obtain information from respondents with fewer questions and to calculate individual utility estimations; this process has led to enhance the information quality, reducing significantly the chances of getting noisy data [12].

Moreover, the strength of the methodologies that estimate individual-level utilities, considering only main effects like CVA, are obtained at the cost of denying the presence of interactions, considering their values to be negligible. Therefore, if significant interactions are found, the conclusions reached by a traditional CA may be invalid, unlike CBC that offers the capability to estimate interactions between attribute levels [13].

The interaction analysis in a CA study can proportionate valuable information to find a model that could be more accurate [3, 4, 7]. Differentiated two distinct composition rules, one considering main effects and other considering interaction terms. The first one is called Simple Additive Model, where the partial utility of attribute levels in a combination are summed up, and the second one is called Interactive Model, where the interaction terms can be added to obtain the total utility or combination preference. However, in the latter model, despite portraying a more realistic situation, it implies more complex calculations [7].

Using individual level estimation approaches has reduced the need for modeling interactions, but this does not directly take for granted the changes that significant interactions can have on the respondents' preferences, so their effects should not be considered negligible [6]. Also significant interactions can be crucial information that needs to be included in the analysis and are often ignored by researches due to time or resources [11].

Practical Applications

Conjoint Analysis is a widespread statistical research technique that can be applied to New Product Development by determining the preference (utility) of a specified product/service through its components and finds an optimal combination of their attributes. There are different Conjoint Analysis methodologies, and a dispute has come along about which methodologies can be more appropriate to use. The traditional full profile Conjoint Analysis or Conjoint Value Analysis (CVA) was the first methodology developed and the most frequently used due to its simplicity in calculations. However, Choice-Based Conjoint (CBC) is gaining popularity due to its way of presenting combinations that simulates what happens in the marketplace more accurately.

Little research has been made in comparing different Conjoint Analysis in an effort to conclude or recommend the utilization of a particular methodology of this technique. Several authors have given guidelines to researchers to know which methodology to choose, basing their recommendations only on the capabilities of each one. There are no conclusive results about which of the methodologies could lead to better results.

Main effects only assumption and the election of one of the elicitation methods have brought interest in using more complex and accurate estimation methods that can include more crucial information for the modeling of consumer's preference for a new product.

Materials and Methods

Objectives of the CA research

The objective of the research was to find which conjoint analysis methodology CVA or CBC got better results taking into account the advantages and disadvantages that each of them have. Specifically, if the advantage of considering interactions terms in the CBC can contribute to the share of preference model explained by the CA, due to the fact that CVA cannot include interaction between the factor that are being explored. Therefore a CVA vs. CBC was conducted and compared to determine which conjoint methodology got better results using an additive composition rule for CVA where only main effects could be considered in contrast with an interactive composition rule in CBC where interaction terms were included. The results were analyzed by goodness of fit, attribute importance values and utility scores.

The theoretical problem is whether to be able to use a more sophisticated and more in use methodology to achieve better results in conjoint analysis through the develop of surveys based on choices (CBC) and then analyze the results; compared against a more straight forward methodology, nowadays less often used where rating specific combinations are used (CVA).

The type of data in the results obtained by each methodology is different. In CBC the data type is nominal so the statistical analysis is completely different from the data obtained in CVA which is ratio data, where traditional statistical basis can be used to approach and obtain results. This distinction is crucial for the theoretical problem stablished and the use of a a complex resolution method such as Hierarchical Bayes in CBC is putted to test against using common statistical knowledge in CVA to obtain people preferences and a model that could predict more accurately these preferences.

With the CBC the theory involved is much more complex as mentioned before than with CVA, so the results in this investigation could provide a guide to whether use a CBC or a CVA based on analytic and technical criteria to discard one of them and not merely discard the CBC for its complexity. The implication of using robust statistical criteria to discard one of the methodologies is a strong fundament of the study in order to make conjectures.

To do so, a study subject was needed in order to generate data to compare both methodologies. Hence the practical problem was to generate a conjoint analysis that allowed combining different levels and factors to apply a share of preference model. Therefore the creation of new mixes of beverages was selected to comply this requirement due to its recent elevated local consumption [14–16].

Next, using a market research, RTD were crafted and those findings were adapted to each conjoint methodology, according to the rules to create unbiased combinations and avoid possible problems due to presentation order.

After limiting the population based on local statistical information found, the surveys were crafted with the help of Sawtooth Software and the experimental designs were carefully selected in order to meet all the technical requirements to get the best possible results in each conjoint methodology.

Market Analysis

Over the years, the consumption of alcoholic beverages has changed in a way that cocktails are now taking more market than before, due to their flavor mixes involved and their low alcoholic content [17]. Also [14] indicated that the propaganda and marketing towards this new trend in consumption of low alcoholic level beverages is directed to the young population, and the core characteristics of these products are: new flavors, quality of ingredients in the mix, and the presentation with the package ready to drink.

The growth in volume of this RTD can be seen in countries like New Zealand, where from early 1990s to 2007, the consumption went from 3% to 14% and assures that this growth will continue in part because of the marketing made and people's consumption habits [14].

Data from the World Health Organization recorded consumption on average values in liters per capita; these values have increased between the periods of 2003-2005 (3.8L) to 2008-2010 (4.2L), which also confirms the growing consumerism of these products [18].

Target Population

The research was conducted in the city of Quito, Ecuador due to high consumption rates per capita [15, 16].

Groups of ages between 18-44 years old were selected to conduct the study because this age group is found to consume more than the others [16]. The socioeconomic population status targeted was high (AB) and mid high

(C+) due to their more frequent consumption [16]. Finally, in order to obtain reliable results, only respondents that have drunk at least once in the last month were targeted, considering consumption frequencies [16].

Considering the parameters mentioned earlier 132,797 people composed the target population [15].

Design of the CA

A comparison for conducting a CVA with Rb elicitation method and a CBC with Cb, are shown in Figure 1 with modifications according to the performed research (the modifications are shown in color).

