

Santiago, Agustín; Juárez, Octaviano; Bouza, Carlos

**Working Paper**

## Un estudio de las potencialidades del mercado para el consumo de agua purificada de la fuente de Azinyahualco

Economic Analysis Working Papers, No. 2008,12

**Provided in Cooperation with:**

Economists Association of A Coruña

*Suggested Citation:* Santiago, Agustín; Juárez, Octaviano; Bouza, Carlos (2008) : Un estudio de las potencialidades del mercado para el consumo de agua purificada de la fuente de Azinyahualco, Economic Analysis Working Papers, No. 2008,12, Colegio de Economistas de A Coruña, A Coruña

This Version is available at:

<https://hdl.handle.net/10419/43406>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **UN ESTUDIO DE LAS POTENCIALIDADES DEL MERCADO PARA EL CONSUMO DE AGUA PURIFICADA DE LA FUENTE DE AZINYAHUALCO**

**Agustín Santiago\*, Octaviano Juárez\* y Carlos Bouza\*\***

**\*Universidad Autónoma de Guerrero**

**\*\*Universidad de La Habana**

## **Resumen**

Actualmente la comercialización del agua embotellada ha cobrado importancia en la mayoría de países de América latina, realizando fuertes inversiones en su desarrollo tanto el capital transnacional como los capitales nativos. El presente trabajo presenta un estudio de mercado solicitado por los ejidatarios del poblado Azinyahualco, Municipio de Chilpancingo, estado de Guerrero, México. Dichos ejidatarios pretenden comercializar el agua proveniente de una fuente natural en la zona urbana y suburbana del municipio de Acapulco, donde se concentra más del 25% de la población del estado de Guerrero. El objetivo de este trabajo es conocer el estado del mercado, para ver si sería rentable la puesta en marcha de una planta purificadora con consecuente introducción de una nueva marca de agua. Se diseñó una encuesta por muestreo. El tamaño de la muestra fue de 2496 viviendas. Se determinaron las características de la oferta, las regularidades del mercado. Respecto a la población de consumidores se estableció su estructura y hábitos de consumo. Se estimó el mercado potencial y se determinaron aquellos nichos donde pudiera penetrar el nuevo producto. Se realizó el análisis de la fidelidad de los consumidores a la marca de su preferencia. La técnica de modelos de árbol de clasificación fue utilizada para establecer los elementos necesarios para el diseño de una estrategia de mercadeo para una nueva marca de agua.

## **Abstract**

Nowadays the sell of bottled water has increased its importance in the majority of the countries of Latin America, important investments of transnational and native capital are being made for its development. This paper presents the results of a study of the market needed by the farmers of the town of Azinyahualco, Municipal of Chilpancingo, Guerrero State, México. These land owners want to commercialize the water, produced by a natural source, in the urban and sub-urban zones for the municipal Acapulco, where more than the 25% of the population of the State of Guerrero live. The objective of this paper is to obtain the needed knowledge of the market, to evaluate the economic feasibility of the development of a new brand, for establishing a purifier plant for introducing it in the market. A survey sampling was designed. The sample size was of 2496 households. The characteristics of offer, the market regularities were determined. The structure and consuming habits were established. The potentialities of the market were estimated and the market niches, where the product could be introduced, were determined. The loyalty of the consumers for the brands was analyzed. Decision trees techniques were used for designing a marketing strategy for the new purified water brand.

## **1. Introducción.**

El mercado de las aguas purificadas es de especial importancia en México y en muchos países de América Latina. Es común que las familias la adquieran para su consumo y para la preparación de sus alimentos. En el país, existen numerosas Plantas purificadoras comerciales, algunas de ellas transnacionales, en las que el agua extraída de pozos o fuentes habilitados y autorizados, es trasladada a grandes o pequeñas plantas purificadoras locales. Este es un buen negocio para las grandes empresas transnacionales establecidas en México y para las pequeñas empresas locales, toda vez que las impurezas de las aguas superficiales y aún las subterráneas, presentan niveles de contaminación que puede significar un riesgo para la salud de quien decide consumirlas directamente. En años recientes, este es un negocio en boga no solo en México, sino en muchos países de América Latina y el mundo [Mena Martínez, Argentina (2008) y Ruiz Talavera, Argentina (2008), Laura Ferrarese, Chile (2006), Columna de Opinión (2008)]. Este es un negocio a todas luces lucrativo, tanto que las empresas transnacionales están en constante lucha por apoderarse del mercado de la región, tal es el caso de la empresa Pepsi Cola que tiene en el mercado el agua “Electropura” y la empresa CocaCola con las marcas Ciel y Manantial. Los empresarios locales han percibido las bondades de este negocio y han proliferado empresas purificadoras en los principales centros urbanos, provocando el riesgo de saturación del mercado.

El método de purificación empleado por la mayoría de empresas purificadoras de agua de la región, dado que la dureza es baja, transcurre de la siguiente manera: el agua llega a plantas purificadoras donde se le adiciona cloro al 12 o 13% por un periodo de 30 minutos y posteriormente es desplazada a unos cilindros que contiene medios filtrantes y donde se retienen de sólidos disueltos e impurezas. Generalmente en el primero se filtra por arena, en el segundo por un filtro de carbón activado y finalmente por un cilindro que contiene agua con sal, llamada salmuera. El proceso continúa y el agua filtrada se hace pasar por una cámara de luz ultravioleta que juega su rol de germicida, finalmente se conduce a la llenadora. En los casos en los cuales la dureza es muy elevada, suelen utilizarse métodos de separación de sales como los sistemas de osmosis inversa y, para mejorar los niveles de purificación, sobre todo en las presentaciones de agua embotellada, se utilizan generadores de ozono como germicidas. En México, las purificadoras legalmente instaladas en el territorio nacional, deben apegarse a normas establecidas por la Secretaría de Salud para el agua purificada de consumo humano, están son NOM-041 y NOM-201.

El presente trabajo ha sido generado por el interés de un grupo de ejidatarios de la comunidad de Azinyahualco, Municipio de Chilpancingo, en el Estado de Guerrero, México, a raíz de un financiamiento otorgado por el proyecto Comunidades, Indígenas y Biodiversidad (COINBIO), ver Coinbio (2008), para realizar un estudio de factibilidad para establecer una planta purificadora y envasadora de agua, con objeto de utilizar las fuentes de agua provenientes de las montañas de la región. Es conocido que personas pertenecientes a diversos sectores del mercado de agua purificada, consideran que el agua que proveen estas fuentes o manantiales es de calidad excepcional, toda vez que la dureza total del agua cruda es de apenas 20 partes por millón, además de que el sabor y la frescura que se perciben al tomarla son muy agradables al gusto. Si a esto agregamos que se le atribuyen propiedades curativas, aumentan las posibilidades de comercializarla y orientar su venta a consumidores de alto poder adquisitivo. Esto hace atractivo el invertir en una planta purificadora que se identifique con esta fuente.

Por lo anterior, desarrollamos un estudio que nos permita obtener la información que haga posible la valoración de diversos aspectos del mercado en que se planea insertar esta nueva marca de agua purificada. Este estudio debe establecer las regularidades del mercado para elaborar estrategias para dar a conocer la marca del producto y, establecer las potencialidades de los diversos nichos de mercado.

