
ANÁLISIS CONJOINT APLICADO AL AGUA EMBOTELLADA

CONJOINT ANALYSIS APPLIED TO BOTTLED WATER

Autor: Ana María Muñoz Chamorro^a
anamunoz@usantotomas.edu.co

Director: José Fernando Zea Castro^b
josezea@usantotomas.edu.co

Resumen

El *Análisis Conjoint* es una metodología estadística que permite establecer las preferencias de los consumidores hacia un producto, a partir de la combinación de ciertos atributos presentados en diferentes niveles y evaluar la importancia que se le asigna a cada una de estas características. El presente estudio permite conocer los atributos del agua embotellada de mayor preferencia por parte de dos tipos de consumidores: intermediarios y consumidores finales del producto, mediante la aplicación de un análisis conjunto se busca la combinación ideal de atributos para la botella de agua a partir de tres de sus principales características: cantidad de agua en la botella (250 ml, 600 ml y litro), característica del agua (agua pura, agua con gas o agua funcional) y diseño de la botella (botella cilíndrica, botella recta, botella tipo champaña).

El análisis de la información recolectada se realiza mediante la utilización de modelos de elección discreta, los resultados de los modelos estimados indican que una botella de agua pura, en una presentación de botella tipo champaña con un tamaño de 600 ml, tiene mayor preferencia por parte de los dos tipos de consumidores en estudio. Los principales hallazgos del estudio manifiestan la importancia que tienen los atributos para acceder al mercado con un *producto ideal* según el perfil del consumidor al que se quiera llegar.

Palabras clave: análisis conjunto, conjunto de elección discreta, regresión logística.

Abstract

The *Conjoint Analysis* is a statistical methodology that allows to establish the preferences of the users for a set of unique characteristics. The present study allows to know the attributes of the bottled water of greater preference for the part of two types of consumers: intermediaries and final consumers of the product, with the application of a conjoint analysis looks for the ideal combination for the bottle of water from three of its main characteristics: quantity of water in the bottle (250 ml, 600 ml and liter), types of water (pure water, sparkling water or functional water) and bottle design (cylindrical bottle, straight bottle, champagne bottle).

The analysis of the collected information can be done by using discrete choice models, the results of the estimated models indicate that a bottle of pure water, in a champagne bottle presentation with a size of 600 ml, has the highest preference by the types of consumers under study. The main findings of the study show the importance of the attributes to access the market with an *ideal product* according to the profile of the consumer you want to reach.

Keywords: conjoint analysis, discrete choice set, logistic regression.

^aEstudiante. Facultad de Estadística, Universidad Santo Tomás

^bDocente. Facultad de Estadística, Universidad Santo Tomás

1. Introducción

El *Análisis Conjoint* es una metodología aplicada en la actualidad en gran parte de estudios de mercado y de viabilidad de productos debido a que la teoría ha sido complementada cada vez más por diferentes autores, haciendo que sea una teoría completa, de fácil acceso y aplicación en casos prácticos; por lo que su implementación en las diferentes áreas es cada vez más frecuente y con mayor impacto, especialmente en los estudios de mercado. A partir de los resultados obtenidos por este análisis, se realiza el diseño de nuevos productos y el ajuste de características o atributos por aquellos de mayor preferencia por parte de los consumidores.

Actualmente en el campo del marketing se ha hecho visible la necesidad de que los productos que se venden en el mercado estén acordes con las necesidades que los consumidores demandan de éstos, por lo cual las organizaciones ajustan sus productos con las preferencias de los consumidores potenciales o reales, con el fin de hacer productos y empresas más competitivas en el mercado y adicionalmente, los consumidores están satisfechos con los productos y servicios adquiridos.

En el área del marketing es necesario disponer de buenas herramientas de tipo estadístico que permitan responder a las necesidades de las empresas; como es el caso de realizar estudios de competencia, predecir rentabilidades, evaluar el impacto de productos nuevos o existentes en el mercado, y para este caso en concreto, se puede responder a preguntas como: ¿cuál es el diseño o características óptimas de un producto?, ¿se deberá modificar alguna característica con base a las preferencias del consumidor?, adicionalmente, se pueden evaluar los impactos generados por campañas publicitarias y predecir el comportamiento de los consumidores ante cambios en los atributos del producto en estudio.

El agua embotellada es un producto de consumo masivo, que permite la diversificación en cuanto a sus características, tamaños y empaques; en el mercado colombiano existen varias empresas con una amplia gama de productos y subproductos fabricados a partir de agua, aquí surge el motivo de la realización del presente trabajo de grado, el cual será de gran utilidad para una empresa embotelladora de agua al analizar las preferencias de sus clientes hacia su producto, evaluando tres de sus principales atributos: cantidad de agua en la botella, característica del agua y diseño de la botella; a partir de lo anterior se llega a conclusiones sobre cuál sería el *producto ideal*¹ para cumplir con las preferencias de los consumidores actualmente.

¹producto con la mejor combinación posible de niveles de atributos.

2. Objetivos

2.1. Objetivo General

Analizar los atributos de mayor preferencia en el agua embotellada, por parte de los consumidores en la ciudad de Bogotá actualmente, a partir de la aplicación de un análisis conjoint.

2.2. Objetivos Específicos

1. Determinar los atributos específicos que tienen mayor importancia y preferencia por parte de los consumidores al momento de comprar agua en botella en la ciudad de Bogotá.
2. Elegir la combinación de atributos viable de mayor preferencia para el agua embotellada por parte de los dos tipos de clientes (consumidor final y cliente intermediario).

3. Revisión de la literatura

Para el presente trabajo se realizó la revisión de la literatura nacional e internacional, teniendo en cuenta las investigaciones sobre la teoría y las aplicaciones de la metodología de análisis conjunto en casos prácticos; con el fin de tener unas bases sólidas tanto teóricamente como computacionalmente al momento de ejecutar el estudio. Quienes dan inicio a esta metodología son Green y Srinivasan (1978), hacen énfasis en la implementación del análisis conjoint con desarrollos técnicos y aplicaciones para esta metodología en diferentes campos. Después de presentar su estudio llegan a la conclusión de que el análisis conjunto proporciona una metodología útil para evaluar la estructura de las preferencias de los consumidores y cierta capacidad para predecir el comportamiento de los mismos hacia ciertos estímulos, y dependiendo de la combinación de alternativas que el investigador presente para maximizar la validez predictiva del estudio, los autores indican que ésta metodología es más usada en el sector privado, pero tiene gran potencial para aplicarse en el sector público.

Green et al (1990), publican un segundo artículo donde actualizan su versión anterior de 1978, y discuten temas de fiabilidad, validez, simuladores de elección y enfoques alternativos ante la presencia de un gran número de atributos en los análisis conjuntos. A partir de este artículo, se concluye que debe existir una conexión entre la teoría y la práctica para lograr resultados acertados y confiables, lo que hace que esta actualización presente mayores alcances y versatilidad en la aplicabilidad de esta metodología en el ámbito empresarial y gubernamental en el mundo real.

Durante los últimos años se han tenido aportes dedicados a la aplicación del método de una forma totalmente práctica, como lo es el caso de los autores: Adasme, Spiller y Díaz, (2006) hacen uso de la técnica de análisis conjunto para determinar la percepción de los atributos de la frutilla por parte de los consumidores, determinar las principales diferencias percibidas entre la frutilla blanca y roja, y finalmente lograr determinar la intención de compra del producto.

En el estudio anteriormente nombrado se tienen en cuenta tres atributos a medir: tipo de frutilla (blanca o roja), denominación de origen y precio, para estos atributos, los investigadores definieron unos niveles y a partir de las posibles combinaciones se obtienen 16 posibles perfiles, de los cuales fueron ocho los presentados a los consumidores. Los resultados obtenidos indican el mercado potencial para la frutilla blanca y permite hacer recomendaciones a los productores locales sobre las cualidades de la fruta que son más valoradas por el consumidor final (calidad, textura, jugosidad, suavidad de la pulpa, sabor y dulzor).

Adicionalmente a los estudios nombrados anteriormente, esta metodología estadística tiene aplicación para casos en los cuales se quieren seleccionar ciertos productos de forma adecuada, esto lo realizaron los autores: Ramírez, Rondán, Cataluña y Guerrero (2007), quienes aplican la metodología para determinar una franquicia de forma correcta. Los investigadores estudian los franquiciados en cuanto a su perfil y que este sea el esperado por parte de los compradores.

Este estudio les permite llegar a la conclusión que al realizar un análisis conjunto se debe partir de una encuesta que sea representativa y que permita obtener el perfil ideal de franquiciado, la cual tiene un alto costo temporal y de tipo económico, se deben tener en cuenta para estudios posteriores que las preferencias de los entrevistados cambian en el tiempo, por ende, se deberá volver a realizar aplicaciones del análisis conjunto si se quiere en un futuro volver a mirar las preferencias de los franquiciados.

Una investigación mucho más reciente fue elaborada por Arias-Rico, R. (2010) en la cual mediante el análisis conjunto se adecuan las características de las sesiones clínicas a las preferencias de los profesionales. El autor lo realiza mediante un modelo que evalúa la utilidad de los diferentes perfiles de sesiones clínicas.

En este estudio es importante resaltar que las encuestas, se desarrollaron con el fin que el entrevistado organizara los perfiles de las sesiones por orden de preferencia. Se encuentra que los atributos más valorados fueron: duración, estructura de contenido, disponibilidad de referencias bibliográficas, archivos de almacenamiento y multimedia. Para llegar a estos resultados primero fueron definidos los atributos a tener cuenta, junto con sus niveles (valores que puede tomar cada atributo).

La conclusión que presenta el autor, es que la metodología de *Análisis Conjoint* es una herramienta válida para la evaluación de preferencias de un segmento de la población, aunque este tipo de análisis no permite conocer la opinión o el nivel de *satisfacción* de un individuo como tal, pero sí representa una opción viable para conocer valores globales o generales de un servicio o producto y lograr disminuir al mínimo razonable la subjetividad que se produciría.

Esta investigación es un caso donde el análisis conjoint sale del campo del marketing para incorporarse en otras áreas del conocimiento donde es útil para la toma de decisiones, los beneficios de esta técnica se deben a la facilidad de interpretar sus resultados y a la fiabilidad al momento de sacar productos al mercado con cierta combinación de características deseadas y valoradas por el consumidor final.

Finalmente los autores Ramírez H, y Paralara, C.(2016), analizan para el caso de los jóvenes españoles los factores que contribuyen a la elección del proveedor de servicios de internet. Se hace uso del análisis conjoint para analizar el comportamiento de la muestra y se obtiene que el precio y la posibilidad de obtener un terminal móvil son los dos atributos más valorados, y estos resultados de la investigación serán útiles tanto para los usuarios como para las empresas del sector proveedor del servicio de internet en España.

En los casos presentados anteriormente se evidencia la versificación de aplicaciones del análisis conjoint y su utilidad en el mercado, para el agua embotellada no se encontraron referencias publicadas debido a que los análisis de mercadeo realizados a este producto son de carácter privado en Colombia.