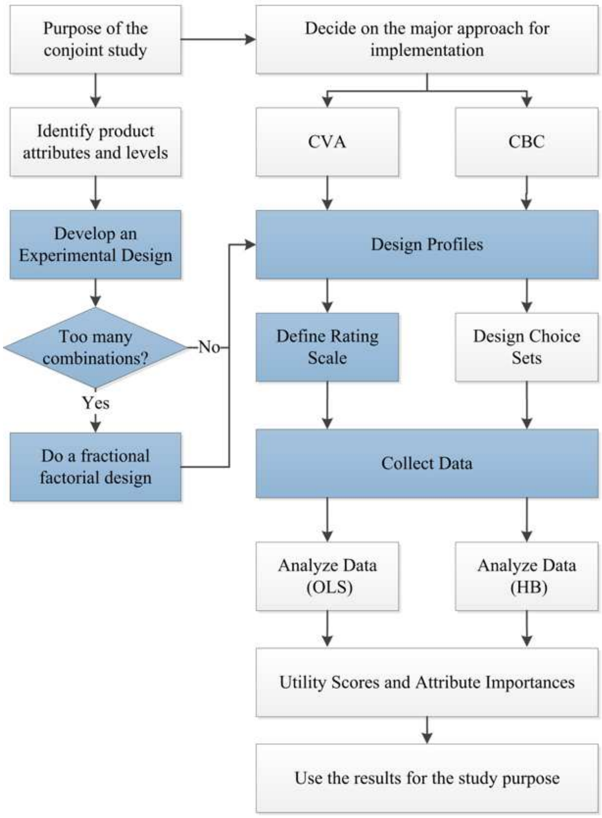

Figure 1: Conducting a Conjoint Analysis RB vs. CB adapted from [9].

Identify product attributes and levels, and Factorial Design Development

Three focus groups were conducted, as well as market research of places where cocktails were sold in order to identify which ingredients were going to be included in the experimental design as attributes and their respective levels.

Each focus group conducted had the purpose of extracting important alcohol consumption information from the participants. The information was about the types of alcohol they used to drink and the preference about each of them. The participants also provided the information about the mixes they preferred to make or buy, and the ingredients and flavors they liked or dislike in a mix. In order to obtain more information about cocktail recipes, many of them were consulted in the market with different bartenders.

The resultant attributes and levels for this study are presented in Table 1.

Attributes	Levels		
Type A Liquor	Whisky	Rum	Beer
Type B Liquor	Tequila	Gin	Vodka
Touch of Flavor	Mentha Spicata	Grenadine	Energy Drinks
Solvent	Lemon Juice	Tonic Water	Lemon Flavored Soda

Table 1: Attributes and levels for CA Scenario 1.

Beer was considered as an attribute in this study due to the high percentage of consumption (79.2%) [15].

All attributes have the same number of levels, representing a symmetric design [9]; these types of designs are the most common ones and have been studied the most. Also they have obtained more representative results than their counterpart asymmetric designs, where the attributes may have different number of levels each. This study was labeled as Scenario 1.

Thus, the resulting factorial design was a 3^4, which generated 81 possible combinations [19]. However respondents often lack the energy or patience to answer many questions in a CA, thus this quantity of combinations can burn the respondents [10].

Typically, the amount of tasks respondents can answer for a CVA is between 20 and 30 [8]. As a guide of the number of questions that could be asked to obtain stabilization in the results, [10] recommended at least two or preferably three times the number of parameters to be estimated. The parameters to be estimated are: $\#levels - \#attributes + 1$, which yielded to 9 parameters. The factor of 3 was applied, thus 27 questions had to be asked in the CVA study. With the CBC methodology there were recommendations to establish a suitable number of questions, but in order to choose the same amount of combinations presented to respondents in both CA, 9 questions with 3 combinations each was proposed for CBC.

The resulting combinations for Scenario 1 could have been difficult to conceptualize for respondents considering that these RTDs have ingredients as attributes and levels that are not typical in the local market mixes. To mitigate this possible bias on the research objectives, a second scenario with most frequent market cocktails was developed to run a complementary study, where now the attributes could be more distinguishable to respondents.

The attributes in this case had different number of levels, resulting now in an asymmetrical design [9]. The

Attribute	Levels			
Cocktail	Mojito	Cuba Libre	Submarine	Michelada
	Blue Margarita	Tequila Sunrise	Padrino	Whisky Sour
	Vodka Tonic	Screw-driver	Tom Collins	Gin Tonic
Container Type	Aluminum Can		Glass Bottle	
Container Size	Medium (330ml)		Small (220ml)	

Table 2: Attributes and levels for CA Scenario 2.

attributes and their levels are shown in Table 2. This study was labeled as Scenario 2.

The study was approached with these two scenarios for the comparison of both methodologies. Two surveys for each scenario were developed, one with CVA and one with CBC respectively. In each of the four surveys, respondents rated or chose among the 27 combinations presented.

Next, each methodology is described in terms of the design involved in each one.

Conducting CVA technique

Considering the number of questions to obtain stable results and to eliminate information overload, a FF design 3^{4-1} was proposed. A 3-level FF design matrix was used to obtain the most efficient fraction of the design, which considered a minimum aberration criterion in order to guarantee a maximum resolution design [20]. Noting that CVA does not consider interactions, the criterion assures not to confuse the main effects between them and with two-way interactions, obtaining a design of resolution IV.

For Scenario 1, ten versions of the design matrix were randomly generated and manually introduced into the SSI Software to avoid order and context effects in the questionnaire, which could have affected people's responses, generating a potential bias in the results [10]. For Scenario 2, the same amounts of versions of Scenario 1 were automatically generated.

A 10-point (10 categories) rating scale was used for the CVA methodology due to its advantages in reliability, validity, discriminating power, and respondents' preferences of rating scales [21].