Es conocido que las empresas en desarrollo deben usar información sobre el mercado en que se quieren insertar, ver Jiménez y García (1999). El conocimiento obtenido permite hacer valoraciones inteligentes que den pie a elaborar estrategias. Esta información puede provenir de Bases de Datos existentes (Sistemas de Información) o de la investigación del mercado en que se desea insertar. A través de la investigación de mercado se debe establecer si hay espacio en el mercado antes de decidirse a producir el nuevo producto, es por esta razón que se efectuó una encuesta en el municipio Acapulco, que es uno de los centros urbanos más importantes del estado de Guerrero, dado que concentra, aproximadamente, el 23% de la población del estado de Guerrero, que según datos del censo 2005 del Instituto Nacional de Estadística Geografía e Informática (INEGI), ver INEGI (2008), asciende a 3 115 202 personas. Los detalles del diseño y ejecución de la misma se describen en la sección 2.

Está establecido que el marketing busca satisfacer los intereses de los clientes a través de elaborar políticas que soporten, por una parte, lo que podríamos denominar de tecnología del desarrollo del producto, y por otra, los intereses de la firma al posicionar el producto en el mercado. A partir de poseer la información requerida se van a fijar los costos y las propiedades del producto, así como la estrategia de precios a seguir y establecer la logística en términos de los lugares en que este se ofertará a la venta. Su inserción en el mercado va a ser más o menos exitosa en función de la promoción que desarrolle. De ahí la necesidad de estudios cuantitativos para insertar nuevos productos, vea Klapper (2005).

La prospección del mercado se llevó a cabo a partir de los resultados de la encuesta por muestreo desarrollada. Los análisis estadísticos de los datos obtenidos permiten desarrollar criterios sobre el número de botellones y botellas de agua requeridos para satisfacer la relación oferta-demanda del municipio de Acapulco, de acuerdo a las necesidades particulares de cada segmento de consumidores. Además se podrán fijar los precios y costes de situar el producto, en términos de la oferta, en los puntos de distribución. Los gerentes del futuro negocio, usarán los análisis estadísticos presentados en la sección 3, para establecer la logística requerida para satisfacer aseguramiento del transporte y almacenaje de los productos.

Es necesario conocer las fortalezas y debilidades del nuevo negocio en el sector del consumo de aguas purificadas, para crear estrategias que permitan aprovechar las debilidades de las actuales firmas para irrumpir en el mercado. Considerando los resultados de Zanutto y Bradlow (2006) consideramos necesario conocer la clientela en términos de sus necesidades, de los factores que caracterizan la calidad del agua purificada y los componentes de la fidelidad a las marcas. El interés de establecer las regularidades del posicionamiento de los consumidores es satisfecho a través de los resultados presentados y discutidos en la sección 3. Estas regularidades se enmarcan en el establecimiento del comportamiento de los clientes respecto a las marcas existentes en el mercado, la frecuencia de consumo, el lugar de compra de los distintos tipos de envase. Respecto a las marcas, es la lealtad de los consumidores la que es estudiada. En la sección 4 se estima la demanda potencial del Mercado de Acapulco para las demandas del agua purificada. Finalmente se dan las conclusiones del trabajo a la luz de estos resultados. Estos resultados utilizan métodos estadísticos clásicos populares en estudios de prospección de mercado basados en información cuantitativa. Véase, por ejemplo Dominick (2003), Jiménez Moreno y García Villaverde (1999) y Klapper (2005).

## **2. Población y diseño muestral.**

Al generar un nuevo producto, se debe determinar el mercado objetivo que se cubrirá. La población objetivo considerada en el presente estudio, es la de los núcleos familiares del municipio Acapulco. El municipio de Acapulco es el núcleo poblacional más importante del estado de Guerrero, concentrando 25.03% del total de viviendas particulares habitadas, en forma más específica, la Ciudad y Puerto de Acapulco concentra 23% del total de la población de la entidad, lo que corresponde a 152, 511 viviendas, según datos del conteo para el año de 2005, ver INEGI (2008).

Para recabar la necesaria información sobre los mercados locales del municipio Acapulco, del Estado de Guerrero de México, se desarrolló una encuesta. Entre sus objetivos estaba el de permitir suministrar, al futuro productor, los necesarios datos, sobre las características económicas de su industria así como de variables geográficas, demográficas, psicológicas y conductuales de los potenciales consumidores.

### **2.1 Diseño de la muestra.**

El diseño de muestreo se llevó a cabo utilizando los criterios usuales de teoría adaptados a la situación político económica de municipio Acapulco. Levy, Paul S. y Stanley Lemeshow (2008), Rueda García et al. (1992) y Rueda Arcos (1998) recomiendan el uso de muestreo polietápico para estudios de poblaciones humanas en problemas similares al que enfrentamos. Según los resultados de la encuesta realizada existen 90.19% viviendas que consumen agua de garrafón lo que representa 140,157 viviendas particulares habitadas para la cabecera municipal de Acapulco. El diseño muestral de la Encuesta de mercado para la botella de agua en las presentaciones de 1.5, 1 y 0.6 litros, es cuasi probabilística estratificada y polietápica. Es una muestra cuasi probabilística porque las unidades finales de selección se obtienen dentro de cada Zona mediante salto sistemático, de manera que se cubra la superficie urbana y suburbana de Acapulco. En las zonas de concentración de la población con alto nivel de ingresos, se entrevistaron a las personas en lugares de concentración como centros comerciales frecuentados por ellos. Es estratificada, porque las unidades de selección son agrupadas por características similares. Es además un diseño polietápico, porque la unidad última de selección (usuario del producto) es seleccionado después de varias etapas de aleatorización y a partir de un salto sistemático en la última etapa.

Para garantizar que las estimaciones obtenidas por la encuesta sean válidas y confiables, es necesario que el tamaño de muestra se defina para cada variable principal y para los correspondientes parámetros de interés. Por ello, el cálculo del tamaño de muestra está en función de los parámetros que se desean estimar, el nivel de confianza deseado y el error relativo máximo que se está dispuesto a aceptar. Además, es necesario incluir una corrección por no respuesta y, como el esquema de selección es diferente al aleatorio simple, se incorpora una cantidad de corrección debida al efecto de diseño. Una detallada discusión de estos problemas puede ser vista en Scheaffer, Lyman y Mendenhall (2006) y Arcos Cebrián y Rueda García (2004). Utilizamos como tamaño de muestra la solución para la ecuación:

$$n = \frac{Z_{\alpha/2}^2(1-P)deff}{r^2PTR}$$

Donde:

n = tamaño de muestra.

P = proporción a estimar

$Z_{\alpha/2}^2$  = Cuantil de una distribución normal asociado a un nivel de confianza deseado.

r = Error relativo máximo que se está dispuesto a aceptar.

deff = Efecto del diseño, que es la pérdida o ganancia en la eficiencia de diseño, por tratarse de uno complejo

TR = Tasa de respuesta

Para el presente trabajo consideramos un valor de  $P = 0.5$ ,  $Z_{\alpha/2}^2 = 1.96$ , que es el error relativo máximo permisible de  $\pm 2\%$ , un efecto de diseño del 1.4 y una tasa de no respuesta del 5%. Toda vez que se aplicará la encuesta a los individuos directamente, en su domicilio y en los centros de concentración, al momento de la aplicación de la encuesta se pedirá al entrevistado que complete sus respuestas o se hará uso de la reentrevista.