4. Caracterización del problema

La iniciativa del presente trabajo nace a partir de la necesidad que actualmente se tiene en el emprendimiento propio: *Green Foods & Fries S.A.S.*, una empresa colombiana dedicada al tratamiento, potabilización, distribución y comercialización de agua embotellada con marca propia *Agua Celestial* o en servicio de maquila (agua personalizada). La compañía nace en marzo de 2016 y ha brindado bienestar a sus clientes y consumidores mediante el producto “Agua potable tratada” con un pH balanceado, libre de minerales, químicos, patógenos y sólidos disueltos en agua.

Dentro de la empresa se llevan a cabo procesos de tratamiento y producción amigables con el ambiente, donde no se hace uso de químicos para tratar el agua, los productos son manipulados bajo estrictos controles de sanidad, la empresa se une al cuidado del medio ambiente, realizando procesos de clasificación de residuos, manejo adecuado de aguas residuales, y los plásticos utilizados en el proceso se reciclan para ser recogidos por empresas que dan un manejo óptimo a estos; disminuyendo así el impacto ambiental negativo del plástico.

Con el presente estudio se busca complementar el trabajo que se viene realizando en la empresa, fortaleciendo el estudio y desarrollo del producto y estableciendo cuáles son los atributos de mayor preferencia en el agua embotellada, por parte de los dos tipos de consumidores en la ciudad de Bogotá actualmente, a partir de la aplicación de un análisis conjoint.

El presente trabajo permite que la empresa llegue a sus clientes con un portafolio de productos más amplio, porque se va a lanzar un nuevo producto y que junto con el que se tiene en el mercado; representen mayor satisfacción para el cliente y el consumidor final, al momento de su elección y consumo al responder con sus preferencias por el agua embotellada. De manera que con este estudio se encuentren las mejores combinaciones de atributos que maximizan la utilidad de los consumidores tenidos en cuenta: clientes intermediarios y consumidores finales.

5. MARCO TEÓRICO

5.1. Análisis Conjoint

El análisis conjoint es desarrollado principalmente en el ámbito del marketing, debido a que es una técnica que permite analizar el sistema de valores implícitos en la elección de una de las alternativas presentes dentro de un conjunto finito de opciones (Hensher DA, Rose JM, Greene WH., 2005). Es una técnica muy útil porque puede llegar a proporcionar estimaciones de la demanda basadas en el diseño del producto o servicio; aunque la calidad de la producción depende en gran medida de la calidad del diseño. Esto significa elegir el atributo correcto o el conjunto de características correcto para usar y garantizar que el diseño de los atributos y sus respectivos niveles cumpla con los objetivos de la investigación y los requerimientos del negocio.

Para llevar a cabo el análisis conjunto existen tres partes fundamentales: el producto o servicio debe poder descomponerse en ciertos atributos y sus respectivos niveles; en segundo lugar, decidir la mejor forma de presentar los perfiles de productos a clientes o consumidores en estudio junto con el método que se utilizará para estimar los productos de mayor preferencia. Por último, elegir el método estadístico para estimar la utilidad de cada atributo y nivel.

Lo anterior depende del número de atributos a considerar, el método de encuestar (escrito, por teléfono o por internet) y el tiempo disponible para aplicación de la encuesta. Adicionalmente se debe considerar el tipo de análisis conjunto que se llevará a cabo, porque cada uno tiene diferentes características en la aplicación, ejecución y análisis de los resultados. A continuación se describen las generalidades de cada uno de los diferentes tipos de análisis conjoint:

1. Análisis de elección (CBC): Choice Based Conjoint Analysis.

CBC es la forma más utilizada actualmente, se basa en un enfoque de perfil completo donde se muestra un nivel de cada atributo en cada alternativa del producto, por lo general cada entrevistado verá de ocho a 12 conjuntos de elección diferentes. Por lo general, CBC muestra más de dos “productos” al mismo tiempo, junto con la alternativa de no elección, que permite que se evalúen decisiones de elección más realistas. El análisis conjunto basado en elección tiende a preferirse si los datos deben ser más sólidos, debido a que los atributos pueden interactuar entre sí, normalmente el CBC se limita como máximo a cinco o siete atributos en estudio y debe analizarse principalmente a nivel agregado.

El valor predeterminado es calcular utilidades e importancias en CBC a través de una muestra como un todo. Para obtener estimaciones a nivel individual, se deben emplear técnicas adicionales para inferir valores individuales. Para casos de segmentación, este tipo de análisis impone algunas limitaciones en los datos de CBC. Sin embargo, el gran beneficio es que los estudios basados en la elección requieren cuestionarios mucho más cortos (de diez a 20 minutos) y se pueden diseñar para que sean en papel.

Las principales ventajas del análisis conjunto basado en la elección, es que le da una mayor solidez a los resultados, junto con un trabajo de levantamiento de información más corto y, por lo tanto, menos costoso. También es favorecido por su rigor académico y porque permite comparaciones con productos fijos que le permiten probar nuevas formulaciones contra una referencia existente o, en la forma más especializada el modelo final puede sintonizarse en gran medida a datos del mundo real aumentando su valor predictivo.

La principal desventaja es la reducida cantidad de atributos posibles a estudiar; a menos que se haga uso de diseños personalizados más complejos que utilicen perfiles parciales, junto con la falta de estimación de utilidades a nivel individual.

2. Perfiles Parciales (ACA): Adaptative Conjoint Analysis.

ACA es un método particularmente usado cuando hay una gran cantidad de atributos que se deben considerar, aunque ACA utiliza los mismos principios generales que CBC, en diseño, implementa-

ción y cálculo es completamente diferente. Mientras que CBC tiene encuestados que seleccionan entre múltiples productos descritos con un conjunto completo de alternativas, ACA muestra descripciones usando dos o tres de los atributos disponibles y es una selección de pares.

adicionalmente, existe un paso de explicación relativamente largo que los encuestados deben completar antes de entrar en la parte principal de elección de “pares”.

El beneficio principal de ACA es que permite utilizar una gran cantidad de atributos (máximo 30) y de niveles (hasta siete por atributo). Sin embargo, ACA requiere una encuesta en computador y debido a la gran cantidad de atributos es común que una entrevista de ACA dure 45 minutos o más, puede llevarse a cabo cara a cara o en línea, porque es difícil realizarse por vía telefónica y en papel no es posible desarrollarlo. Además, se necesita cierto cuidado al elegir y diseñar los atributos para obtener resultados confiables debido a la gran cantidad de atributos en estudio.

3. Perfiles Completos (CVA): Conjoint Value Analysis.

El perfil completo es la forma original de conjoint. Al igual que el análisis conjunto basado en elección, CVA utiliza un número más limitado de atributos para describir el producto o servicio, pero se muestran suficientes tarjetas o alternativas a un encuestado para que se puedan calcular las utilidades a nivel individual. Un diseño factorial fraccionado se utiliza para especificar un conjunto fijo de perfiles que deben mostrarse para el análisis. La dificultad es que esto limita el número de atributos severamente. Es probable que se use con tres o cuatro atributos, pero incluso en esta cantidad, puede implicar la evaluación de 16 a 24 tarjetas de perfiles de producto.

Los beneficios de la aplicación de este tipo de técnica en las investigaciones de producto consisten en medir la importancia latente de los distintos atributos, identificar qué rasgos de los atributos de un producto son más valorados por sus clientes potenciales, estimar las posibilidades de aceptación y éxito antes del lanzamiento de un producto al mercado, calcular cuál puede ser la cuota de mercado de un nuevo producto antes de lanzarlo y tener claro cuáles son los productos más rentables dentro del catálogo, en función de criterios como la utilidad y el costo.

Durante este estudio se aplica el análisis conjunto basado en elección (choice based conjoint (CBC), el cual se emplea cuando el individuo debe elegir una opción dentro de un conjunto de varias alternativas, donde cada una de ellas está formada por ciertos atributos. Los *atributos* son las características que se están evaluando de un producto en el experimento con el fin de saber cómo afectan o influyen en la elección de una botella con agua, el *nivel de atributo* indica los diferentes valores que pueden llegar a tomar cada uno de los atributos en estudio, según Aizaki, 2012 cada atributo puede tener dos o más niveles; cuando se tiene una combinación de niveles de todos los atributos, se denomina *alternativa* y el conjunto de alternativas es un *conjunto de opciones*.

El diseño experimental es la compilación de los conjuntos de opciones, y cada conjunto de opciones es una pregunta donde el individuo va a seleccionar su alternativa de mayor preferencia. El *conjoint* obliga al entrevistado a sospechar productos reales, encontrando aspectos que le gustan y otros que no, así se revelan qué atributos son realmente claves en su toma de decisión al momento de seleccionar un producto.

5.2. Modelo Logit Condicional

El modelo logit condicional es de los modelos más importantes para los datos de respuesta binaria y se ha utilizado para una amplia variedad de aplicaciones cuando las variables explicativas se refieren a atributos de las diferentes alternativas de elección, y no a las características específicas de los individuos, en este caso el valor de cada variable varía para cada alternativa. Inicialmente a modo general se explicará un modelo logit multinomial en términos de una variable latente subyacente para seguir con el modelo de interés.

5.2.1. Un modelo general de elección

Para la toma de decisiones dentro de un ejercicio de elección, como el que se presenta en el desarrollo del trabajo, los individuos realizan sus decisiones bajo un mecanismo de maximización de utilidad, es decir, el individuo elige el producto que cree le producirá mayor satisfacción. Así la alternativa elegida es aquella que presenta los atributos que son más convenientes del producto, y que por ende brindan mayor utilidad al individuo. En este proceso de elección hay un nivel de aleatoriedad dado por que no se conocen ni se pueden estimar todos los aspectos que influye en la toma de decisión de los individuos. Así la utilidad U es descompuesta en dos partes: una determinística η (es observable y explicada por las características de las opciones elegidas) y otra estocástica ϵ (relacionada con las características no observadas).

Suponga que Y_i representa una elección discreta entre J alternativas. U_{ij} representa el valor de la *utilidad* de la j -ésima elección para el i -ésimo individuo. U_{ij} será una variable aleatoria independiente con un componente sistemático η_{ij} y un componente aleatorio ϵ_{ij} tal que:

$$U_{ij} = \eta_{ij} + \epsilon_{ij}. \quad (1)$$

Se asume que los individuos actúan de forma racional, maximizando su utilidad. Por lo tanto, el individuo i elegirá la alternativa j si U_{ij} es mayor que $U_{i1}, \dots, U_{iJ}$. Se debe tener en cuenta que cada elección tiene un componente aleatorio, debido que depende de utilidades aleatorias. La *probabilidad* que el sujeto i elija la alternativa j es:

$$\pi_{ij} = Pr\{Y_i = j\} = Pr\{\max(U_{i1}, \dots, U_{iJ}) = U_{ij}\} \quad (2)$$

Se puede demostrar que si los términos de error ϵ_{ij} tienen distribuciones de valores extremos estándar tipo I con densidad:

$$f(\epsilon) = \exp\{-\epsilon - \exp\{-\epsilon\}\} \quad (3)$$

entonces,

$$\pi_{ij} = \frac{\exp\{\eta_{ij}\}}{\sum \exp\{\eta_{ij}\}}, \quad (4)$$

La cual es la ecuación básica que define el modelo logit multinomial.