CVA Utility Estimation by OLS

The basic weighted additive model for CVA methodology using rating tasks can be stated as follows [22]

$$r_k = \beta_0 + \sum_{j=1}^{J} \sum_{m=1}^{M} \beta_{jm} x_{jm} + e_k$$

where

r_k = Response for option k.
β_0 = Intercept or constant.
β_{jm} = Part worth utility of level m of attribute j.
$x_{jm} = 1$ if option k has level m on attribute j (otherwise) else $x_{jm} = 0$.
e_k = Error term.

Part worth utilities were estimated by applying multiple regressions with OLS, using a dummy variable coding in which deleting one level of each attribute from the computation was done. Otherwise, a linear dependence among the variables describing the levels of each attribute would lead to indeterminacy in the computation [23].

Conducting CBC technique

As mentioned earlier, the condition was to test the same amount of combinations between the two CA methodologies. In order to achieve that, each choice set was composed by 3 combinations plus the "none option", obtaining a total of 9 tasks; in addition, two fixed tasks were added to improve the estimation (Sawtooth Software, 2014) giving a total of 11 tasks instead of the 27 tasks on the CVA. This was done to comply with the typical number of choice task on a CBC, which is about 8 to 12 [24].

Next, a method to construct the choice sets was selected. Several methods exist to do the construction, depending on their capabilities, exposed in Table 3.

The random method was selected to compose the CBC sets because it is the most complete design in estimating interaction effects, in spite of being the least efficient when estimating main effects [13]. For computerized interviewing, in order to gain in design efficiency, but most importantly to decrease order and context effects, 300 different version sets were automatically created [11].

Each version showed 27 different combinations among the 81 possible ones, unlike the CVA methodology where the 27 combinations were part of the same fraction but the order was randomized.

CBC Utility Estimation by HB

The utility associated with a combination can be stated as follows [22]

$$U_{ij} = \beta x_{ij} + e_{ij}$$

where

U_{ij} = The utility of respondent i associated with profile j (this could be a combination A or B).
β = A vector of parameters to be estimated.
x_{ij} = A vector of attributes of profile j presented to respondent i.
e_{ij} = The stochastic portion of the utility function.

	Design Method						
Effects	**FF Shift**	**FF Mix & Match**	**FF L^{MN}**	**CBC Complete Enum**	**CBC Shortcut**	**CBC Random**	**CBC Balanced Overlap**
Main Effects Only	X	X	X	X	X	X	X
Interactions		X	X	X	X	X	X
Prohibitions				X	X	X	X
Alternative Specific Effects		X	X		X	X	
Cross Effects			X			X	

Table 3: Comparison of Capabilities [4].

Respondent i would choose profile A over profile B if $U_{Ai} > U_{iB}$ and the probability of such choice is $P_i(A) = Prob\{\beta x_{iA} + e_{iA} \geq \beta x_{iB} + e_{iB}\}$

CBC is based on a maximum utility model MNL, which is part deterministic and part random, borrowing information from the rest of the sample to estimate the probabilistic part of the utility [25].

Using MNL, the HB estimation method performs an iterative process using Bayesian Analysis to draw the parameters of the prior distributions that the data is assumed to follow, in this case the partial utility weights of the preference model.

Using the Bayesian Analysis implies to turn the statistical estimation process around; this is done instead of assuming that the data is described by a particular model with specified parameters, and then investigate if the data is consistent with those assumptions; now the assumption about the model that describes the data remains, and computations are done to see if the data is consistent with the assumptions a priori. The difference lies in the fact that now the probability distribution of the parameters is investigated given the data [6, 25].

In Appendix A, a more detailed explanation of the Hierarchical Bayes model used to estimate the utilities is presented based on the work made by [25, 26].

In order to detect if the interactions were significant, an interaction analysis was performed using the modified 2-log likelihood test (2LL); this process has demonstrated to be very effective in finding significant interactions [11].

Sample Size

Reducing the possible errors generated by the data in a cost-effective way is of big concern for researchers that want to implement CA and obtain representative results [10]. Current literature on the topic presents several options for determining valid sample sizes in CA. For developing hypothesis for a market, values of 30 to 60 are recommended [10]. A range of 150 to 1200 respondents is suggested for experimentation [10]. [27] Mentions a sample size for full factorial designs between 50-100 respondents to achieve meaningful, robust, and projectable data. Other examples include the work performed by Torres, Paz & Salazar that resulted in a sample size of 246 people using a mathematical formulation. Finally, in a study with limited sample size, 250 respondents for CBC had a stable performance [2]. Based on these results, a sample size of 250 was used for Scenario 1 and 150 for scenario 2 due to the more natural and known nature of the attributes to respondents.

Construction of the survey and data collection

SSI Web module was used to generate the questions for each scenario. One introduction page was made to guide the respondents through the survey. Then, the screening questions were presented to guarantee the respondents were in fact part of the target population.

Validated CA questions were presented afterwards. The validation process consisted of a pilot test where original questions were presented to the 25 respondents to determine if the questions were clear to answer. Results showed that respondents did not understand how to answer to different elicitation tasks presented and got confused about the amount of combinations, tending to think that they were repeated along the survey. In order to solve this issue, a description was added, indicating how to qualify the task with a notice that all of the combinations in the survey were different and at least one level made the difference among them.

All surveys were fielded through the Internet and stored in Sawtooth Software servers. However, given the low rate response of 3% for data collection using Internet [27], the use of tablets was employed to collect data on the field as well.

Results & Discussion

Estimating the Conjoint model and assessing overall fit

The highest incomplete surveys were from the CVA methodology due to the quantity of rates that they involved; in contrast, in CBC choices facilitated the tasks to respondents. The total data collected across surveys are shown in Table 4.

	CBC1	CVA1	CBC2	CVA2	Total
Qualified / Complete	251	250	150	150	801
Disqualified	101	130	106	145	482
Incomplete	100	145	75	104	424
Total	452	525	331	399	1707

Table 4: Number of fielded surveys.

In order to compare similar situations among the used conjoint methodologies, the screening questions obtained from each survey were equivalent. The results from the screening questions corroborated what several authors mentioned, which is that this product is aimed to younger group ages; thus, the inclusion of this group was more pronounced. The results table is exposed in Appendix B.