## 2.2 Tamaño y distribución de la muestra

El tamaño de muestra utilizado fue de 2496 individuos en Acapulco, estratificados por grupo de edad, sexo, personalidad, cultura, motivos de compra, nivel socioeconómico, ocupación y nivel de escolaridad, distribuidos de la siguiente forma:

Tabla 2.1 Estructura del tamaño de muestra utilizado por edad y sexo

|                   | Distribución de la Muestra en Acapulco por grupo de edad y Sexo |         |                 |              |         |         |                 |              |
|-------------------|---|---------|-----------------|--------------|---------|---------|-----------------|--------------|
|                   | Hombre  |         |                 |              | Mujer   |         |                 |              |
|                   | Jóvenes   | Adultos | Adultos Mayores | Tercera Edad | Jóvenes | Adultos | Adultos Mayores | Tercera Edad |
| Grupo de Edad     | 15-24   | 25-36   | 37-59           | 60 o +       | 15-24   | 25-36   | 37-59           | 60 o +       |
| Tamaño de muestra | 203   | 494     | 209             | 41           | 294     | 884     | 287             | 82           |
| TOTAL             | 947   |         |                 |              | 1547    |         |                 |              |

La distribución de la muestra por zonas, grupo de edad y sexo es dada en la Tabla 2.2.

Tabla 2.2 Estructura del tamaño de muestra utilizado por zona, edad y sexo

| Zona                           | Grupos de Edad |         |         |             |             | Total |
|--------------------------------|----------------|---------|---------|-------------|-------------|-------|
|                                | Género         | Jóvenes | Adultos | Ad. mayores | Tercera ed. |       |
| Anfiteatro: Costa Azul-Icacos  | Hombre         | 20      | 90      | 35          | 5           | 150   |
|                                | Mujer          | 30      | 120     | 45          | 5           | 200   |
|                                | Totales        | 50      | 210     | 80          | 10          | 350   |
| Anfiteatro Centro              | Hombre         | 36      | 108     | 33          | 6           | 183   |
|                                | Mujer          | 60      | 184     | 43          | 21          | 308   |
|                                | Totales        | 96      | 292     | 76          | 27          | 491   |
| Anfiteatro: Barrios Históricos | Hombre         | 26      | 67      | 35          | 7           | 135   |
|                                | Mujer          | 27      | 113     | 48          | 20          | 208   |
|                                | Totales        | 53      | 180     | 83          | 27          | 343   |
| Ciudad Renacimiento            | Hombre         | 33      | 61      | 33          | 9           | 136   |
|                                | Mujer          | 47      | 105     | 51          | 9           | 212   |
|                                | Totales        | 80      | 166     | 84          | 18          | 348   |
| Col. Emiliano Zapata           | Hombre         | 27      | 50      | 28          | 11          | 116   |
|                                | Mujer          | 38      | 115     | 44          | 22          | 219   |
|                                | Totales        | 65      | 165     | 72          | 33          | 335   |
| Col. Jardín                    | Hombre         | 16      | 33      | 17          | 2           | 68    |
|                                | Mujer          | 35      | 63      | 20          | 2           | 120   |
|                                | Totales        | 51      | 96      | 37          | 4           | 188   |
| Sabana-Máquina                 | Hombre         | 38      | 55      | 15          | 0           | 108   |
|                                | Mujer          | 43      | 78      | 13          | 1           | 135   |
|                                | Totales        | 81      | 133     | 28          | 1           | 243   |
| Coloso-Colosio                 | Hombre         | 7       | 30      | 13          | 1           | 51    |
|                                | Mujer          | 14      | 106     | 23          | 2           | 145   |
|                                | Totales        | 21      | 136     | 36          | 3           | 196   |

### 2.3 Esquema de selección.

Los pasos del proceso de selección fueron los siguientes:

- 1.- Asignación proporcional del número de estratos en la muestra, por grupo de edad, que para nuestro caso son 4 diferenciadas por sexo y por Zona,
- 2.- Selección de m individuos dentro de los 4 grupos (por sexo y por Zona) con igual probabilidad.

La colección de la información fue diferenciada, porque existen, clientes potenciales a los cuales el acceso se dificulta por la accesibilidad a las viviendas y por razones de seguridad en las zonas exclusivas o fraccionamientos privados. Por ello, en las zonas de fácil acceso la entrevista se realizó en el domicilio del cliente potencial, en entrevista cara a cara y se preguntó a una persona con capacidad de decisión de compra, por hogar. Para las zonas de difícil acceso, la información fue recabada en los sitios de compra mediante la entrevista de los clientes que acudían a los mismos.

Independientemente de la zona en la cual se entrevistó a los clientes potenciales, la selección se hizo usando salto sistemático, considerando grupo de edad y sexo.

Tabla 2.3 Esquema del proceso seguido de entrevistas

|                                |  |
|--------------------------------|--|
| <b>Tipo de entrevista</b>      | Cara a cara en hogares.  |
| <b>Selección de la muestra</b> | Hombres y Mujeres mayores de 15 años de edad entrevistadas en hogares, seleccionadas mediante salto sistemático en zonas de alta densidad poblacional, manzana, hogar e individuo. |
| <b>Público Objetivo</b>        | Hombres y mujeres mayores de 15 años residentes en la Ciudad y en la zona correspondiente.   |
| <b>Muestra</b>                 | 2494 entrevistas en hogares  |
| <b>Error Muestral</b>          | (+/-) 2.8 Puntos porcentuales, 95% de Confianza  |
| <b>Fecha de levantamiento</b>  | 2-3 de Junio de 2007.  |
| <b>Error Muestral</b>          | (+/-) 2.8 Puntos porcentuales, 95% de Confianza  |
| <b>Fecha de levantamiento</b>  | 2-3 de Junio de 2007.  |

#### 2.4 Diseño del cuestionario.

Toda vez que el estudio debe permitir:

- Establecer las regularidades del mercado para elaborar estrategias para dar a conocer la marca del producto.
- Identificar las potencialidades de los diversos nichos de mercado,
- Establecer valoraciones inteligentes que den pie a elaborar estrategias para su comercialización,
- Si hay espacio en el mercado antes de decidirse a producir el nuevo producto,
- Si es posible satisfacer los intereses de los clientes y los intereses de la firma,
- Obtener la información apropiada para fijar los costos y las propiedades del producto, así como la estrategia de precios a seguir y establecer la logística en términos de los lugares en que este se ofertará a la venta,
- Desarrollar criterios sobre el número de botellones y botellas de agua requeridos para satisfacer la relación oferta-demanda del municipio de Acapulco,
- Fijar los precios y costes de situar el producto, en términos de la oferta, en los puntos de distribución,
- Establecer la logística requerida para satisfacer aseguramiento del transporte y almacenaje de los productos,
- Conocer las fortalezas y debilidades del nuevo negocio en el sector del consumo de aguas purificadas, para crear estrategias que permitan aprovechar las debilidades de las actuales firmas para irrumpir en el mercado,
- Conocer la clientela en términos de sus necesidades, de los factores que caracterizan la calidad del agua purificada y los componentes de la fidelidad a las marcas.
- comportamiento de los clientes respecto a las marcas existentes en el mercado, la frecuencia de consumo, el lugar de compra de los distintos tipos de envase.



El cuestionario fue diseñado de acuerdo con los objetivos planteados anteriormente (Ver anexo 1). Este satisface los requerimientos metodológicos establecidos por diferentes autores como Jiménez Moreno. y García Villaverde (1999), Levy y Lemeshow (2008), Rueda García, Arcos Cebrián y González Carmona (1992), Rueda García y Arcos Cebrián (1998), Scheaffer , Lyman y Mendenhall (2006) y Zou, Feng y Qin (2002).