En el caso especial donde $J = 2$, el individuo i elegirá la primera alternativa si $U_{i1} - U_{i2} > 0$. si las utilidades aleatorias U_{ij} tienen distribuciones independientes de valores extremos, se puede demostrar que su diferencia tiene una distribución logística, y de allí se obtiene el modelo de regresión logística estándar.

5.2.2. Logit multinomial

En el modelo logit multinomial habitual, las utilidades esperadas η_{ij} son modeladas en términos de las características de los individuos, de modo que,

$$\eta_{ij} = x_i' \beta_j \quad (5)$$

Aquí los coeficientes de regresión β_j pueden interpretarse como una evidencia de los efectos de las covariables sobre las probabilidades de realizar una elección dada (como se ilustró en el caso anterior) o sobre las utilidades subyacentes de las diversas elecciones.

5.2.3. Logit condicional

Este modelo permite modelar las utilidades esperadas η_{ij} en términos de las características de las alternativas en lugar de los atributos de los individuos. si z_j representa un vector de características de la j -ésima alternativa, entonces se obtiene el siguiente modelo:

$$\eta_{ij} = z_j' \gamma \quad (6)$$

Este modelo se conoce como **modelo logit condicional**, y resulta ser equivalente a un modelo log-lineal donde el efecto principal de la respuesta es representado en términos de las covariables z_j .

Se debe tener presente que con J categorías de respuesta, el margen de respuesta se puede reproducir usando cualquier $J - 1$ atributos linealmente independientes de las elecciones. En general, se desea que la dimensión de z_j sea sustancialmente menor que J . En consecuencia, los modelos logit condicionales se utilizan frecuentemente cuando el número de opciones posibles es grande.

En esta etapa del estudio se ejecuta el siguiente modelo mediante la función *clogit* para cada uno de los dos tipos de encuestados (clientes intermediarios y consumidor final), y un modelo general:

$$U_{ji} = ASC_j + \beta_{600ml} X_{600ji} + \beta_{litro} litro_{ji} + \beta_{gas} gas_{ji} + \beta_{funcional} funcional_{ji} + \beta_{recta} recta_{ji} + \beta_{lagrima} lagrima_{ji} \quad (7)$$

A diferencia de la regresión logística ordinaria, no hay intercepto en el modelo, esta es una consecuencia del condicionamiento en los conglomerados: se están comparando los efectos dentro de ellos; y en este caso el intercepto se cancela y hace referencia a la utilidad del producto con los atributos de referencia; en este caso: agua pura, tamaño de 600 ml, y botella tipo convencional.

El cálculo de las utilidades se realiza después de las estimaciones al reemplazar los betas por los resultados del modelo, estas pueden ser positivas o negativas, dependiendo de los signos de los parámetros estimados, lo que indica el nivel de preferencia o de no preferencia por el nivel del atributo en particular; a partir de estos resultados se pueden tomar decisiones en la empresa sobre el producto y sus respectivos atributos.

6. MARCO METODOLÓGICO

Para el presente trabajo se llevaron a cabo las etapas que se evidencian a continuación.

6.1. Identificar y describir los atributos

Los atributos y niveles forman la base fundamental del análisis conjunto; la idea es que un producto o servicio se puede dividir en sus partes constitutivas, esta descomposición de productos en atributos y niveles es una herramienta muy poderosa para examinar lo que ofrece una empresa y lo que debería ofrecer. Para el desarrollo de nuevos productos, la combinación de este desglose de productos con una comprensión de lo que el cliente valora más significa que la empresa puede enfocar sus esfuerzos en aquellas áreas de mayor importancia para los clientes.

En el análisis conjoint, los atributos y niveles tienen que comportarse de cierta manera para que el análisis conjunto sea válido, y en ciertas otras formas para hacer que el conjunto sea útil. En primer lugar, cada atributo debe ser independiente, es decir, los atributos no deben estar correlacionados entre sí. Cada

nivel también debe poder leerse y entenderse por sí mismo, porque lo que el encuestado verá solo los niveles; aunque en el análisis los atributos se usan para ayudar a descomponer un producto cuando se presenta a los encuestados.

Los niveles independientes y legibles son importantes desde el punto de vista del análisis, pero para que el conjunto sea útil también debe garantizar que el rango de atributos cubra todas las áreas que son importantes para el cliente, y que el rango de niveles cubra todas las posibilidades desde el peor de los casos hasta el producto ideal.

En esta etapa se deben identificar grupos focales con personas ampliamente representativas de la población que serán incluidas en la muestra o realizar una prueba piloto con el fin de identificar los atributos relevantes del producto que serán introducidos en el estudio, en esta etapa también es importante decidir cuántos atributos se tendrán en cuenta y determinar los niveles particulares que cada atributo puede tomar; se debe mantener el conjunto de atributos tan simple como sea posible.

Se deben seleccionar atributos que no estén correlacionados entre sí para que no existan combinaciones poco verídicas entre los niveles de atributo, y evitar que algún atributo presente más niveles que los demás porque en el momento del análisis de los resultados se puede llegar a caer en el problema del *efecto del número de niveles*, haciendo que ese atributo tenga mayor importancia que los demás en el estudio.

Para este caso teniendo en cuenta el producto *agua embotellada*, a partir de la experiencia en la venta del producto y del conocimiento de algunos clientes intermediarios, se seleccionaron los siguientes tres atributos del producto y los respectivos niveles que pueden llegar a tomar (ver tabla 1):

ATRIBUTO	DESCRIPCIÓN	NIVELES
Cantidad de agua en la botella	Porción de agua contenida dentro de la botella	250 ml
		600 ml
		litro
Característica del agua	Tipos de agua que se ofrecen en el mercado del agua embotellada	Agua pura
		Agua con gas
		Agua Funcional
Diseño de la botella	Estilo y forma de la botella PET disponible para embotellar agua	Botella cilíndrica
		Botella recta
		Botella tipo champaña

Figura 1: Atributos en estudio. Fuente: *Elaboración propia*

6.2. Desarrollar el diseño experimental

Se utiliza el diseño experimental para construir las alternativas que serán presentadas a los encuestados, en la mayoría de los casos es imposible presentar todas las combinaciones de atributos y niveles; por lo cual el diseño experimental debe utilizarse para identificar subconjuntos de las posibles combinaciones que identifiquen las preferencias de los atributos.

El presente estudio se desarrolló con el software R, se planeó el diseño mediante el paquete “*support.CEs*” el cual proporciona las funciones básicas que respaldan la implementación de los experimentos de elección, la función con la que se creó el diseño fue “*rotation.design*”; basada en una matriz de efectos principales ortogonales.

Para usar la función se deben introducir los siguientes argumentos:

- **attribute.names**: lista con todos los nombres y niveles que toma cada uno de los atributos.
- **nalternatives**: número de alternativas que se presetarán por pregunta, se debe excluir la opción de no elección.

- **nblocks**: número de bloques en los que se debe dividir el diseño de experimentos.
- **row.renames**: se asigna “TRUE” si los nombres de los atributos están incluidos.
- **randomize**: se asigna “TRUE” si la función ejecutará el método de selección.
- **seed**: semilla generadora.

El siguiente paso consiste en elaborar la matriz de diseño, mediante la función *“make.design.matrix”*; a partir del diseño elaborado anteriormente. Se deben tener en cuenta los siguientes argumentos:

- **choice.experiment.design**: contiene el experimento de elección creado con la función anterior.
- **optout**: se asigna “TRUE” si la opción de exclusión está incluida.
- **categorical.attributes**: vector que contiene los nombres de los atributos que son tratados como variables categóricas independientes.
- **continuous.attributes**: se asigna el valor de “NULL” debido a que no se tienen atributos que sean tratados como variables continuas.
- **unlabeled**: variable que describe el tipo de diseño de experimento asignado con la función: *“choice.experiment.design”*.
- **common**: se asigna el valor de “NULL” debido a que no se tiene una combinación fija para todas las preguntas.
- **binary**: se asigna el valor de “FALSE”, si se aplica un modelo logit condicional.

A continuación se presentan los primeros valores de la matriz de diseño, se omite un nivel por cada atributo para evitar multicolinealidad:

BLOCK	QES	ALT	ASC	X600	litro	gas	funcional	recta	lágrima
1	1	1	1	0	1	0	0	1	0
1	1	2	1	0	0	0	1	1	0
1	1	3	1	0	1	1	0	0	0
1	1	4	0	0	0	0	0	0	0
1	2	1	1	1	0	1	0	1	0
1	2	3	1	0	0	0	0	0	0
1	2	4	0	0	0	0	0	0	0
1	3	1	1	0	1	1	0	0	0
1	2	2	1	1	0	0	0	0	1
1	3	2	1	0	0	1	0	0	1

Figura 2: Primeros valores de la matriz de diseño. Fuente: *Elaboración propia*

6.3. Desarrollar el cuestionario

En este tipo de diseños lo ideal es lograr representar una situación real del mercado, para lo cual es aconsejable incluir la opción: *“no elegiría ninguna de las alternativas disponibles”*. En este caso esta opción permite la posibilidad de que cuando los individuos se presentan con alternativas que no son satisfactorias para ellos, puedan optar por no tomar ninguna de ellas. Si no se presenta esta alternativa de exclusión voluntaria, cada elección proporciona información sobre las preferencias, condicionada a la

elección de una de las alternativas, pero no proporciona información sobre si el individuo elegiría o no una de las alternativas. Se considera que en los escenarios de elección del mundo real, las personas no están en una situación de “elección forzada” y tienen la opción de elegir o no elegir. (Champ, Boyle y Brown, 2003).

Haciendo uso de la función “*questionnaire*” se obtiene el cuestionario con tres alternativas y la combinación de niveles de los atributos que se deben presentar a los encuestados, solamente se debe introducir el diseño creado en la primera etapa. A continuación se presentan las tres primeras preguntas de la salida de la función:

```

Question 1
cantidad alt.1 alt.2 alt.3
agua      "litro" "250"  "litro"
botella   "pura"  "funcional" "gas"
          "recta" "recta"  "convencional"

Question 2
cantidad alt.1 alt.2 alt.3
agua      "600"  "600"  "250"
botella   "gas"  "pura"  "pura"
          "recta" "lágrima" "convencional"

Question 3
cantidad alt.1 alt.2 alt.3
agua      "litro"  "250"  "litro"
botella   "gas"    "gas"   "funcional"
          "convencional" "lágrima" "lágrima"

```

Figura 3: Primeros valores del cuestionario. Fuente: *Elaboración propia*

En la siguiente etapa del trabajo se deben transformar estas preguntas del cuestionario en una encuesta como tal y que sea aplicable; de manera que sea de fácil visualización y que sea entendida por los individuos que conforman la muestra. Adicionalmente a las alternativas presentadas anteriormente en la figura 3, se debe agregar la alternativa de no elección dentro de la encuesta.