The utilities, importance values, and goodness-of-fit were obtained for each respondents using SSI web (Sawtooth Software version 8.3.6). In Appendix C, the utility scores for Scenario 1 are presented as reference. Next the results obtained will be exposed for each Scenario.

Scenario 1

Importance values were obtained for both CA and are shown in Figure 2.

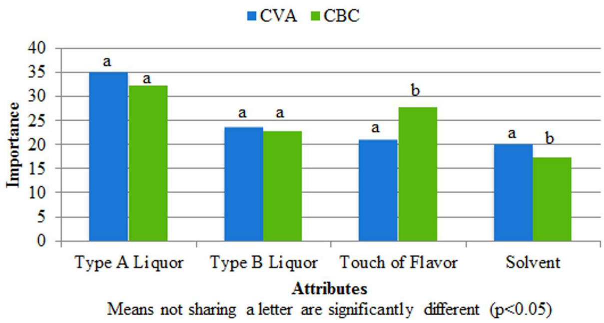

Figure 2: Importance values for Scenario 1.

As it can be seen in Figure 2, the importance values obtained by each CA were not equal. The importance values of Touch of Flavor were different at 95% significance as well as in the Solvent importance values; in addition, considering the order of high to low importance values, the second importance, Type B Liquor for CVA and Touch of Flavor for CBC, indicated that both CA estimations did not lead to similar results.

[8] Mentioned the determination coefficient as a measure of the goodness-of-fit for the estimation of the model as well as the study performed by [2]. The average determination coefficients obtained for Scenario 1 were 0.49 for CVA and 0.63 for CBC. The last value includes significant interaction, which will be expanded later, and represents a much better fitting model estimation for CBC. Overall utilities for both conjoint methodologies

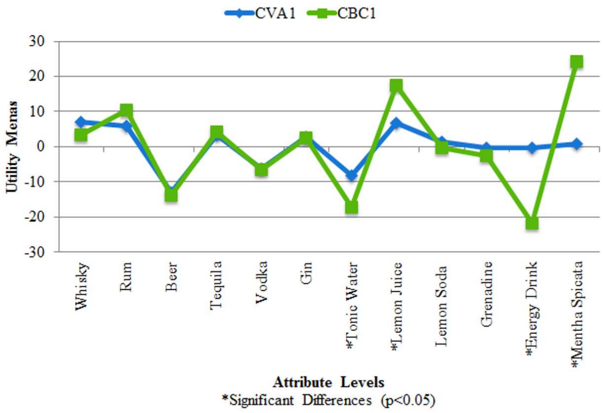

Figure 3: Zero centered average utility values for Scenario 1.

are presented in Figure 3, along with the respective comparisons.

Figure 3 shows the same utility trend; however, the analysis of significant differences at 95% confidence showed that some utility averages were different. Hence, this was another evidence that CBC and CVA did not lead to similar results.

Figure 3 also showed that for Liquor Type A, the least preferred level was Beer for both methodologies and the most preferred ones were Rum and Whisky for CBC and CVA, respectively; for Liquor Type B, the least and most preferred attribute levels were Vodka and Tequila for both CA. Similarly, for Solvent attribute, the least and more preferred levels were Tonic Water and Lemon Juice; for Touch of Flavor, the least and more preferred levels were Energy Drink and Mentha Spicata.

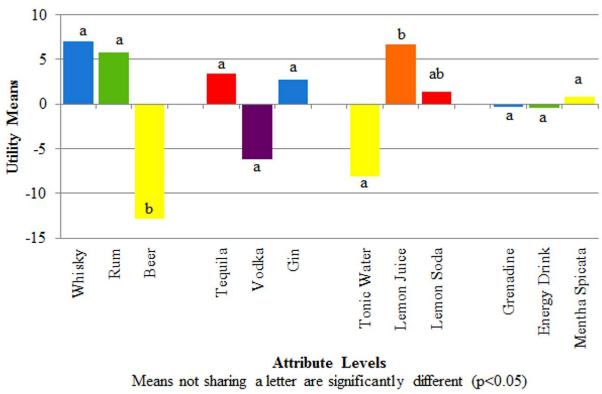

Figure 4: Zero Centered utility values per attribute levels for CVA1.

Similarly, the utility means for each CA methodology were tested to determine if each CA could find significant differences between utilities, yielding in the impact over the total preference model. The results are shown in Figures 4 and 5.

As it can be seen, despite the fact that both CA revealed the same positive and negative trends for all levels, CBC considered more significant differences between levels of the same attribute than the CVA. For example, for

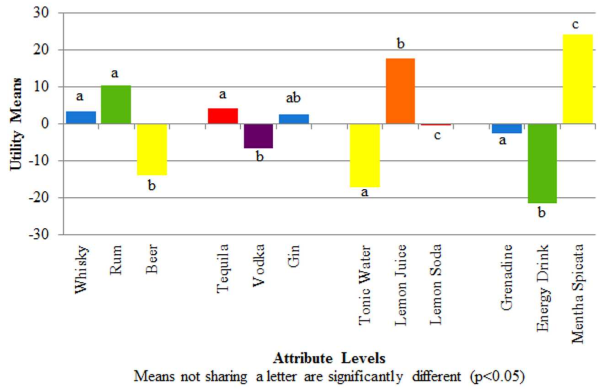

Figure 5: Zero Centered utility values per attribute levels for CBC1.

Type A Liquor, CVA found the same significant differences as CBC; however, for the rest of the attributes, the results were not the same, especially for the Touch of Flavor attribute, where CBC determined all significant differences between its levels, yet the CVA declares none.

One important milestone in the results obtained from the CBC study was to look for possible interactions that could have been affecting the preference model. An interaction plot was constructed using Minitab Software to identify possible interactions and is shown in Figure 6.