### 3. Análisis de la demanda

La compra de agua y refrescos se incrementa anualmente debido a las altas temperaturas que se reportan en los estados de la mayor parte del país. Este fenómeno se reporta en zonas de América Latina y principalmente en México, vea Mena Martínez (2008) y Ruiz Talavera (2008), Argentina, vea Laura Ferrarese (2006), y Chile Columna de Opinión (2008) por solo reportar algunos artículos periodísticos.

Se estudiaron los ratios y hábitos de compra de la población de 15 años y más en la Ciudad de Acapulco. Las viviendas en la ciudad de Acapulco suman la cantidad de 413,268, por lo que en la encuesta se reporta que las personas entrevistadas que han manifestado que consumen agua embotellada representan el 83.02%, por lo que se estima que 343,095 de esas personas consumen agua embotellada. Esto refleja el potencial de consumidores. Vea detalles en la Tabla 3.1.

Tabla 3.1 Potencial de consumo

| <b>Usualmente, ¿consume usted agua embotellada?</b> |         |            |                               |        |
|---|---------|------------|-------------------------------|--------|
|   | Consumo | Porcentaje | Intervalo de confianza al 95% |        |
|   |         |            | LI                            | LS     |
| Sí  | 2073    | 83,02%     | 81,50%                        | 84,50% |
| No  | 407     | 16,30%     |                               |        |
| No contestó   | 17      | 0,68%      |                               |        |
| Total   | 2497    | 100%       |                               |        |

De acuerdo a los resultados mostrados en esa tabla, se tiene que, la proporción de la población que consume agua embotellada en Acapulco se encuentra entre 81.50% y 84.50%, lo cual representa un número estimado de consumidores que va de 336802 a 349223, con una confiabilidad del 95%. Observe que hay solo un 0,68% de observaciones perdidas, lo que se considera no afecte significativamente la validez de las conclusiones de acuerdo al diseño inicial.

#### 3.1 Posicionamiento de las marcas en el mercado

La lealtad de un cliente a una marca es una de las problemáticas básicas en la práctica del marketing de relaciones. Su importancia reside en que un cliente leal comprará solo el agua de una marca. Muchos clientes no son absolutamente leales a una marca y esto puede ser usado para establecer un adecuado marketing de relaciones. Por ello es importante establecer el comportamiento de la fidelidad de los clientes de las marcas en el mercado. Los clientes que no muestran fidelidad por una marca particular son sensibles a las ofertas de la nueva marca, por ello

es necesario establecer a que valores (calidades) del agua son mas sensibles al fijar su preferencia. Haremos un estudio de los datos obtenidos con el fin de establecer niveles de lealtad a las marcas. Las técnicas usadas pueden ser consultadas en Webster (2000). El software utilizado fue el SPSS, dadas sus bondades para estudios de este tipo, Consúltese Ferran Aranaz, (1996) de ser necesario detalles sobre este paquete.

En el Estudio del Agua purificada en presentación de Garrafón de 20 litros el resultado de la encuesta determinó que la demanda de agua de garrafón se concentra fundamentalmente en el producto distribuido por la Pepsi Cola, la marca Electropura, con un 66.4% lo que significa que 101,267 viviendas particulares habitadas consumen este producto. Le siguen en orden de importancia la marca Diana con 9.2% y los establecimientos de agua para rellenar los garrafones llamados Florida Water con 4.0%. Las empresas embotelladoras de agua de garrafón pequeñas les corresponde el 20.4% de la demanda total, ver Tabla 3.2

Tabla 3.2 Estructura del consumo por marca

| Marca         | Viviendas | Porcentaje | Intervalo de confianza al 95% |        |
|---------------|-----------|------------|-------------------------------|--------|
|               |           |            | Li                            | Ls     |
| Electropura   | 1501      | 66,45%     | 66,40%                        | 66,49% |
| Diana         | 208       | 9,21%      | 9,18%                         | 9,23%  |
| Peyper        | 12        | 0,53%      | 0,52%                         | 0,54%  |
| Luvi          | 16        | 0,71%      | 0,70%                         | 0,72%  |
| Florida Water | 90        | 3,98%      | 3,97%                         | 4,00%  |
| La sierra     | 28        | 1,24%      | 1,23%                         | 1,25%  |
| Shanito       | 6         | 0,27%      | 0,26%                         | 0,27%  |
| No contestó   | 238       | 10,54%     | 10,51%                        | 10,56% |
| Otra          | 398       | 17,62%     | 17,58%                        | 17,65% |
| Total         | 2259      | 100,00%    |                               |        |

Estos resultados hacen evidente que hay una diferencia significativa entre electropura y las demás marcas, es decir, la proporción de hogares que consumen agua electropura es mayor que la de todas las demás juntas. Se confirma en la prueba que el verdadero valor de la proporción poblacional de viviendas en las cuales se consume agua electropura se encuentra entre 66.40% y 66.49%. Véase Tabla 3.3

Tabla 3.3 Prueba de hipótesis sobre la diferencia entre Electropura y las demás marcas para agua en garrafón

## Prueba binomial

|  |             | Categoría | N    | Proporción observada | Prop. de prueba | Sig. asintót. (unilateral) |
|--|-------------|-----------|------|----------------------|-----------------|----------------------------|
| ¿Cuál es la marca de agua de garrafón que se consume en esta vivienda? | Electropura | <= 1      | 1501 | ,66                  | ,66             | ,336 <sup>a</sup>          |
|  | Otras       | > 1       | 758  | ,34                  |                 |                            |
|  | Total       |           | 2259 | 1,00                 |                 |                            |

a. Basado en la aproximación Z.

En las presentaciones de agua embotellada, el 83.6% de las personas entrevistadas dijo que consumía agua embotellada, de las cuales el 44.0% de este conjunto señaló que consumía agua de la marca Manantial, el 34.5% consume agua de la marca Bonafón, el 15.0% agua de la marca e-Pura, la marca Diana participa con 1.6%, la Ciel con 1.2%, Aguafiel con 0.6% y otras marcas participan con 3.0%. Entonces, la competencia más fuerte la enfrentaría con las primeras tres marcas. Hubo un mayor porcentaje de observaciones perdidas, 16.7%, que representa la proporción de personas que no contestaron a esta parte del cuestionario porque manifestaron no consumir agua embotellada.

Analizando el porcentaje válido tenemos que los intervalos de confianza para las proporciones de viviendas en la cuales manifestaron consumir agua embotellada, son como se muestran en la tabla 3.4 y, es evidente que las diferencias son significativas para las tres marcas principales, toda vez que los intervalos de confianza no se traslapan. Por esta razón carece de sentido realizar alguna prueba para la diferencia de proporciones en estas tres marcas.