6.4. Recolectar los datos

En esta etapa del estudio se hace uso de los conjuntos de elección en donde cada individuo debe elegir dentro de un conjunto de alternativas aquella que por sus características se ajuste mejor a sus necesidades y preferencias, es decir, la alternativa que le represente mayor utilidad, o elegir la opción que no prefiere ninguna de las alternativas presentadas en cada una de las preguntas. De esta forma se obliga a elegir una entre las opciones existentes, para el participante se hace familiar a las decisiones que se toman en la vida real, porque al escoger una se renuncia a elegir otra, o por lo menos en un mismo momento del tiempo.

En la imagen inferior aparece un experimento de elección en el que los encuestados tendrían que elegir una de las tres alternativas, existiendo la posibilidad de no elegir ninguna. Cada alternativa está formada por la combinación de diferentes niveles de tres atributos referidos al agua embotellada. Se puede utilizar imágenes de cada nivel en vez de texto, para mayor comprensión por parte de los clientes potenciales.

Para este caso se realizaron nueve tarjetas (preguntas), donde dos de los tres atributos están escritos y uno en imagen (tipo de botella). Cada pregunta contiene cuatro alternativas, y el cuestionario se conforma de nueve preguntas, para un total de 27 alternativas evaluadas por individuo.

A continuación se presenta la pregunta dos del cuestionario para evidenciar la forma en que es presentado el cuestionario a los individuos encuestados:

2. Por favor seleccione su alternativa de mayor preferencia:

Q2	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	600 ml	600 ml	250 ml	
AGUA	Con Gas	Pura	Pura	
BOTELLA				
ELECCIÓN				

Figura 4: Pregunta del cuestionario. Fuente: *Elaboración propia*

Fue importante realizar una prueba piloto para asegurar que los encuestados comprendían claramente la información presentada dentro del cuestionario. Se debe tener en cuenta que el número de registros es diferente al número de encuestados; debido a que a cada encuestado se le hacen varias preguntas y el tamaño de muestra está dado al multiplicar el número de personas encuestadas por el número de preguntas realizadas. En este caso el cuestionario constaba de nueve preguntas y fue aplicado de forma anónima a 90 individuos de la ciudad de Bogotá, obteniendo un número de 810 registros en total.

6.5. Respuestas

Después de ser aplicada la encuesta de forma presencial a los dos tipos de clientes, se deben reunir las respuestas en un archivo de datos en excel.

6.5.1. Base de datos con respuestas

Con las respuestas al cuestionario recolectadas en el numeral anterior, se crea una base de datos que contiene las siguientes variables:

- **ID**: identificador del individuo.
- **BLOCK**: número de la serie de bloques a los que se asigna a cada encuestado.
- **q1, q2, ..., q9**: variables de respuestas a cada una de las nueve preguntas, puede tomar el valor de 1, 2, 3 o 4, dependiendo de la alternativa que el individuo seleccionó como de mayor preferencia.
- **EDAD**: edad del individuo encuestado.
- **SEXO**: género.
- **ESTRATO**: estrato socioeconómico.
- **NIVELEDU**: nivel educativo.
- **PROFESION**: ocupación.

A continuación se presenta como quedó la base de datos con las respuestas de los individuos encuestados:

ID	BLOCK	q1	q2	q3	q4	q5	q6	q7	q8	q9	EDAD	SEXO	ESTRATO	NIVELEDU	PROFESION
1	1	3	2	3	1	2	1	3	3	2	40	0	2	3	3
2	1	1	2	1	2	2	3	3	1	2	48	0	2	2	3
3	1	2	2	4	3	1	1	3	4	1	46	1	2	3	2
4	1	1	2	4	1	4	1	3	3	1	20	1	3	4	2
5	1	2	2	3	3	2	1	3	1	1	26	0	3	5	2
...															
87	1	4	3	4	4	4	4	1	4	2	17	1	3	4	1
88	1	1	3	3	2	1	3	1	2	2	18	0	5	3	1
89	1	1	3	3	2	1	3	1	2	2	18	0	4	3	1
90	1	1	2	1	1	2	1	3	3	2	20	1	3	3	1
91	1	2	2	3	3	2	3	1	1	1	18	0	1	3	1

Figura 5: Respuestas. Fuente: *Elaboración propia*

En este archivo se tienen las respuestas de todo el conjunto de individuos, los primeros registros corresponden a los clientes intermediarios, y los siguientes a los consumidores finales. Existen 91 respuestas en este archivo debido a que uno de los registros marcó nunca haber comprado agua en el mercado, por ende no siguió con la encuesta y este registro fue excluido posteriormente para continuar con el análisis.

6.5.2. Base de datos por tipo de consumidor

Dentro de los 90 individuos encuestados que afirmaron haber comprado agua embotellada, 14 fueron clientes intermediarios (tenderos y propietarios de establecimientos con los que la empresa tiene relaciones comerciales) y los restantes 76 fueron consumidores finales de agua embotellada de aquellos establecimientos. Obteniendo un tamaño de muestra de 126 para los clientes intermediarios, y de 684 para los consumidores finales.

Se realizó la división de la base de datos con el fin de aplicar los siguientes pasos del estudio diferenciando cada uno de los grupos de interés. En seguida se muestra una gráfica por tipo de cliente:

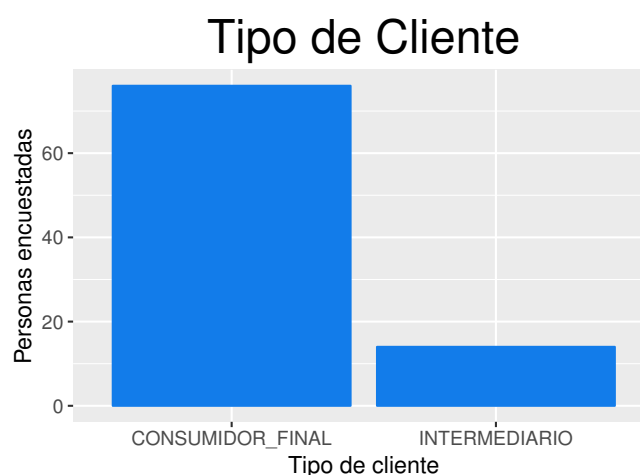


Figura 6: Tipo de cliente. Fuente: *Elaboración propia*

6.5.3. Conjunto de datos final

La base de datos del numeral anterior no está adecuada para ejecutar el modelo, por esto se hace necesario crear un conjunto de datos que combine la información de las respuestas a las preguntas q1 a q9, y la matriz de diseño relacionada a estas preguntas. El conjunto de datos final se realiza mediante la función “*make.dataset*”, la cual tiene los siguientes argumentos:

- **respondent.dataset**: conjunto de datos que contiene las respuestas de la encuesta.
- **choice.indicators**: vector que indica las variables de las preguntas de elección: q1 a q9.
- **design.matrix**: contiene la matriz de diseño.

El conjunto de datos final queda de la siguiente manera:

	ID	EDAD	SEXO	ESTRATO	NIVELEDU	PROFESION	BLOCK	QES	ALT	RES	ASC	X600	litro	gas	funcional	recta	lágrima	STR
1	1	40	0	2	3	3	1	1	1	FALSE	1	0	1	0	0	1	0	101
2	1	40	0	2	3	3	1	1	2	FALSE	1	0	0	0	1	1	0	101
3	1	40	0	2	3	3	1	1	3	TRUE	1	0	1	1	0	0	0	101
4	1	40	0	2	3	3	1	1	4	FALSE	0	0	0	0	0	0	0	101
5	1	40	0	2	3	3	1	2	1	FALSE	1	1	0	1	0	1	0	102
6	1	40	0	2	3	3	1	2	2	TRUE	1	1	0	0	0	0	1	102
7	1	40	0	2	3	3	1	2	3	FALSE	1	0	0	0	0	0	0	102
8	1	40	0	2	3	3	1	2	4	FALSE	0	0	0	0	0	0	0	102

Figura 7: Conjunto de datos final. Fuente: *Elaboración propia*

Se debe tener presente que el orden de las preguntas en el conjunto de datos de respuestas deber ser el mismo que el de la variable **QES** en el conjunto de datos de la matriz de diseño. Este nuevo conjunto de datos tiene las siguientes variables adicionales a las de la base anterior:

- **QES**: variable que describe el número de serie de preguntas de acuerdo con el valor de la variable *BLOCK*.
- **ALT**: variable que describe el número de serie de alternativas según la variable *QES*.
- **RES**: variable lógica que toma el valor de *TRUE* cuando el individuo selecciona esa alternativa y *FALSE* en caso contrario.
- **ASC**: constante específica alternativa.
- **X600**: variable que representa si el nivel de *600 ml* del atributo *cantidad de agua en la botella* está presente en la alternativa evaluada, en este caso toma el valor de 1, y 0 en caso de estar ausente.
- **litro**: variable que representa si el nivel de *litro* del atributo *cantidad de agua en la botella* está presente en la alternativa evaluada, en este caso toma el valor de 1, y 0 en caso de estar ausente.
- **gas**: variable que representa si el nivel de *agua con gas* del atributo *característica del agua* está presente en la alternativa evaluada, en este caso toma el valor de 1, y 0 en caso de estar ausente.
- **funcional**: variable que representa si el nivel de *agua funcional* del atributo *característica del agua* está presente en la alternativa evaluada, en este caso toma el valor de 1, y 0 en caso de estar ausente.
- **recta**: variable que representa si el nivel de *botella recta* del atributo *diseño de la botella* está presente en la alternativa evaluada, en este caso toma el valor de 1, y 0 en caso de estar ausente.

- **lágrima**: variable que representa si el nivel de *botella tipo champaña* del atributo *diseño de la botella* está presente en la alternativa evaluada, en este caso toma el valor de 1, y 0 en caso de estar ausente.
- **STR**: variable que es creada para el argumento *strata* en la función *clogit* para identificar en orden cada combinación de encuestado y pregunta.

En el conjunto de datos no se presentan las variables pertenecientes a todos los niveles de los atributos debido a que para la aplicación del modelo logit condicional se debe excluir un nivel por atributo para que los parámetros del modelo sean estimables.

7. Resultados

En esta parte del documento se incluyen los resultados obtenidos después de todos los procedimientos realizados; iniciando con los análisis descriptivos, análisis univariados, resultados de la aplicación del modelo general y diferenciado por tipo de consumidor y por último cálculo de las utilidades.