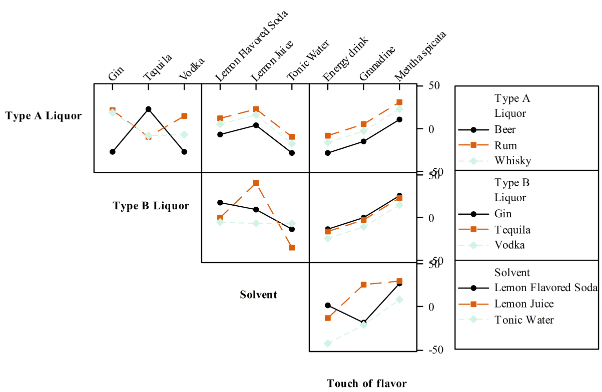

Figure 6: Interaction Plot for Scenario 1 CBC.

Figure 6 showed a potential interaction between the attributes Type A Liquor and Type B Liquor, and similarly, with Type B Liquor and Solvent, as well as Solvent and Touch of Flavor. After running the modified 2LL test, the results found in Figure 6 were corroborated, and the 3 significant interactions agreed. The modified 2LL test results are shown in Table 5 with the respective p-values considered to establish significant interactions at 95% of confidence.

Composition rule model

The utility scores for the significant interactions are shown in Appendix D as reference.

The 81 possible combinations were obtained considering the additive model for CVA and the interaction model

Interaction	Chi-square Value	2LL p-value
Type A Liquor * Type B Liquor	19.8055	0.0005
Type B Liquor * Solvent	17.6101	0.0015
Solvent * Touch of Flavor	11.6912	0.0198
Type A Liquor * Touch of Flavor	6.4600	0.1673
Type B Liquor * Touch of Flavor	5.2857	0.2592
Type A Liquor * Solvent	5.0089	0.2864

Table 5: Interaction Search Tool Results.

for CBC [8]. The total utility scores for each CA methodology were sorted from highest to lowest; the two highest scores are displayed in Table 6 as reference.

CA1	Ord.	Type A Liquor	Type B Liquor	Touch of Flavor	Solvent	Total utility
CVA1	1	Whisky	Tequila	Lemon Juice	Mentha Spicata	17.96
	2	Whisky	Gin	Lemon Juice	Mentha Spicata	17.34
CBC1	1	Beer	Tequila	Lemon Juice	Mentha Spicata	71.33
	2	Beer	Tequila	Lemon Juice	Grenadine	66.05

Table 6: Best two Combinations for CVA and CBC Scenario 1.

The determinant appearance of Beer in both combinations with the higher utility scores for the CBC in contrast with the total absence of this level in the CVA showed a clear interaction that explains preference more accurately in the CBC with the interaction composition rule. When considering interactions, the performance of the CBC was greater in terms of information about the respondents' preferences.

For Scenario 1, the two highest utility scores of CVA were positioned in CBC in places 6 and 15. In contrast, the highest utility scores of CBC were positioned in places 49 and 51 in CVA.

Scenario 2

The same analysis was conducted for Scenario 2. Importance values are shown in Figure 7.

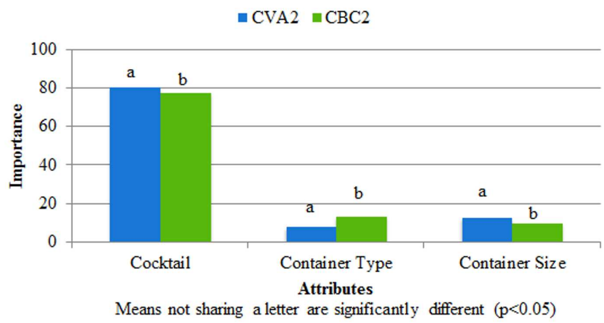

Figure 7: Importance values for Scenario 2.

Cocktail showed, that in both CA methodologies, was by far the most preferred attribute when comparing it with Container Type and Container Size. Also, the order of attribute importance in both CA was the same.

The average correlation coefficients obtained were 0.75 for CVA and 0.67 for CBC, representing a better fitting estimation for CVA. The comparisons of the results for the utilities calculated are shown below.

One of the possible explanations to the sudden change in Scenario 2, was due to three main factors: with lower sample size, CVA performs better estimation predictions than CBC as the literature suggested; no significant interactions were found and the nature of the attributes were more known.

The same analysis of mean comparisons were performed between both CA and are shown in Figure 8 were the results presented significant differences between all utility levels, except for Tequila Sunrise, Vodka Tonic and Michelada.

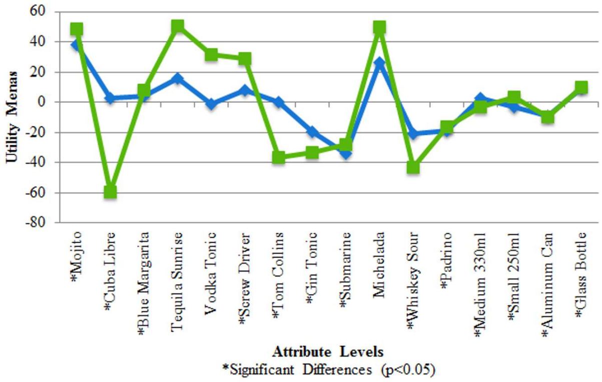

Figure 8: Zero centered average utility values for Scenario 2.

According to Figure 8, the least Cocktail level preferred was Cuba Libre and the most preferred one was Tequila Sunrise for CBC. In the CVA, the least preferred one was Submarine, and the most preferred one was Mojito. For the Container Type attribute, the most preferred level was Glass Bottle for both methodologies. Finally, for Container Size, the most and least preferred attribute levels were contrasted between the two conjoint studies performed.

Like in Scenario 1, a mean comparison between levels for each attribute was considered, revealing the same pattern for CBC being the technique that led to determine significant differences between levels. For both CA, the results were different; for example, for the attribute Cocktail in the CVA, Mojito, Tequila Sunrise, and Michelada had the highest utility scores, in contrast with the results from the CBC where only Mojito had the highest utility score. For Container Size, in CBC there was a significant difference between the values for 250ml with a higher value than for the 330 ml; on the other hand in CVA significant differences were not found between these two levels. At last, for Container Type, significant differences between these two were found for both CA.