Tabla 3.4. Estudio del consumo de las marcas para garrafón

| Marca      | Viviendas | Porcentaje | % válido | Intervalo de confianza al 95% |        | E. estándar |
|------------|-----------|------------|----------|-------------------------------|--------|-------------|
|            |           |            |          | Li                            | Ls     |             |
| Manantial  | 916       | 36,68%     | 44,04%   | 42,11%                        | 45,97% | 0,96%       |
| Bonafón    | 718       | 28,75%     | 34,52%   | 32,71%                        | 36,33% | 0,91%       |
| e-pura     | 312       | 12,49%     | 15,00%   | 13,68%                        | 16,32% | 0,66%       |
| Diana      | 34        | 1,36%      | 1,63%    | 1,17%                         | 2,10%  | 0,23%       |
| Ciel       | 24        | 0,96%      | 1,15%    | 0,76%                         | 1,54%  | 0,20%       |
| Agua fiel  | 13        | 0,52%      | 0,63%    | 0,34%                         | 0,91%  | 0,14%       |
| Otra       | 63        | 2,52%      | 3,03%    | 2,40%                         | 3,66%  | 0,31%       |
| Total      | 2080      | 83,30%     | 100,00%  |                               |        |             |
| No consume | 417       | 16,70%     |          |                               |        |             |

Sin embargo, las marcas de agua Diana y Ciel presentan proporciones en las cuales parece que la diferencia no es significativa. Realicemos la prueba correspondiente:

$$H_0: p_1 = p_2$$

$$H_a: p_1 \neq p_2$$

Para esta prueba se obtiene un valor del estadístico  $Z_0 = 0.1382$  y el valor crítico es  $Z = 1.96$ . Por lo anterior se tiene que no existe evidencia al nivel  $\alpha = 0.05$  para rechazar la hipótesis de que la proporción de personas que dicen consumir agua de las marcas Diana y Ciel, es la misma.

### 3.2 Análisis de la frecuencia de compra

A partir de la experiencia acumulada en estudio de marketing, ver Berry y Linoff (2004) por ejemplo, conocer los procedimientos de compra ayuda a influir en la opción de compra de la nueva marca de agua purificada. Los resultados de la encuesta fijan que el 29.9% de las viviendas consumen 2 garrafones de agua por semana. Este porcentaje equivale a 41,907 viviendas particulares habitadas sobre el total de viviendas que han manifestado consumir agua de garrafón. Casi el 70% de las viviendas consumen entre 1 y 3 garrafones de agua. El promedio del número de garrafones consumidos es de 3.11 y su desviación típica es de 1.842. Aceptando la normalidad podemos decir con un 95% de confiabilidad que el consumo promedio de garrafones está entre 3.03 y 3.18. Para el agua embotellada el consumo promedio por semana es de 4.7 unidades y su desviación típica es de 3.30. Asumiendo normalidad de los datos podemos afirmar que el promedio de botellas consumidas está entre 4.55 y 4.84, con una confianza del 95%. Ver detalles en la tabla 3.6.

Tabla 3.6 Estudio del consumo de garrafón y botellas

| Variable                                | Mínimo | Máximo | Medi<br>a | Error típico | Desv.<br>típ. | Intervalo de confianza al 95% |      |
|---|--------|--------|-----------|--------------|---------------|-------------------------------|------|
|   |        |        |           |              |               | Li                            | Ls   |
| Consumo semanal de Garrafones de 20 lts | 1      | 20     | 3,11      | 0,04         | 1,84          | 3,03                          | 3,18 |
| Consumo semanal de botellas de agua     | 1      | 30     | 4,70      | 0,07         | 3,30          | 4,55                          | 4,84 |

Ahora, si consideramos la proporción de personas que consumen desde uno a 20 garrafones, la población se distribuiría como se muestra en la tabla 3.7.

Tabla 3.7 Estudio del número de garrafones consumidos

| Garrafones  | Viviendas | Porcentaje | % válido | Intervalo de confianza al 95% |        | E. Estándar |
|-------------|-----------|------------|----------|-------------------------------|--------|-------------|
|             |           |            |          | Li                            | Ls     |             |
| 1G          | 311       | 12,45      | 13,79    | 13,510                        | 14,061 | 0,13785461  |
| 2G          | 674       | 26,99      | 29,88    | 29,278                        | 30,473 | 0,29875887  |
| 3G          | 579       | 23,19      | 25,66    | 25,152                        | 26,178 | 0,25664894  |
| 4G          | 320       | 12,82      | 14,18    | 13,901                        | 14,468 | 0,14184397  |
| 5G          | 164       | 6,57       | 7,27     | 7,124                         | 7,415  | 0,07269504  |
| 6G          | 106       | 4,25       | 4,70     | 4,605                         | 4,793  | 0,04698582  |
| 7G          | 54        | 2,16       | 2,39     | 2,346                         | 2,441  | 0,02393617  |
| 8G          | 17        | 0,68       | 0,75     | 0,738                         | 0,769  | 0,00753546  |
| 9G          | 5         | 0,20       | 0,22     | 0,217                         | 0,226  | 0,00221631  |
| 10G         | 14        | 0,56       | 0,62     | 0,608                         | 0,633  | 0,00620567  |
| 12G         | 3         | 0,12       | 0,13     | 0,130                         | 0,136  | 0,00132979  |
| 14G         | 4         | 0,16       | 0,18     | 0,174                         | 0,181  | 0,00177305  |
| 15G         | 3         | 0,12       | 0,13     | 0,130                         | 0,136  | 0,00132979  |
| 18G         | 1         | 0,04       | 0,04     | 0,043                         | 0,045  | 0,00044326  |
| 20G         | 1         | 0,04       | 0,04     | 0,043                         | 0,045  | 0,00044326  |
| Total       | 2256      | 90,35      | 100      |                               |        |             |
| No consumen | 241       | 9,65       |          |                               |        |             |
| Total       | 2497      | 100        |          |                               |        |             |

Es posible que la proporción de hogares en los que se consume 1G de 20 litros a la semana es igual a la proporción de viviendas en la cuales se consumen 4G de 20 litros semanales. Algo similar ocurre con las proporciones de viviendas en las cuales se consumen 2 y 3 garrafones. Haremos las pruebas de comparación correspondiente:

$$H_0: p_1 = p_4$$

$$H_a: p_1 \neq p_4$$

Para esta prueba se tiene un valor de  $Z_0 = -0.144$  y  $Z = 1.96$ , por lo que no existe evidencia al nivel  $\alpha = 0.05$  para rechazar la hipótesis que las proporciones de viviendas en las cuales se consume 1 y 4 garrafones, es la misma.

$$H_0: p_2 = p_3$$

$$H_a: p_2 \neq p_3$$

En esta prueba  $Z_0 = 1.66 <$  que  $Z = 1.96$ , por lo que se concluye que no existe evidencia al nivel de significación  $\alpha = 0.05$  para rechazar la hipótesis que las proporciones de viviendas en las cuales se consume 2 y 3 garrafones, es la misma.

A partir de las respuestas de los muestreados se estima la estructura de consumo de los consumidores de agua embotellada. En la muestra se observa que el 31.8% compra la presentación de un litro, el 31.6% la presentación de 600 ml., el 13.8% la presentación de 1.5 litros, el 13.7% la de dos litros y solamente el 9.1% la de medio litro, ver tabla 3.8.