7.1. Análisis descriptivos

Los resultados descriptivos de las variables socio-demográficas que se tuvieron presentes en el estudio se presentan a continuación, y son las que permiten mejorar la comprensión de la estructura de las preferencias en estudio en general para todos los encuestados y para cada uno de los dos tipos de clientes:

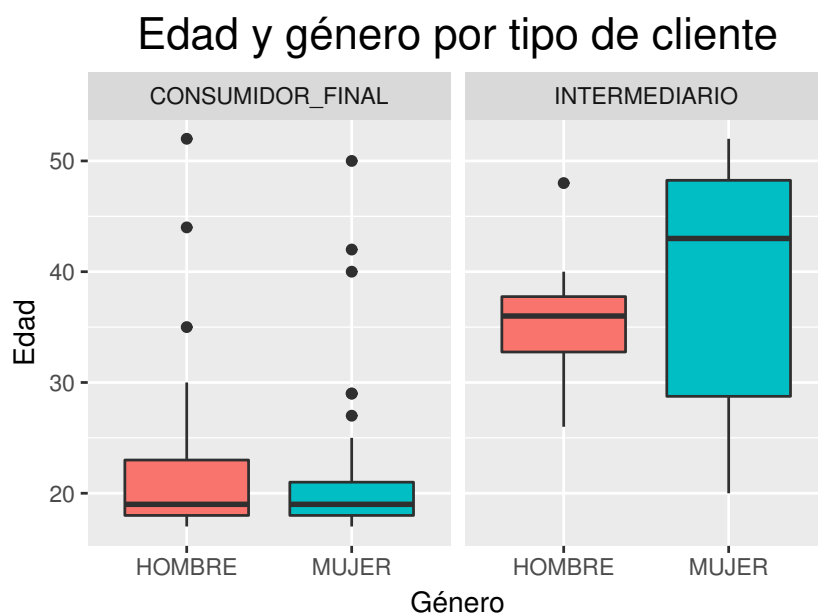


Figura 8: Edad por sexo y tipo de cliente. Fuente: *Elaboración propia*

A partir de la gráfica 8 se observa que para el grupo de consumidores finales las edades son menores con respecto al grupo de cliente intermedio, el promedio de edad para el primer grupo es de 21.7 años; mientras que para el segundo grupo de individuos es de 37 años, esto se debe a que la mayoría de los

clientes en los establecimientos fueron estudiantes, mientras que los dueños de dichos establecimientos son personas con más edad.

A continuación se presentan los gráficos descriptivos de las demás variables socio-demográficas diferenciado por tipo de cliente:

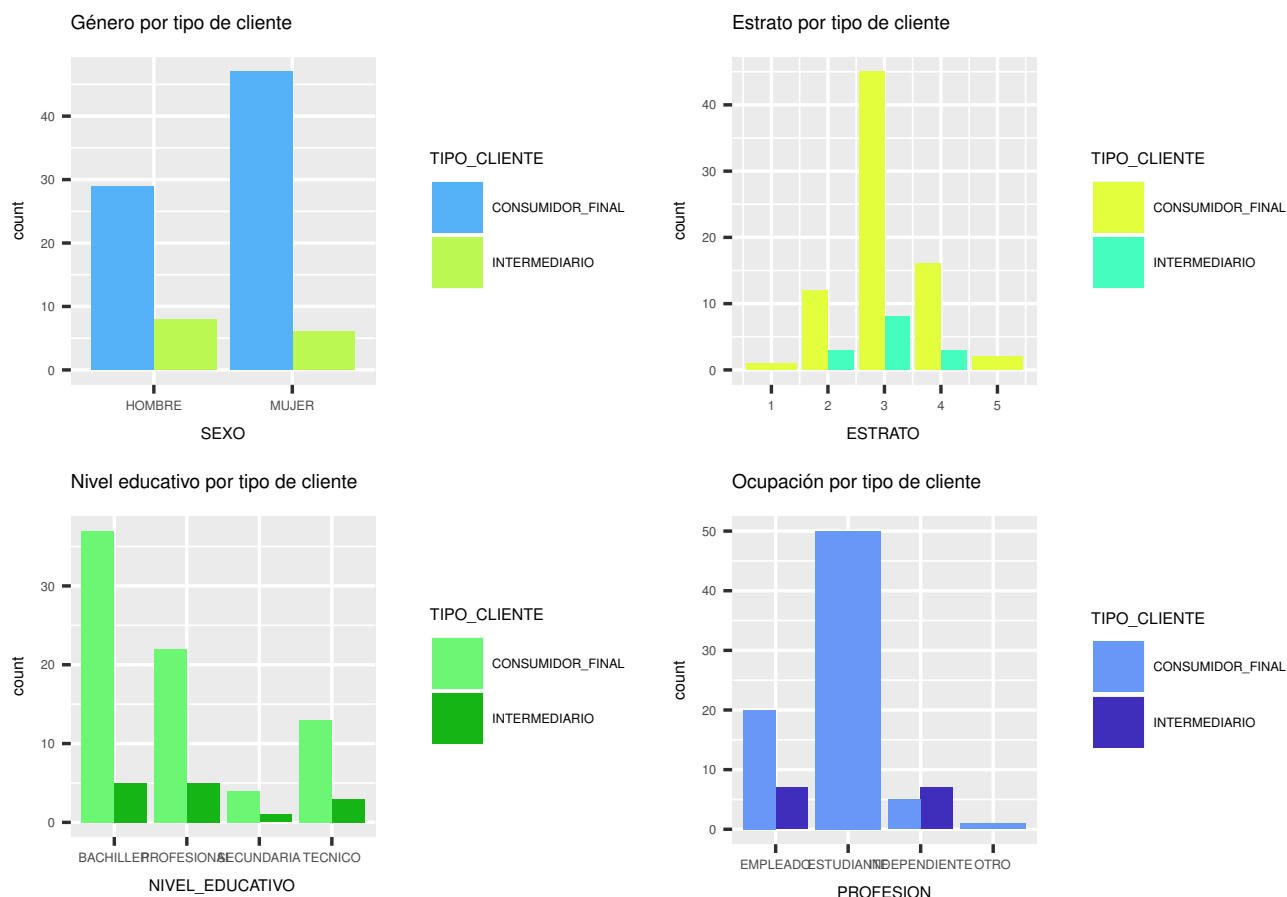


Figura 9: Gráficas descriptivas. Fuente: *Elaboración propia*

En la gráfica 9 se observa que dentro de la muestra hay más mujeres que hombres en los dos tipos de clientes, el estrato socioeconómico predominante es el tres, se observa que la mayoría de personas que participaron en este estudio son bachilleres, seguido de los profesionales y técnicos; por último se evidencia que la mayoría de consumidores finales son estudiantes y empleados, mientras que para el cliente intermediario no existen mayores diferencias entre los que tienen ocupación de empleados y los independientes, esto se debe a que fueron entrevistados algunos propietarios y administradores de los establecimientos.

7.2. Análisis Univariados

El análisis univariado consiste en conocer y determinar los atributos que tienen mayor importancia dentro del conjoint, a partir de la elección de las opciones presentadas en las diferentes tarjetas.

A continuación se muestran los resultados para los clientes intermediarios:

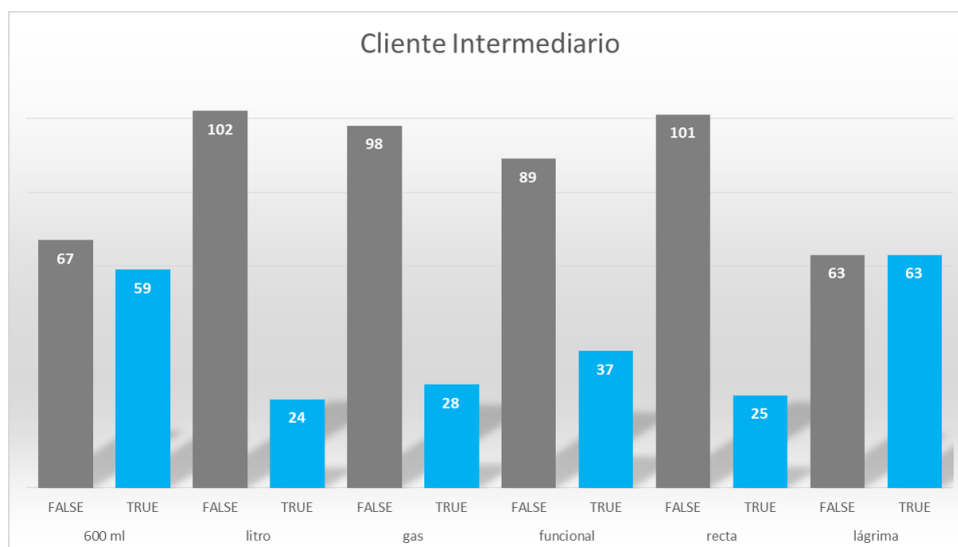


Figura 10: Análisis Univariado clientes intermediarios. Fuente: *Elaboración propia*

En la figura 10 a partir de un análisis univariado, se muestra la elección (TRUE) o no (FALSE) de las alternativas en la variable respuesta con cada uno de los atributos en estudio, se evidencia que para el tipo de cliente intermediario, el atributo que mayor afecta a la elección es que el diseño de la botella sea tipo champaña, seguido por la cantidad de agua en la botella que sea de 600 ml, y el de menor interés es la cantidad de litro.

En seguida se muestran los resultados para los consumidores finales:

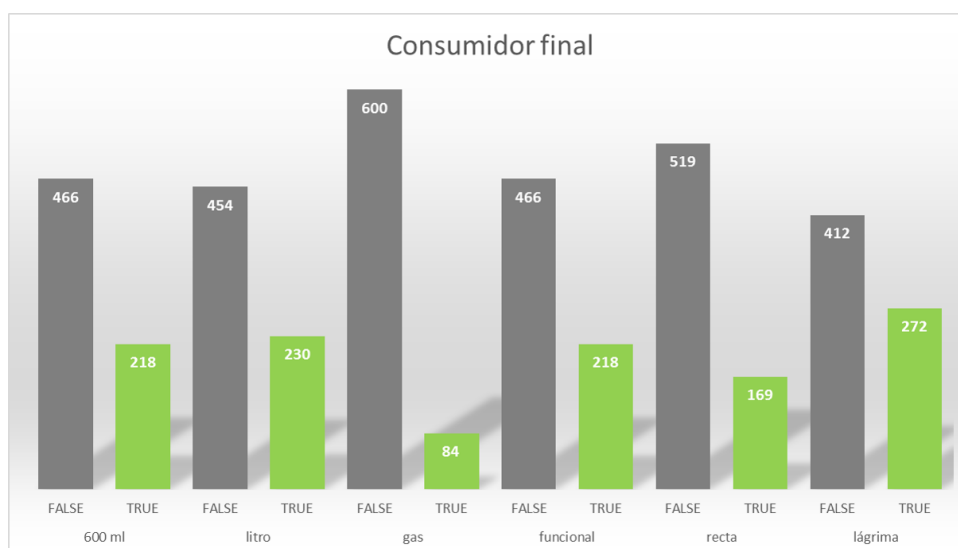


Figura 11: Análisis Univariado consumidor final. Fuente: *Elaboración propia*

Según el mismo análisis univariado aplicado anteriormente de la variable respuesta con cada una de las variables en estudio, se evidencia que para el consumidor final, el atributo que mayor afecta a la elección al igual que para los clientes intermediarios es que el diseño de la botella sea tipo champaña y en segundo lugar que la cantidad de agua en la botella sea de un litro, finalmente el atributo de menor interés es el agua con gas como se observa en la figura 11.