The interaction search tool did not find any significant interaction to be included in the model. Table 7 shows the higher utility values obtained after applying the additive model with no interactions.

CA2	Ord.	Cocktail	Container Type	Container Size	Total Utility
CVA2	1	Mojito	Glass Bottle	Medium 330ml	50.04
	2	Mojito	Glass Bottle	Small 250ml	44.21
CBC2	1	Tequila Sunrise	Glass Bottle	Small 250ml	64.18
	2	Michelada	Glass Bottle	Small 250ml	63.37

Table 7: Two Higher Utility values for CA Scenario 2.

Despite the fact that an Additive model was used for both methodologies, due to the absence of significant interactions, the combinations showed in Table 7 were different between both CA.

As reference, for Scenario 2, the two highest utility scores of CVA were positioned in CBC in places 6 and 3. In contrast, the highest utility scores of CBC, were positioned in places 4 and 8 in CVA. As it can be noticed, the highest utility scores in each methodology switched places, considering that Container Size Small was present in three of the four combinations and had an individual estimation that was significantly different with a high utility score.

These results can be explained because the CBC could better represent more respondents' heterogeneity due to the significant interactions encountered within attribute levels. With this differentiation, the CBC can distinguish which attribute levels impact the respondents' preferences better than the CVA.

Conclusions

The inclusion of significant interactions led to different combinations with higher total utility in the composition rule for CBC Scenario1. If these interactions had been excluded from the analysis, the appearance of Beer level would have never been considered due to its negative partial utility score in both methodologies. Therefore, when considering only the main effects, information can be ignored; thus, the analysis of interactions has to be a fundamental part in a CA study.

When the NPD process involves inner characteristics, like in scenario 1, the early knowing of which attribute levels could have significant interactions with others can be a challenging task that could be biased by the researchers. Thus, in this case using a CA methodology that permits the estimation of interactions should always be considered.

In this study, CVA and CBC methodologies did not lead to similar results due to different importance orders attributes and significant differences across utility level estimations. Moreover, CBC reflected more capabilities in finding significant differences within attribute levels that definitely aided to differentiate the levels that contributed to respondent preferences. Furthermore, the goodness-of-fit when interactions were significant was higher for CBC, obtaining a model of preference with more information.

Designing a CBC study is a more complex task that involves specialized software aid for HB estimation and extra experimental design knowledge because it does not only considerate creation of combinations but also the design of sets. Consequently, the study gains the inclusion of significant interactions to predict as accurately as possible the consumer's preferences.

The HB estimation method helped the CBC to obtain results at an individual level with less information from respondents. This led to capture more heterogeneity and significant interactions across the respondents, which is a positive fact about this method. However, the complexity involved with this estimation is high and without the help of specialized software, the estimation could be extensive, meaning that the iterative process performed is computationally extensive and trying to simulate it can take longer periods of time and effort. Thus, using complex estimation methods should be balanced with using specialized software.

The use of choices as an elicitation method instead of ratings, gives less information about the preference of the respondent due to the fact that choices represent the preference of the selected combination in the set, but it does not state how high or low that preference is, like rating tasks do. This fact can be balanced with higher sample sizes, as it can be seen in the results of scenario 2, in which a smaller sample size led to obtain a lower goodness-of-fit for CBC.

The goodness-of-fit of each methodology was compared and analyzed, and higher values were obtained when interactions were included; thus, CBC performed better due to the quantity of variance explained by the model, in contrast with the goodness-of-fit of the CVA methodology. When interactions were not found, the model explained greater variance with the CVA methodology possibly due to the random design method used for estimating the Conjoint instead of the Balance Overlap Method that is better at estimating main effects.

In the CVA methodology, the amount of categories in rating scales can affect the results obtained; thus, the selection of the categories is a variable to be considered. Also, the number of tasks that is directly correlated with the number of profiles tested could overwhelm respondents, thus the use of fractional factorial designs helps the CVA methodology in reducing drastically the number of tasks presented to respondents to obtain better results.

On the other hand, the construction of sets in CBC drastically reduces the number of tasks presented in contrast with CVA, therefore reducing a potential burden of respondents. However, it is important to mention that the cognitive effort in a choice task is greater than in a rating task; therefore, this effect has to be studied in more detail.

Design concerns for the creation of stimuli are a fundamental part of any CA study that needs to be completed to obtain the best possible results. The proper uses of experimental design tools in each of the design phases are key for the results gathered.

The results exposed in this study has led to recommend using CBC methodology, acknowledging the fact that interactions can not be foreseen, and balancing the complexity involved in HB estimation. An important result is the capture of heterogeneity in CBC, which means that the difference across respondents about their preferences was revealed; leading to know which levels in each attribute contributed more to the preference model.

The results obtained cannot be generalized and the guidelines that several authors give should be taken into consideration and analyzed deeply. Nevertheless, the decision has to be supported on quantitative data, and more investigation should be encouraged to see the benefits and drawbacks of using different CA methodologies.

Acknowledgements

This research was submitted in partial fulfillment of the requirements for the undergraduate degree of Industrial Engineering in Universidad San Francisco de Quito (Ecuador).

Sawtooth Software Inc. provided the software as a student research grant to Alvaro Enríquez and Andrés Paredes.

The authors thank Sawtooth Software Inc. for providing the SSI Web to carry out this study.

References

[1] Karniouchina, E.; Moore, W.; van der Rhee, B. 2008. "Issues in the use of ratings-based versus Choice-based conjoint analysis" *in operations management research. European Journal of Operational Research*, 197(1): 340-348.

[2] Jervis, S.; Ennis, J.; Drake, M. 2012. "Comparison of adaptive choice-based conjoint and choice-based conjoint to determine key choice attributes of sour cream with limited sample size". *Raleigh: Journal of Sensory Studies*, 27: 451-562.

[3] Lambin, J. J.; Schulling, I. 2008. "Market-Driven Management: Supplementary web resource material". Palgrave Macmillan: 117-125.