Tabla 3.8 Estudio del consumo de los tamaños de botella ofertados  
¿Cuál es el tamaño de botella de agua que consume con mayor frecuencia?

| Tamaño        | Unidades | Porcentaje | % válido | Intervalo de confianza al 95% |        | Error estándar |
|---------------|----------|------------|----------|-------------------------------|--------|----------------|
|               |          |            |          | Li                            | Ls     |                |
| Medio litro   | 190      | 7,609      | 9,139    | 7,999                         | 10,279 | 0,582          |
| 600 ml        | 657      | 26,312     | 31,602   | 29,709                        | 33,495 | 0,966          |
| Un litro      | 661      | 26,472     | 31,794   | 29,898                        | 33,691 | 0,968          |
| Litro y medio | 287      | 11,494     | 13,805   | 12,434                        | 15,176 | 0,700          |
| 2 litros      | 284      | 11,374     | 13,660   | 12,296                        | 15,025 | 0,696          |
| Total         | 2079     | 83,260     | 100      |                               |        |                |
| No contestó   | 418      | 16,740     |          |                               |        |                |
| Total         | 2497     | 100        |          |                               |        |                |

Observe que posiblemente no existen diferencias significativas entre la proporción de personas que prefieren la presentación de 600 ml y un litro, así como entre las presentaciones de 1,5 y 2 litros. Haremos las pruebas correspondientes:

$$H_0: p_{(600\text{ ml})} = p_{(\text{un l})} \quad \text{o} \quad H_0: p_{(600\text{ ml})} - p_{(\text{un l})} = 0$$

$$H_0: p_{(600\text{ ml})} \neq p_{(\text{un l})} \quad \text{o} \quad H_0: p_{(600\text{ ml})} - p_{(\text{un l})} \neq 0$$

Si consideramos un nivel de significación  $\alpha = 0.05$  el valor crítico  $Z = - 1.96$  que es menor que el valor computado que fue  $Z_0 = - 0.075058$ , por lo que no existe evidencia para rechazar la hipótesis de que las proporciones son iguales. Es decir, la proporción de personas que prefiere las presentaciones de 600 ml y de un litro es la misma.

Para las presentaciones de 1,5 y 2 litros se tiene:

$$H_0: p_{(1,5)} = p_{(2)} \quad \text{o} \quad H_0: p_{(1,5)} - p_{(2)} = 0$$

$$H_0: p_{(1,5)} \neq p_{(2)} \quad \text{o} \quad H_0: p_{(1,5)} - p_{(2)} \neq 0$$

Si consideramos un nivel de significación  $\alpha = 0.05$  el valor crítico  $Z = 1.96$  supera a  $Z_0 = 0.05009$ , por lo que no existe evidencia para rechazar la hipótesis de que las proporciones son iguales o que la diferencia entre ellas es cero. Es decir, la proporción de personas que prefiere las presentaciones de 1,5 litros y de 2 litros es la misma.

### 3.3 Análisis del lugar de compra del agua purificada

Una variable importante en el marketing es la distribución, vea Berry y Gordon (2004). O sea sobre cómo hacer llegar el producto al cliente. Esto incluye desde los mismos vendedores hasta la manera de transportar el producto al lugar de consumo. En la distribución del agua purificada se distingue.

Considerando el total de viviendas donde se consume agua de garrafón, se tiene que el 39.7 % de las viviendas sus habitantes han manifestado que compran el agua de garrafón en las misceláneas, el 34.9% lo adquieren en los camiones y el resto (25.4%) compran el agua de garrafón en otros establecimientos. Ver tabla 3.9.

Tabla 3.9. Estudio del lugar de venta de los garrafones  
Usualmente, ¿Dónde compra el agua de garrafón?

| ¿Dónde compra el agua de garrafón? | Menciones | Porcentaje | % válido | Intervalo de confianza al 95% |        | E. Estándar |
|------------------------------------|-----------|------------|----------|-------------------------------|--------|-------------|
|                                    |           |            |          | Li                            | Ls     |             |
| Miscelánea                         | 895       | 35,843     | 39,707   | 37,687                        | 41,356 | 1,0306      |
| Tienda de Abarrotes                | 267       | 10,693     | 11,846   | 10,512                        | 12,935 | 0,6806      |
| Supermercado                       | 81        | 3,244      | 3,594    | 2,825                         | 4,221  | 0,3921      |
| Camión                             | 786       | 31,478     | 34,871   | 32,904                        | 36,477 | 1,0038      |
| Otros                              | 225       | 9,011      | 9,982    | 8,745                         | 10,992 | 0,6314      |
| Total                              | 2254      | 90,268     | 100      |                               |        |             |
| No contestó                        | 243       | 9,732      |          |                               |        |             |
| Total                              | 2497      | 100        |          |                               |        |             |

Al momento de decidir sobre los lugares de venta del nuevo producto deberemos tener en cuenta estos resultados, es decir, deben diseñarse políticas de ventas en misceláneas y distribución directa en el domicilio del consumidor empleando camiones de reparto.

Para la compra de botellas de agua se obtiene los resultados de la tabla 3.10:

Tabla 3.10. Estudio del lugar de venta de las botellas

| ¿Dónde compra el | Menciones | Porcentaje | % válido | Intervalo de confianza al 95% | E. |
|------------------|-----------|------------|----------|-------------------------------|----|
|------------------|-----------|------------|----------|-------------------------------|----|

| agua embotellada?   |      | e      |         | Li     | Ls     | Estándar |
|---------------------|------|--------|---------|--------|--------|----------|
| Miscelánea          | 1463 | 58,590 | 70,4381 | 68,476 | 72,040 | 1,0013   |
| Tienda de Abarrotes | 370  | 14,818 | 17,8142 | 16,169 | 19,157 | 0,8396   |
| Supermercado        | 158  | 6,328  | 7,6071  | 6,467  | 8,538  | 0,5817   |
| Camión              | 44   | 1,762  | 2,1184  | 1,499  | 2,624  | 0,3160   |
| Otros               | 42   | 1,682  | 2,0221  | 1,417  | 2,516  | 0,3089   |
| Total               | 2077 | 83,180 | 100     |        |        |          |
| No contestó         | 420  | 16,820 |         |        |        |          |
| Total               | 2497 | 100    |         |        |        |          |

Observemos que las mayores ventas ocurren en las misceláneas y en segundo lugar, las tiendas de abarrotes.

### 3.4 Lealtad a los productos en el mercado.

Los consumidores tienen preferencias y se inclinan por ciertas marcas. Esto es destacado por diversos autores como Jiménez Moreno y García Villaverde (1999), Klapper (2005), Webster (2000). El resultado de nuestro estudio mostró como se comporta la lealtad el mercado del agua purificada. Es conocido que la lealtad hacia la marca no es aleatoria y que se refleja en la conducta ante la compra. En el agua purificada se espera poca firmeza en la lealtad dadas las características del producto ofertado. La lealtad de marca se puede explicar a través de las características del comprador y su sector social. Para el caso de agua de garrafón se tiene:

Tabla 3.11 Homogeneidad de la razón de posible cambio de marca por tipo de ingreso para garrafón.

Tabla de contingencia

|                   | ¿Cual sería la principal razón para cambiar de agua de garrafón? |                       |                     |                        |                        |                       | Total                |                        |
|-------------------|--|-----------------------|---------------------|------------------------|------------------------|-----------------------|----------------------|------------------------|
|                   | Por su sabor   | Por su comodidad      | Por su presentación | Por su calidad         | Por su precio          | Otra razón            |                      |                        |
| Nivel de ingresos | Bajos ingresos   | 203<br>25,9%<br>34,2% | 28<br>3,6%<br>35,0% | 4<br>,5%<br>23,5%      | 299<br>38,1%<br>34,0%  | 166<br>21,1%<br>37,4% | 85<br>10,8%<br>42,3% | 785<br>100,0%<br>35,5% |
|                   | Medianos Ingresos  | 259<br>27,7%<br>43,7% | 35<br>3,7%<br>43,8% | 8<br>,9%<br>47,1%      | 375<br>40,1%<br>42,7%  | 170<br>18,2%<br>38,3% | 88<br>9,4%<br>43,8%  | 935<br>100,0%<br>42,2% |
|                   | Altos Ingresos   | 131<br>26,5%<br>22,1% | 17<br>3,4%<br>21,3% | 5<br>1,0%<br>29,4%     | 205<br>41,5%<br>23,3%  | 108<br>21,9%<br>24,3% | 28<br>5,7%<br>13,9%  | 494<br>100,0%<br>22,3% |
| Total             | 593<br>26,8%<br>100,0%   | 80<br>3,6%<br>100,0%  | 17<br>,8%<br>100,0% | 879<br>39,7%<br>100,0% | 444<br>20,1%<br>100,0% | 201<br>9,1%<br>100,0% | 2214<br>100,0%       |                        |

Los consumidores potenciales clasificados como de bajos ingresos cambiarían el agua de garrafón que actualmente consumen, principalmente por la calidad y el sabor de un nuevo producto, la tercera variable en importancia es el precio. Los consumidores potenciales clasificados como de medianos ingresos consideran que cambiarían el agua que actualmente consumen si el nuevo producto es mejor en calidad y sabor. En el grupo de altos ingresos se repite este comportamiento, pero, ¿la intención de cambio es independiente del nivel de ingresos? Para contestar a esta

pregunta realizamos una prueba de independencia-homogeneidad entre las variables mediante la utilización del estadístico  $\chi^2$ . Vea Webster (2000) para una discusión sobre esta prueba.