7.3. Resultados aplicación del modelo

7.3.1. Modelo General

Se realizó la aplicación del modelo con todo el conjunto de datos, sin realizar diferenciación por tipo de cliente. En seguida se muestran los resultados obtenidos de la estimación del modelo con 810 registros:

	<i>coef</i>	<i>exp(coef)</i>	<i>se(coef)</i>	<i>z</i>	<i>p</i>
ASC	1,87823	6,54194	0,17283	10,867	< 2e-16 ***
X600	0,48337	1,62153	0,11819	4,090	4.32e-05 ***
litro	0,02295	1,02322	0,11706	0,196	0,84500
gas	-1,83706	0,15929	0,13799	-13,313	< 2e-16 ***
funcional	-0,96967	0,37921	0,11824	-8,201	2.22e-16 ***
recta	-0,07441	0,92829	0,11775	-0,632	0,52700
lágrima	0,44945	1,56745	0,10547	4,262	2.03e-05 ***

Figura 12: Modelo general. Fuente: *Elaboración propia*

La ecuación 7 indica los resultados anteriores:

$$U_{ji} = 1.87823 + 0.48337 X600_{ji} + 0.02295 \text{litro}_{ji} - 1.83706 \text{gas}_{ji} - 0.96967 \text{funcional}_{ji} - 0.07441 \text{recta}_{ji} + 0.44945 \text{lagrima}_{ji} \quad (7)$$

Las estimaciones de los parámetros indican que para la *cantidad de agua en la botella*, al tener 600 ml se incrementan las probabilidades de elección por parte de los encuestados con un factor de 0.48337, lo que significa que esta cantidad parece ser favorable al estar presente en la botella con agua, y su efecto es estadísticamente significativo ($p = 4.32e - 05$). Para el caso del *diseño de la botella*, las probabilidades de elección son 0.44945 veces mayores para la botella de agua con diseño *tipo champaña* en comparación con la botella cilíndrica. Finalmente se puede evidenciar que en la *característica del agua* se tiene mayores probabilidades de elección al presentarle al sujeto *agua pura* en comparación con un *agua con gas* o *funcional*.

Mediante la función *gofm* se obtiene el cálculo de las medidas de bondad de ajuste para el modelo anterior como se presenta en seguida:

Rho-squared	0.1943714
Adjusted rho-squared	0.1881375
Akaike information criterion (AIC)	1823,278
Bayesian information criterion (BIC)	1856,157
Number of coefficients	7
Log likelihood at start	-1122,898
Log likelihood at convergence	-904,6391

Figura 13: Bondad de ajuste del modelo general. Fuente: *Elaboración propia*

7.3.2. Clientes intermediarios

En seguida se muestran los resultados obtenidos del modelo realizado para los clientes intermediarios con 126 registros:

	<i>coef</i>	<i>exp(coef)</i>	<i>se(coef)</i>	<i>z</i>	<i>p</i>
ASC	1,08456	2,95814	0,46091	2,353	0,018619 *
X600	1,36185	3,90340	0,31458	4,329	1,497E-05 ***
litro	-0,56010	0,57115	0,30757	-1,821	0,068601 .
gas	-1,10679	0,33062	0,34572	-3,201	0,001368 **
funcional	-0,79612	0,45108	0,33737	-2,36	0,018285 *
recta	0,01846	1,01863	0,33233	0,056	0,955713
lágrima	1,15643	3,17855	0,30358	3,809	0,0001390 ***

Figura 14: Modelo para los clientes intermediarios. Fuente: *Elaboración propia*

La ecuación 8 indica los resultados anteriores:

$$U_{ji} = 1.085 + 1.362 X600_{ji} - 0.5601 litro_{ji} - 1.107 gas_{ji} - 0.7961 funcional_{ji} + 0.01846 recta_{ji} + 1.156 lagrima_{ji} \quad (8)$$

Las estimaciones de los parámetros indican que para la *cantidad de agua en la botella*, al tener 600 ml se incrementan las probabilidades de elección por parte de los clientes intermediarios con un factor de 1.362, lo que significa que esta cantidad parece ser favorable al estar presente en la botella con agua, y su efecto es estadísticamente significativo ($p = 1,497E - 05$). Para el caso del *diseño de la botella*, las probabilidades de elección son 1.156 veces mayores para la botella de agua con diseño *tipo champaña* en comparación con la botella cilíndrica. Finalmente se puede evidenciar que en la *característica del agua* se tiene mayores probabilidades de elección al presentarle al sujeto *agua pura* en comparación con un *agua con gas* o *funcional*.

Mediante la función *gofm* se obtiene el cálculo de las medidas de bondad de ajuste para el modelo anterior como se presenta en seguida:

Rho-squared	0.2571933
Adjusted rho-squared	0.2171184
Akaike information criterion (AIC)	273,4967
Bayesian information criterion (BIC)	293,3507
Number of coefficients	7
Log likelihood at start	-174,6731
Log likelihood at convergence	-129,7483

Figura 15: Bondad de ajuste del modelo para los clientes intermediarios. Fuente: *Elaboración propia*

7.3.3. Consumidor final

A continuación se muestran los resultados obtenidos del modelo realizado para el tipo de cliente consumidor final, con 684 registros:

	<i>coef</i>	<i>exp(coef)</i>	<i>se(coef)</i>	<i>z</i>	<i>p</i>
ASC	1,9945	7,3482	0,1895	10,526	< 2e-16 ***
X600	0,3534	1,4239	0,1309	2,701	0,00692 **
litro	0,1292	1,1379	0,1292	1,000	0,31745
gas	-2,0147	0,1334	0,1556	-12,949	< 2e-16 ***
funcional	-0,9899	0,3716	0,1290	-7,675	1.65e-14 ***
recta	-0,1027	0,9024	0,1287	-0,798	0,42505
lágrima	0,3578	1,4302	0,1158	3,090	0,00200 **

Figura 16: Modelo para los consumidores finales. Fuente: *Elaboración propia*

La ecuación 9 indica los resultados anteriores:

$$U_{ji} = 1.994 + 0.3534 X600_{ji} + 0.1292 litro_{ji} - 2.015 gas_{ji} - 0.9899 funcional_{ji} - 0.1027 recta_{ji} + 0.3578 lagrima_{ji} \quad (9)$$

Las estimaciones de los parámetros indican que para la *cantidad de agua en la botella*, al tener 600 ml se incrementan las probabilidades de elección por parte de los consumidores finales con un factor de 0.3534, lo que significa que esta cantidad parece ser favorable al estar presente en la botella con agua, y su efecto es estadísticamente significativo ($p = 0.00692$). Para el caso de *característica del agua* se tiene mayores probabilidades de elección al presentarle al sujeto *agua pura* en comparación con un *agua con gas* o *funcional*, siendo estadísticamente significativos los efectos que tienen las dos opciones nombradas anteriormente. Finalmente para el *diseño de la botella*, las probabilidades de elección son 0.3578 veces mayores para la botella de agua con diseño *tipo champaña* en comparación con la botella cilíndrica.

Estos fueron los cálculos obtenidos de las medidas de bondad de ajuste para el modelo anterior:

Rho-squared	0.2057317
Adjusted rho-squared	0.1983495
Akaike information criterion (AIC)	1520,291
Bayesian information criterion (BIC)	1551,986
Number of coefficients	7
Log likelihood at start	-948,2253
Log likelihood at convergence	-753,1453

Figura 17: Bondad de ajuste del modelo para los consumidores finales. Fuente: *Elaboración propia*

En consecuencia, la cantidad de agua en la botella y el diseño de la botella se definen como los atributos de mayor preferencia al momento de la elección de una botella con agua en el mercado, por parte de los dos tipos de clientes en estudio.

7.4. Cálculo de las utilidades

El cálculo de las utilidades se realizó a partir de los modelos descritos anteriormente; un modelo general y dos modelos diferenciando por tipo de cliente. Se realizó un listado en excel, donde se puede elegir el modelo y en listas desplegables se indican las características que se desean tomar, y el programa automáticamente nos indica la utilidad de esa combinación dada.

A continuación se muestra el formato:

MODELO	cantidad de agua en la botella	característica del agua	diseño de la botella	UTILIDAD
general	600 ml	agua pura	botella recta	2,28719
cliente intermediario	600 ml	agua con gas	botella tipo champaña	2,496
consumidor final	600 ml	agua funcional	botella cilíndrica	1,3575

Figura 18: Formato Cálculo de las utilidades. Fuente: *Elaboración propia*

Se debe tener en cuenta que cada vez que se cambie alguno de los posibles niveles de atributo, cambiará la utilidad, ésta puede ser positiva o negativa dependiendo de los signos dados para los parámetros en cada uno de los modelos. Y en los diferentes modelos cambia el orden de preferencia de las combinaciones según el tipo de cliente que se esté analizando. Para observar de una forma más clara esto se construyeron gráficas por modelo de cada una de las alternativas con su respectiva utilidad como se muestra a continuación:

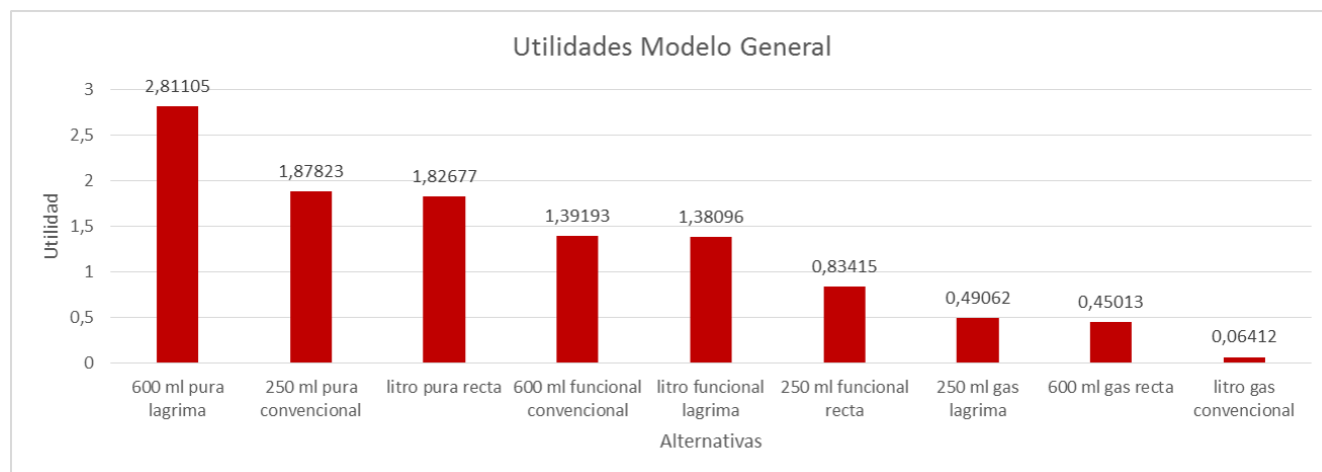


Figura 19: Utilidades Modelo General. Fuente: *Elaboración propia*

Se observa en la figura 19 que la alternativa con mayor utilidad para todos los encuestados es el agua pura en presentación de 600 ml y botella tipo champaña, seguido de agua pura de 250 ml en botella convencional (tipo cilíndrica), en último lugar y con la menor utilidad se encuentra la botella de agua con gas, en presentación de litro y botella convencional.