[4] Orme, B.; Chrzan, K. 2000. "An overview and Comparison of Design Strategies for Choice-Based Conjoint Analysis". Sawtooth Software Research Paper Series, Washington, USA. `https://www.sawtoothsoftware.com/download/techpap/desgncbc.pdf`.

[5] Barone, S.; Lombardo, A.; Tarantino, P. 2007. "A Weighted Logistic Regression for Conjoint Analysis and Kansei Engineering". *Quality and Reliability Engineering International*, 23: 689-706.

[6] Sawtooth Software. 2009. "The CBC/HB System for Hierarchical Bayes Estimation Version 5.0 Technical Paper". Sawtooth Software Technical Paper Series. Washington, USA. `https://www.sawtoothsoftware.com/download/techpap/hbtech.pdf`.

[7] Green, P. E. 1990. "Conjoint Analysis in Marketing: New Developments with Implication for Research and Practice". *Journal of Marketing*, 54(4): 1-19.

[8] Hair Jr, J. F.; Black W. C.; Babin, B. J.; Anderson, R. E. (2007). "Multivariate Data Analysis". Prentice-Hall.

[9] Rao, V. R. 2010. "Conjoint analysis". John Wiley & Sons, USA.

[10] Orme, B. 2010. "Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research". Research Publishers LLC: Madison, USA.

[11] Sawtooth Software Inc. 2014. "SSI Web Help Manual". Sawtooth Software: Orem, USA. `https://sawtoothsoftware.com/support/manuals/ssi-web-help`.

[12] Orme, B.; Alpert, M.; Christensen, E. 1997. "Assessing the validity of conjoint analysis - continued". Sawtooth Software Research Paper Series. Washington, USA. `https://www.sawtoothsoftware.com/download/techpap/assess2.pdf`.

[13] Sawtooth Software, Inc. 2013. "The CBC System for Choice-Based Conjoint Analysis Version 8". Utah, USA. `https://sawtoothsoftware.com/download/techpap/cbctech.pdf`.

[14] Brown-Forman. 2011. "Ready-To-Drinks". Brown-Forman: Kentucky, USA. `http://www.ourthinkingaboutdrinking.com/uploadedfiles/PDF_Downloads/Read-y-to-Drinks.pdf`.

[15] [15] Instituto Nacional de Estadìstica y Censos (INEC). 2011-2012. "Encuesta Nacional de Ingresos y Gastos en hogares Urbanos y Rurales (ENIGHUR)". Ecuador en cifras: Quito, Ecuador. `http://www.inec.gob.ec/Enighur_/Analisis_ENIGHUR%202011-2012_rev.pdf`.

[16] Sojo, O. C. 2012. "Patrones de consumo de alcohol en el Ecuador". FLACSO, Investigación: San José, Costa Rica.

[17] King, E. 2012. "Regulatory Impact Statement: Alcohol Reform Bill - Policy Amendments for Inclusion in the Government Supplementary Order Paper". Minitry of Justice: New Zeland. `http://www.treasury.govt.nz/publications/informationreleases/ris/pdfs/ris-justice-arbp-aug12.pdf`.

[18] World Health Organization (WHO). 2014. "Global Alcohol Report - Ecuador". World Health Organization: Geneva, Suiza. `http://www.who.int/substance_abuse/publications/global_alcohol_report/profiles/ecu.pdf?ua=1`.

[19] Addelman, S. 1962. "Symmetrical and Asymmetrical Fractional Factorial Plans". *American Society for Quality*, 4 (1): 47-58.

[20] Xu, H. 2005. "A catalogue of three-level regular fractional factorial designs". *Metrika*, 62: 259-281.

[21] Preston, C. C.; Colman, A. M. 2000. "Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preference". *Acta Phychologica*, 104(1): 1-15.

[22] Igomereho, O. S. 2011. "Conjoint Analysis: A Strategic Tool for Product Research". *International Journal of Economic Development Research and Investment*, 2 (3): 1-9.

[23] Sawtooth Software Inc. 2002. "Conjoint Value Analysis (CVA) Version 3.0". Sawtooth Software Technical Paper Series. Washington, USA. `https://www.sawtoothsoftware.com/download/techpap/cva3tech.pdf`.

[24] Howell, J. 2009. "CBC/HB for beginners. Sequim: Sawtooth Software". Sawtooth Software Research Paper Series. Washington, USA. `http://www.sawtoothsoftware.com/download/techpap/CBCHBbeginners.pdf`.

[25] Allenby, G. M.; Rossi, P. E.; McCulloch, R. E. 2005. "Hierarchical Bayes Models: A Practitioners Guide". Social Science Research Network. Ohio, USA. `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=655541`.

[26] Lenk, P.; DeSarbo, W.; Green, P.; Young, M. 1996. "Hierarchical Bayes conjoint analysis: recovery of part worth heterogeneity from reduced experimental designs". *Marketing Science*, 15 (2): 173-191.

[27] Moskowitz, H. R.; Silcher, M. 2006. "The applications of conjoint analysis and their possible uses in Sensometrics". *Food Quality and Preference*, 17: 145-165.

[28] Chib, S.; Greenberg, E. 1995. "Understanding the Metropolis-Hastings Algorithm". *The American Statistician*, 49(4): 327-335.

Supplementary Material

Appendix A: Hierarchical Bayes

The use of this Bayesian statistical analysis is based on the Bayes Theorem for conditional probabilities and it gives the capability to update the estimations made a priori with information from the data. Only an intuitive explanation is given here, and for more information refer to Sawtooth Software references for Bayesian Data Analysis.

Now, the HB is called "hierarchical" because of the two levels that the estimation has where:

D = A matrix of variances and covariances of the distribution of part worth across individuals.

At the lower level, once the individual's part worth's are given, the probabilities that a respondent chooses a particular alternative is assumed to be governed by a Multinomial Logit Model.

The probability of the ith individual choosing the kth alternative in a particular task is

$$p_k = \frac{e^{x'_k \beta_i}}{\sum_j e^{x'_j \beta_i}}$$

where

$p_k =$ The probability of and individual choosing the kth combination in a particular task.