La hipótesis se puede formular de la siguiente forma:

H<sub>0</sub>: La intención de cambio es independiente del nivel de ingresos de la población.

H<sub>a</sub>: La intención de cambio depende del nivel de ingresos de la población.

Como se puede observar en la tabla 3.12, no existe evidencia al nivel de significación  $\alpha = 0.05$  para rechazar la hipótesis de que “La intención de cambio es independiente del nivel de ingresos de la población”, toda vez que  $0.141 > 0.05$ . Por lo anterior podemos concluir que independientemente del nivel de ingresos, la población estará dispuesta a cambiar el producto que ahora consume si es que existe una mejora en el sabor, calidad y precio.

Tabla 3.12 Prueba para la Homogeneidad de la razón de posible cambio de marca por tipo de ingreso para garrafón

**Pruebas de chi-cuadrado**

|                              | Valor               | gl | Sig. asintótica<br>(bilateral) |
|------------------------------|---------------------|----|--------------------------------|
| Chi-cuadrado de Pearson      | 14,757 <sup>a</sup> | 10 | ,141                           |
| Razón de verosimilitud       | 15,663              | 10 | ,110                           |
| Asociación lineal por lineal | 1,834               | 1  | ,176                           |
| N de casos válidos           | 2214                |    |                                |

a. 1 casillas (5,6%) tienen una frecuencia esperada inferior a 5.  
La frecuencia mínima esperada es 3,79.

Para el agua de botella en sus distintas presentaciones se observa que la intención de cambio de la marca que actualmente se consume se centra en la calidad, el sabor y el precio del nuevo producto como se puede comprobar en la tabla 3.13, pero, a diferencia del caso de agua de garrafón, al hacer la prueba de independencia entre las variables nivel de ingresos e intención de cambio del agua que actualmente consume, existe evidencia suficiente al nivel de significación  $\alpha = 0.05$  para rechazar la hipótesis de que “La intención de cambio es independiente del nivel de ingresos de la población”, toda vez que el p valor obtenido de la muestra es menor que el valor de prueba, es decir que como el P-valor es 0.007 es menor que el nivel  $\alpha = 0.05$ , como se muestra en la tabla 3.14. Podemos concluir que la intención de cambio del agua embotellada que actualmente se consume, depende del nivel de ingresos. Este aspecto debe tenerse en cuenta al momento de diseñar la estrategia de publicidad del nuevo producto.

Tabla 3.13 Homogeneidad de la razón de posible cambio de marca por tipo de ingreso para botella



Tabla de contingencia

|                   |                   | ¿Cual sería la principal razón para cambiar de marca de agua embotellada? |                      |                      |                        |                        | Otra razón             | Total                    |
|-------------------|-------------------|---|----------------------|----------------------|------------------------|------------------------|------------------------|--------------------------|
|                   |                   | Por su sabor  | Por su comodidad     | Por su presentación  | Por su calidad         | Por su precio          |                        |                          |
| Nivel de ingresos | Bajos ingresos    | 196<br>28,6%<br>34,8%   | 15<br>2,2%<br>21,4%  | 6<br>,9%<br>16,7%    | 228<br>33,3%<br>31,8%  | 135<br>19,7%<br>35,4%  | 105<br>15,3%<br>42,3%  | 685<br>100,0%<br>34,0%   |
|                   | Medianos Ingresos | 235<br>27,7%<br>41,7%   | 34<br>4,0%<br>48,6%  | 23<br>2,7%<br>63,9%  | 306<br>36,1%<br>42,7%  | 152<br>17,9%<br>39,9%  | 98<br>11,6%<br>39,5%   | 848<br>100,0%<br>42,1%   |
|                   | Altos Ingresos    | 133<br>27,6%<br>23,6%   | 21<br>4,4%<br>30,0%  | 7<br>1,5%<br>19,4%   | 182<br>37,8%<br>25,4%  | 94<br>19,5%<br>24,7%   | 45<br>9,3%<br>18,1%    | 482<br>100,0%<br>23,9%   |
| Total             |                   | 564<br>28,0%<br>100,0%  | 70<br>3,5%<br>100,0% | 36<br>1,8%<br>100,0% | 716<br>35,5%<br>100,0% | 381<br>18,9%<br>100,0% | 248<br>12,3%<br>100,0% | 2015<br>100,0%<br>100,0% |

Tabla 3.14 Prueba para la Homogeneidad de la razón de posible cambio de marca por tipo de ingreso para botella

Pruebas de chi-cuadrado

|                              | Valor               | gl | Sig. asintótica (bilateral) |
|------------------------------|---------------------|----|-----------------------------|
| Chi-cuadrado de Pearson      | 24,128 <sup>a</sup> | 10 | ,007                        |
| Razón de verosimilitud       | 24,676              | 10 | ,006                        |
| Asociación lineal por lineal | 2,014               | 1  | ,156                        |
| N de casos válidos           | 2015                |    |                             |

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5.  
La frecuencia mínima esperada es 8,61.

Otro de los aspectos que deben tenerse en cuenta antes de comercializar el producto, aunque el precio debe ser el mismo, son las zonas de distribución, dado que si conocemos la intención de cambio por distribución geográfica, estaremos en condiciones de diseñar estrategias de comercialización específicas para cada zona. Se confirma que la población en estudio cambiaría a un nuevo producto si es mejor en calidad, sabor y precio, aunque es en las zonas donde se concentran las personas de medianos y altos ingresos (zona del anfiteatro), que dan menos importancia al precio y una mayor importancia a la calidad y al sabor, en ese orden.