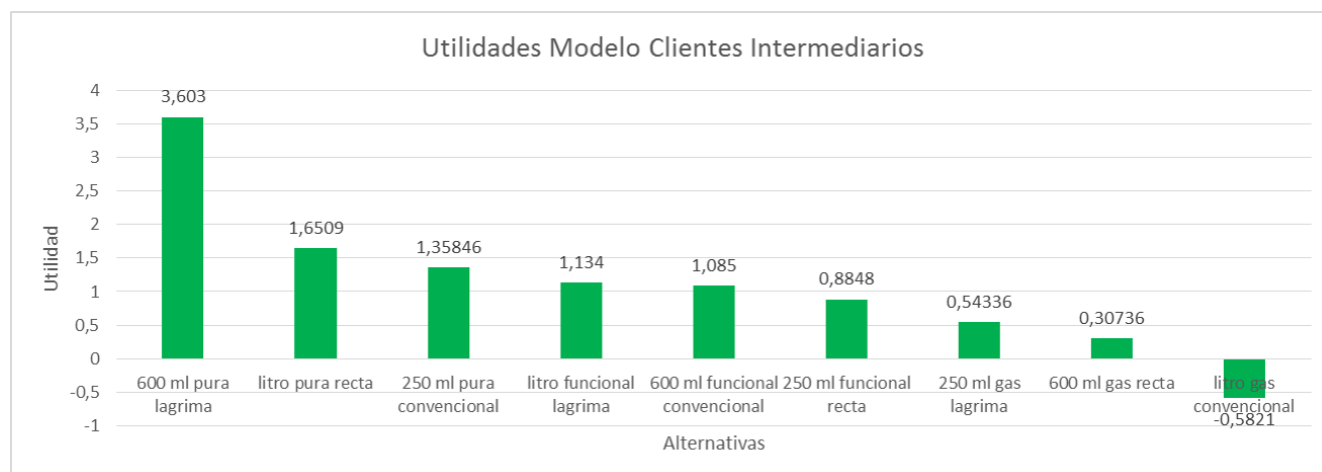


Figura 20: Utilidades Modelo clientes intermediarios. Fuente: *Elaboración propia*

Se observa en la figura 20 que la alternativa con mayor utilidad para los clientes intermediarios es el agua pura en presentación de 600 ml y botella tipo champaña, seguido de agua pura de un litro en botella recta, en último lugar y con la menor utilidad, en este caso un nivel de utilidad negativo; se encuentra la botella de agua con gas, en presentación de litro y botella convencional.

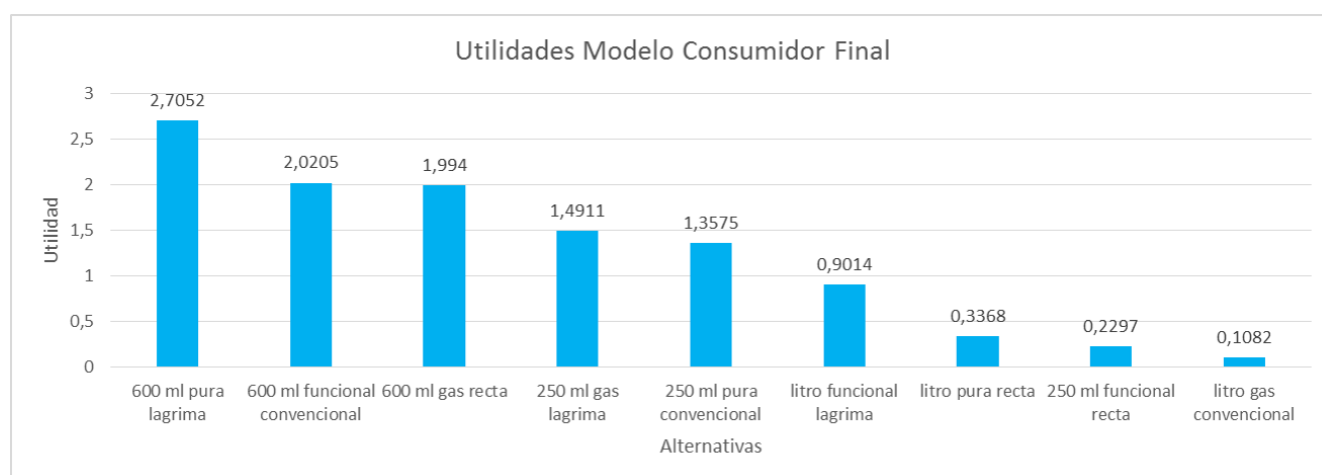


Figura 21: Utilidades Modelo Consumidor final. Fuente: *Elaboración propia*

Se observa en la figura 21 que la alternativa con mayor utilidad para los consumidores finales es el agua pura en presentación de 600 ml y botella tipo champaña, seguido de agua funcional de 600 ml en botella convencional, en último lugar y con la menor utilidad, en este caso un nivel de utilidad negativo; se encuentra la botella de agua con gas, en presentación de litro y botella convencional.

Se evidencia que en los tres modelos la alternativa de mayor y menor preferencia se mantiene en común para todos los tipos de clientes.

8. Discusión

8.1. Conclusiones

El análisis conjoint es una técnica estadística que permite llegar a conclusiones generales sobre un producto o servicio, en este caso: agua embotellada, permitiendo evaluar la importancia relativa de cada atributo que conforma el producto, a través de la elección que realizan los individuos que fueron encuestados para el estudio; este análisis si es aplicado de manera adecuada permite establecer cuáles son las características del producto que aportan valor y cuáles lo extraen al momento de ser elegido por parte de los consumidores en el mercado.

Los hallazgos del trabajo permiten determinar cuáles son los criterios más evaluados al momento de comprar agua embotellada por parte de los consumidores y los atributos específicos de mayor atractivo en el producto. Las estimaciones se realizaron mediante la determinación de modelos estadísticos logit condicionales. De esta manera, los factores que más influyen en la probabilidad de elección de agua embotellada son la cantidad de agua en la botella y el diseño de la botella.

Se puede concluir que según el análisis realizado; la combinación de atributos de mayor preferencia por parte de los dos tipos de clientes es una botella de agua pura de 600 ml con un diseño de botella tipo champaña. Y la alternativa de menor preferencia es la botella de agua con gas, en presentación de litro y botella convencional. Las demás combinaciones del ranking varían dependiendo del tipo de cliente, por lo que se propone a la empresa que se tengan en cuenta las combinaciones con mayor utilidad para el lanzamiento del nuevo producto.

8.2. Limitaciones

Durante el desarrollo del proyecto se presentó la limitante de que el análisis conjoint al ser una técnica donde solo se evalúan algunas de las características del producto que desde el inicio fueron establecidas por el investigador, resulta ser una metodología rígida al no existir la posibilidad de que existan otros factores relevantes que los individuos tendrían en cuenta al momento de tomar la decisión.

8.3. Trabajos futuros

Teniendo en cuenta las limitaciones del presente estudio y las conclusiones a las cuales se llegó, se evidencia que la técnica utilizada tiene aspectos positivos, pero también negativos, para próximos estudios se podría complementar la información con estudios preliminares al público objetivo, mediante entrevistas o encuestas abiertas, que permitan conocer hasta qué punto los atributos y los respectivos niveles que son incorporados en el experimento son verdaderamente relevantes de acuerdo a la estructura de preferencias de los consumidores.

Referencias

- [1] ADASME, C., SPILLER, A., DIAZ, J. (2006). *Determinación de preferencias del consumidor de la región metropolitana hacia la frutilla blanca (fragaria chiloensis). un análisis conjunto y una prueba sensorial*. Economía Agraria, 10.
- [2] AIZAKI, H. (2012). *Basic functions for supporting an implementation of choice experiments in R*. Journal of Statistical Software, Vol 50 doi:10.18637/jss.v050.c02
- [3] ARIAS-RICO, R. (2010). *Aplicación del análisis conjunto en la formación continua de un servicio de farmacia*. Farmacia Hospitalaria, 34(4), 181-187. doi:10.1016/j.farma.2009.11.007
- [4] CHAMP, P. A., BOYLE, K. J., BROWN, T. C. (2003). *A primer on nonmarket valuation* Springer. ISBN 978-94-007-0826-6.
- [5] GREEN, P. E., SRINIVASAN, V. (1978). *Conjoint analysis in consumer research: Issues and outlook*. Journal of Consumer Research, 5(2), 103-123. Retrieved from <http://www.jstor.org/stable/2489001>
- [6] GREEN, P. E., SRINIVASAN, V. (1990). *Conjoint analysis in marketing: New developments with implications for research and practice*. Journal of Marketing, 54(4), 3-19. doi:10.2307/1251756.
- [7] HENSHER DA, ROSE JM, GREENE WH (2005). *Applied Choice Analysis: A Primer*. Cambridge. University Press, Cambridge, UK.
- [8] RAMÍREZ HURTADO, J. M., RONDÁN CATALUÑA, F. J., GUERERO CASAS, F M ^A. (2007). *Selección de franquiciados mediante simulación con análisis conjunto*. Cuadernos De Economía Y Dirección De La Empresa, 10(31), 171-201. doi:10.1016/S1138-5758(07)70087-7
- [9] RAMÍREZ HURTADO, J. M., PARALERA MORALES, C. (2016). *Preferencias de los estudiantes universitarios en la elección del proveedor de internet*. Revista Latina De Comunicación Social, (71), 413-427.