$x'_j =$ A vector of values describing the jth alternative in that choice task.

At the higher level the individual's part worth is assumed to be described by a multiva riate normal distribution.

$$\beta_i \; Normal(\alpha, D)$$

where

$\beta_i =$ A vector of part worth for the ith individual.

$\alpha =$ A vector of means of the distribution of individual's part worth.

The estimation is an iterative process where the parameters are uploaded until convergence of the parameters are obtained, and it uses a Metropolis Hastings Algorithm which is based on the Markov Chain Monte Carlo methods used to simulate complex, nonstandard multivariate distributions according to [28]. As mentioned before, the introduction of such methods were not possible due to the computational intensiveness required, but in the 90's these type of estimations were beginning to be developed with the technologic advances made and now it is used in the present research with the help of Sawtooth Software. One important aspect to point out about the HB estimation, has the capability to estimate individual part-worth for respondents, which was not possible with MNL or latent class utility estimations by itself when conducting a CBC; and it is this capability that permits a more detailed and accurate contrast with the results obtained from the CVA surveys.

Also, through HB estimation interactions affecting the utility scores can be measured, which in the context of this research is valuable to obtain results that could add information, resulting in a more accurate model.

Appendix B: Screening Questions Responses

11 SCREENING QUESTIONS		CBC1 (n=250) (%)	CVA1 (n=250) (%)	CBC2 (n=150) (%)	CVA2 (n=150) (%)
Male		58	44	52	38
Female		42	56	48	62
Under 18 years old		0	0	0	0
18-24 years old		60	47	69	77
25-29 years old		29	35	11	14
30-34 years old		5	14	6	6
35-39 years old		3	3	7	3
40-44 years old		0	0	6	0
45 years old and older		0	0	0	0
Indicate if you or your house has the following options:	Internet Access	98	97	97	97
	Smartphone	79	90	95	81
	Laptop	86	75	86	91
	None of the Above	0	0	0	0
Indicate what type of floor has at home	Floating floor	29	32	27	34
	Parquet	30	32	26	26
	Board	17	13	20	20
	Wooden Stave	2	1	2	0
	Ceramic	27	15	23	19
	Tile	32	15	30	31
	Vinyl	2	0	0	1
	Porcelain / Marbel	6	7	15	16
	Other	0	0	0	0
Do you drink alcoholic beverages?	Yes	100	100	100	100
	No	0	0	0	0
If consumed, when was the last time you consumed alcohol?	In the last week	69	66	61	66
	In the last two weeks	15	20	26	22
	In the last month	16	15	13	12
	More than one month	0	0	0	0
Would you be willing to try a new blend of alcoholic cocktail?	Yes	100	100	100	100
	No	0	0	0	0
What are the places where you drink alcohol more often?	Bars and discos	76	67	65	70
	At home	26	24	38	27
	Friends home	53	39	69	61
	Restaurants	21	23	23	16
	Other	0	2	0	1
Indicate where do you get alcoholic beverages more frequently	Neighborhood shops	39	24	28	34
	Supermarkets	47	40	61	50
	Friends	27	22	25	33
	Liquor stores	53	52	49	46
	Other	2	4	3	3
Do you want to be contacted for a future consumer test?	Yes	39	40	42	39
	No	61	60	68	61

Appendix C: Utilities Scores for Scenario 1

CVA												
Factor	**Liquor Type A**			**Liquor Type B**			**Touch of Flavor**			**Solvent**		
Average Importance Values	35.06			23.68			21.06			20.20		
Level	**Whisky**	**Rum**	**Beer**	**Tequila**	**Gin**	**Vodka**	**Mentha Spicata**	**Grenadine**	**Energy Drink**	**Lemon Juice**	**Lemon Flavored Soda**	**Tonic Water**
Average Utilities	5.79	7.06	-12.85	3.38	2.76	-6.15	0.79	-0.36	-0.43	6.72	1.40	-8.13
CBC												
Factor	**Liquor Type A**			**Liquor Type B**			**Touch of Flavor**			**Solvent**		
Average Importance Values	32.20			27.68			22.81			17.30		
Level	**Rum**	**Whisky**	**Beer**	**Mentha Spicata**	**Grenadine**	**Energy Drink**	**Tequila**	**Gin**	**Vodka**	**Lemon Juice**	**Lemon Flavored Soda**	**Tonic Water**
Average Utilities	10.39	3.49	-13.88	24.17	-2.58	-21.60	4.10	2.53	-6.63	17.60	-0.38	-17.21

Appendix D: Utility Scores for Significant Interactions

Interaction Utilities		
Interaction Term	**Level Interaction**	**Average Utilities**
Liquor Type A Liquor Type B	Whisky x Tequila	-11.50
	Whisky x Vodka	0.92
	Whisky x Gin	10.58
	Rum x Tequila	-20.16
	Rum x Vodka	16.41
	Rum x Gin	3.75
	Beer x Tequila	31.67
	Beer x Vodka	-17.34
	Beer x Gin	-14.33
Liquor Type B Solvent	Tequila x Tonic Water	-17.00
	Tequila x Lemon Juice	18.25
	Tequila x Lemon Flavored Soda	-1.25
	Vodka x Tonic Water	10.73
	Vodka x Lemon Juice	-6.44
	Vodka x Lemon Flavored Soda	-4.29
	Gin x Tonic Water	6.27
	Gin x Lemon Juice	-11.81
	Gin x Lemon Flavored Soda	5.55
Solvent Touch of Flavor	Tonic Water x Grenadine	7.47
	Tonic Water x Energizer	-13.00
	Tonic Water x Mentha Spicata	5.53
	Lemon Juice x Grenadine	10.89
	Lemon Juice x Energizer	-0.31
	Lemon Juice x Mentha Spicata	-10.58
	Lemon Flavored Soda x Grenadine	-18.36
	Lemon Flavored Soda x Energizer	13.31
	Lemon Flavored Soda x Mentha Spicata	5.05