9. Anexo 1: Código en R

```

library(support.CEs)
library(survival)
library(gridExtra)
library(grid)
library(ggplot2)
library(lattice)
library(readr)
library(ggplot2)

# diseño
des1 <- rotation.design(
  attribute.names = list(
    cantidad = c("250", "600", "litro"),
    agua = c("pura", "gas", "funcional"),
    botella = c("convencional", "recta", "lágrima")),
  nalternatives = 3,
  nblocks = 1,
  row.renames = TRUE,
  randomize = TRUE,
  seed = 987)

# cuestionario
questionnaire(choice.experiment.design = des1)

# nombrar el diseño
choice.experiment.design = des1

# matriz de diseno
desmat1 <- make.design.matrix(
  choice.experiment.design = des1,
  optout = TRUE, #la opción de exclusión SI esta incluida
  categorical.attributes = c("cantidad", "agua", "botella"),
  continuous.attributes = NULL,
  unlabeled = TRUE,
  common = NULL,
  binary = FALSE)

# subir el archivo de respuestas COMPLETO
RESPUESTAS <- read_delim("~/ANA/Clases/tesis/RESPUESTAS.csv",
                        ";", escape_double = FALSE, trim_ws = TRUE)

## modelo completo
# haciendo el conjunto de datos adecuado para el modelo
dataset1 <- make.dataset(
  respondent.dataset = RESPUESTAS,
  choice.indicators =
    c("q1", "q2", "q3", "q4", "q5", "q6", "q7", "q8", "q9"),
  design.matrix = desmat1)

# MODELO LOGIT general
clogout1 <- clogit(RES ~ ASC + X600 + litro + gas + funcional + recta + lágrima
                  + strata(STR), data = dataset1)

summary(clogout1)
gofm(clogout1)

# BASE DE DATOS DE LOS CLIENTES
RESPUESTAS_CLIENTES = RESPUESTAS[1:14,]

# haciendo el conjunto de datos adecuado para el modelo
dataset1_CLIENTES <- make.dataset(
  respondent.dataset = RESPUESTAS_CLIENTES,
  choice.indicators =
    c("q1", "q2", "q3", "q4", "q5", "q6", "q7", "q8", "q9"),

```

```

design.matrix = desmat1)

# MODELO LOGIT PARA CLIENTES
clogout1_CLIENTES <- clogit(RES ~ ASC + X600 + litro + gas + funcional + recta + lágrima
+ strata(STR), data = dataset1_CLIENTES)
summary(clogout1_CLIENTES)
gofm(clogout1_CLIENTES)

# BASE DE DATOS PARA CONSUMIDOR
RESPUESTAS_CONS = RESPUESTAS[15:90,]

# haciendo el conjunto de datos adecuado para el modelo
dataset1_CONS <- make.dataset(
  respondent.dataset = RESPUESTAS_CONS,
  choice.indicators =
    c("q1", "q2", "q3", "q4", "q5", "q6", "q7", "q8", "q9"),
  design.matrix = desmat1)

# MODELO LOGIT PARA CONSUMIDOR FINAL
clogout1_CONS <- clogit(RES ~ ASC + X600 + litro + gas + funcional + recta + lágrima
+ strata(STR), data = dataset1_CONS)
summary(clogout1_CONS)
gofm(clogout1_CONS)

## ANALISIS DESCRIPTIVOS

descriptivos <- read.csv("~/ANA/Clases/tesis/ARCHIVOS TESIS/descriptivos.csv", sep=";")

# tipo de cliente

cliente <- ggplot(data = descriptivos, aes(x = factor(TIPO_CLIENTE))) +
  theme (text = element_text(size = 12)) +
  geom_bar(stat = "count", fill = "#207EE9", colour = "#207EE9", position = position_dodge()) +
  ggtitle("Tipo de Cliente") +
  theme(legend.position = "bottom") +
  theme (plot.title = element_text(size=rel(2),
                                   hjust=0.5,
                                   face="plain",
                                   lineheight=1.5)) +
  labs(x = "Tipo de cliente", y = "Personas encuestadas")

cliente

# edad por tipo de cliente y sexo

desedadygen <- ggplot(descriptivos, aes(x=factor(SEX0), y=EDAD, fill=SEX0)) +
  geom_boxplot() +
  facet_grid(. ~ TIPO_CLIENTE) +
  ggtitle("Edad y género por tipo de cliente") +
  labs(x = "Género", y = "Edad") +
  theme (plot.title = element_text(size=rel(1.5),
                                   hjust=0.5,
                                   face="plain",
                                   lineheight=1.5)) +
  theme(legend.position="none")

desedadygen

# sexo por tipo de cliente

sexocliente <- ggplot(descriptivos, aes(x=SEX0, fill=TIPO_CLIENTE)) +
  geom_bar(stat="count", position=position_dodge()) +
  ggtitle("Género por tipo de cliente") +
  scale_fill_manual(values=c("#59B4F7", "#BBF759")) +

```

```

    theme (text = element_text(size=10), plot.title = element_text(hjust=2))

sexocliente

# estrato por tipo de cliente

estratocliente <- ggplot(descriptivos, aes(x=ESTRATO, fill=TIPO_CLIENTE)) +
  geom_bar(stat="count", position=position_dodge()) +
  ggtitle("Estrato por tipo de cliente") +
  scale_fill_manual(values=c("#E2FD47", "#47FDC1")) +
  theme (text = element_text(size=10), plot.title = element_text(hjust=2))

estratocliente

# nivel educativo por tipo de cliente

nivelcliente <- ggplot(descriptivos, aes(x=NIVEL_EDUCATIVO, fill=TIPO_CLIENTE)) +
  geom_bar(stat="count", position=position_dodge()) +
  ggtitle("Nivel educativo por tipo de cliente")+
  scale_fill_manual(values=c("#6AF677", "#16B624")) +
  theme (text = element_text(size=10), plot.title = element_text(hjust=5))

nivelcliente

# ocupación por tipo de cliente

ocuciente <- ggplot(descriptivos, aes(x=PROFESION, fill=TIPO_CLIENTE)) +
  geom_bar(stat="count", position=position_dodge()) +
  ggtitle("Ocupación por tipo de cliente")+
  scale_fill_manual(values=c("#6A99F6", "#4433BB")) +
  theme (text = element_text(size=10), plot.title = element_text(hjust=3))

ocuciente

grid.arrange(sexocliente, estratocliente, nivelcliente, ocuciente, ncol=2)

# edad promedio por tipo de cliente

d <- data.frame(descriptivos$TIPO_CLIENTE, descriptivos$EDAD, descriptivos$SEXO)

with(d, aggregate(d$descriptivos.EDAD, list(d$descriptivos.SEXO, d$descriptivos.TIPO_CLIENTE), mean))
with(d, tapply(d$descriptivos.EDAD, list( d$descriptivos.TIPO_CLIENTE), mean))

# análisis univariados
# Un primer análisis univariado:
# Cómo la escogencia se afecta por el nivel del precio CLIENTES

pander(xtabs(RES ~ X600, data=dataset1_CLIENTES))
pander(xtabs(RES ~ litro, data=dataset1_CLIENTES))
pander(xtabs(RES ~ gas, data=dataset1_CLIENTES))
pander(xtabs(RES ~ funcional, data=dataset1_CLIENTES))
pander(xtabs(RES ~ recta, data=dataset1_CLIENTES))
pander(xtabs(RES ~ lágrima, data=dataset1_CLIENTES))

# Cómo la escogencia se afecta por el nivel del precio CFINALES

pander(xtabs(RES ~ X600, data=dataset1_CONS))
pander(xtabs(RES ~ litro, data=dataset1_CONS))
pander(xtabs(RES ~ gas, data=dataset1_CONS))
pander(xtabs(RES ~ funcional, data=dataset1_CONS))
pander(xtabs(RES ~ recta, data=dataset1_CONS))
pander(xtabs(RES ~ lágrima, data=dataset1_CONS))

```

10. Anexo 2: Cuestionario

A continuación se ilustra las preguntas que fueron presentadas a los encuestados:

1. Pregunta 1:

1. Por favor seleccione su alternativa de mayor preferencia:

Q1	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	Litro	250 ml	Litro	
AGUA	Pura	Funcional	Gas	
BOTELLA				
ELECCIÓN				

2. Pregunta 2:

2. Por favor seleccione su alternativa de mayor preferencia:

Q2	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	600 ml	600 ml	250 ml	
AGUA	Con Gas	Pura	Pura	
BOTELLA				
ELECCIÓN				

3. Pregunta 3:

3. Por favor seleccione su alternativa de mayor preferencia:

Q3	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	Litro	250 ml	Litro	
AGUA	Con Gas	Con Gas	Funcional	
BOTELLA				
ELECCIÓN				

4. Pregunta 4:

4. Por favor seleccione su alternativa de mayor preferencia:

Q4	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	250 ml	Litro	250 ml	
AGUA	Con Gas	Funcional	Funcional	
BOTELLA				
ELECCIÓN				

5. Pregunta 5:

5. Por favor seleccione su alternativa de mayor preferencia:

Q5	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	Litro	600 ml	Litro	
AGUA	Funcional	Funcional	Pura	
BOTELLA				
ELECCIÓN				

6. Pregunta 6:

6. Por favor seleccione su alternativa de mayor preferencia:

Q6	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	600 ml	Litro	600 ml	
AGUA	Pura	Pura	Funcional	
BOTELLA				
ELECCIÓN				

7. Pregunta 7:

7. Por favor seleccione su alternativa de mayor preferencia:

Q7	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	250 ml	600 ml	600 ml	
AGUA	Pura	Con Gas	Pura	
BOTELLA				
ELECCIÓN				

8. Pregunta 8:

8. Por favor seleccione su alternativa de mayor preferencia:

Q8	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	600 ml	Litro	250 ml	
AGUA	Funcional	Con Gas	Con Gas	
BOTELLA				
ELECCIÓN				

9. Pregunta 9:

9. Por favor seleccione su alternativa de mayor preferencia:

Q9	ALTERNATIVA 1	ALTERNATIVA 2	ALTERNATIVA 3	NO PREFIERE NINGUNA
CANTIDAD	250 ml	250 ml	600 ml	
AGUA	Funcional	Pura	Con Gas	
BOTELLA				
ELECCIÓN				

11. Anexo 3: Carta de habeas data

Bogotá D.C., Diciembre 4 de 2017

Señores

FACULTAD DE ESTADÍSTICA
UNIVERSIDAD SANTO TOMÁS
Bogotá D.C.

REF: Manejo de información

Respetados señores:

Bajo los términos de derechos de autor, procedencia y manejo de los datos en el presente trabajo de grado:

1. Afirmo que el material es original; que no contiene ningún contenido que sea difamatorio o ilegal o que invade la privacidad individual o infrinja cualquier derecho de propiedad o cualquier derecho legal; garantizo que no se ha asignado previamente ningún derecho protegido por derechos de autor al trabajo.
2. Afirmo que los datos empleados en el presente trabajo, fueron recolectados por el autor, a través de encuestas escritas y aplicadas presencialmente a los participantes; manteniendo la anonimidad de los mismos dentro del estudio.
3. Garantizo que el Trabajo no ha sido publicado en ninguna otra parte, en su totalidad o en parte, y que no he dado permiso a nadie más para publicar o explotar cualquier parte de la Obra.

Título del trabajo: Análisis Conjoint aplicado al agua embotellada.

Autor: Ana María Muñoz Chamorro

Director: José Fernando Zea



ANA MARÍA MUÑOZ CHAMORRO

C.C. 1.018.466.886 de Bogotá

Estudiante de Estadística