Tabla 3.15 Homogeneidad de la razón de posible cambio de marca por zona para garrafón

Tabla de contingencia

| Zona                           | ¿Cual sería la principal razón para cambiar de agua de garrafón? |                  |                     |                |               |            | Total  |
|--------------------------------|--|------------------|---------------------|----------------|---------------|------------|--------|
|                                | Por su sabor   | Por su comodidad | Por su presentación | Por su calidad | Por su precio | Otra razón |        |
| Anfiteatro: Costa Azul-Icacos  | 42   | 19               | 4                   | 133            | 49            | 26         | 273    |
|                                | 15,4%  | 7,0%             | 1,5%                | 48,7%          | 17,9%         | 9,5%       | 100,0% |
| Anfiteatro Centro              | 7,1%   | 23,8%            | 23,5%               | 15,1%          | 11,0%         | 12,9%      | 12,3%  |
|                                | 166  | 20               | 2                   | 168            | 66            | 25         | 447    |
| Anfiteatro: Barrios Históricos | 37,1%  | 4,5%             | ,4%                 | 37,6%          | 14,8%         | 5,6%       | 100,0% |
|                                | 27,9%  | 25,0%            | 11,8%               | 19,0%          | 14,9%         | 12,4%      | 20,2%  |
| Ciudad Renacimiento            | 103  | 14               | 2                   | 97             | 64            | 36         | 316    |
|                                | 32,6%  | 4,4%             | ,6%                 | 30,7%          | 20,3%         | 11,4%      | 100,0% |
| Col. Emiliano Zapata           | 17,3%  | 17,5%            | 11,8%               | 11,0%          | 14,4%         | 17,9%      | 14,2%  |
|                                | 65   | 10               | 4                   | 118            | 68            | 32         | 297    |
| Col. Jardín                    | 21,9%  | 3,4%             | 1,3%                | 39,7%          | 22,9%         | 10,8%      | 100,0% |
|                                | 10,9%  | 12,5%            | 23,5%               | 13,4%          | 15,3%         | 15,9%      | 13,4%  |
| Sabana-Máquina                 | 53   | 7                | 2                   | 120            | 62            | 41         | 285    |
|                                | 18,6%  | 2,5%             | ,7%                 | 42,1%          | 21,8%         | 14,4%      | 100,0% |
| Coloso-Colosio                 | 8,9%   | 8,8%             | 11,8%               | 13,6%          | 14,0%         | 20,4%      | 12,8%  |
|                                | 37   | 3                | 1                   | 73             | 34            | 31         | 179    |
| Anfiteatro: Barrios Históricos | 20,7%  | 1,7%             | ,6%                 | 40,8%          | 19,0%         | 17,3%      | 100,0% |
|                                | 6,2%   | 3,8%             | 5,9%                | 8,3%           | 7,7%          | 15,4%      | 8,1%   |
| Anfiteatro: Barrios Históricos | 71   | 1                | 1                   | 74             | 81            | 5          | 233    |
|                                | 30,5%  | ,4%              | ,4%                 | 31,8%          | 34,8%         | 2,1%       | 100,0% |
| Anfiteatro: Barrios Históricos | 12,0%  | 1,3%             | 5,9%                | 8,4%           | 18,2%         | 2,5%       | 10,5%  |
|                                | 57   | 6                | 1                   | 99             | 20            | 5          | 188    |
| Anfiteatro: Barrios Históricos | 30,3%  | 3,2%             | ,5%                 | 52,7%          | 10,6%         | 2,7%       | 100,0% |
|                                | 9,6%   | 7,5%             | 5,9%                | 11,2%          | 4,5%          | 2,5%       | 8,5%   |
| Total                          | 594  | 80               | 17                  | 882            | 444           | 201        | 2218   |
|                                | 26,8%  | 3,6%             | ,8%                 | 39,8%          | 20,0%         | 9,1%       | 100,0% |
|                                | 100,0%   | 100,0%           | 100,0%              | 100,0%         | 100,0%        | 100,0%     | 100,0% |

La hipótesis de independencia para la intención de cambio y zonas s formula de la siguiente forma:

H<sub>0</sub>: La intención de cambio es independiente de la zona donde vive el cliente potencial.

H<sub>a</sub>: La intención de cambio depende de la zona donde vive el cliente potencial.

Como se puede observar en la tabla 3.16, existe evidencia al nivel de significación  $\alpha = 0.05$  para rechazar la hipótesis de que “La intención de cambio es independiente de la zona donde vive el cliente potencial”, toda vez que el P-valor es cero y es menor que 0.05. Por lo anterior podemos concluir que la intención de cambio depende de la zona donde vive el cliente potencial.

Tabla 3.16 Prueba para la Homogeneidad de la razón de posible cambio de marca por zona para garrafón

## Pruebas de chi-cuadrado

|                              | Valor                | gl | Sig. asintótica<br>(bilateral) |
|------------------------------|----------------------|----|--------------------------------|
| Chi-cuadrado de Pearson      | 192,115 <sup>a</sup> | 35 | ,000                           |
| Razón de verosimilitud       | 198,557              | 35 | ,000                           |
| Asociación lineal por lineal | 1,559                | 1  | ,212                           |
| N de casos válidos           | 2218                 |    |                                |

a. 8 casillas (16,7%) tienen una frecuencia esperada inferior a 5.  
La frecuencia mínima esperada es 1,37.

Para el caso de agua embotellada se observa un comportamiento similar al observado en el agua de garrafón, es decir, cambiarán solamente si el nuevo producto es mejor en calidad, sabor y precio. Ver tabla 3.17.

Tabla 3.17 Homogeneidad de la razón de posible cambio de marca por zona para botella

|       |                                | ¿Cual sería la principal razón para cambiar de marca de agua embotellada? |                  |                     |                |               |            | Total  |
|-------|--------------------------------|---|------------------|---------------------|----------------|---------------|------------|--------|
|       |                                | Por su sabor  | Por su comodidad | Por su presentación | Por su calidad | Por su precio | Otra razón |        |
| Zona  | Anfiteatro: Costa Azul-Icaicos | 62  | 15               | 9                   | 105            | 48            | 32         | 271    |
|       |                                | 22,9%   | 5,5%             | 3,3%                | 38,7%          | 17,7%         | 11,8%      | 100,0% |
|       |                                | 11,0%   | 21,4%            | 25,0%               | 14,6%          | 12,6%         | 12,9%      | 13,4%  |
|       | Anfiteatro Centro              | 160   | 14               | 4                   | 152            | 67            | 29         | 426    |
|       |                                | 37,6%   | 3,3%             | ,9%                 | 35,7%          | 15,7%         | 6,8%       | 100,0% |
|       |                                | 28,3%   | 20,0%            | 11,1%               | 21,2%          | 17,6%         | 11,6%      | 21,1%  |
|       | Anfiteatro: Barrios Históricos | 86  | 13               | 5                   | 80             | 39            | 48         | 271    |
|       |                                | 31,7%   | 4,8%             | 1,8%                | 29,5%          | 14,4%         | 17,7%      | 100,0% |
|       |                                | 15,2%   | 18,6%            | 13,9%               | 11,1%          | 10,2%         | 19,3%      | 13,4%  |
|       | Ciudad Renacimiento            | 67  | 8                | 4                   | 82             | 40            | 52         | 253    |
|       |                                | 26,5%   | 3,2%             | 1,6%                | 32,4%          | 15,8%         | 20,6%      | 100,0% |
|       |                                | 11,9%   | 11,4%            | 11,1%               | 11,4%          | 10,5%         | 20,9%      | 12,5%  |
|       | Col. Emiliano Zapata           | 65  | 1                | 3                   | 60             | 49            | 49         | 227    |
|       |                                | 28,6%   | ,4%              | 1,3%                | 26,4%          | 21,6%         | 21,6%      | 100,0% |
|       |                                | 11,5%   | 1,4%             | 8,3%                | 8,4%           | 12,9%         | 19,7%      | 11,2%  |
|       | Col. Jardín                    | 21  | 3                | 4                   | 57             | 41            | 30         | 156    |
|       |                                | 13,5%   | 1,9%             | 2,6%                | 36,5%          | 26,3%         | 19,2%      | 100,0% |
|       |                                | 3,7%  | 4,3%             | 11,1%               | 7,9%           | 10,8%         | 12,0%      | 7,7%   |
|       | Sabana-Máquina                 | 51  | 10               | 6                   | 79             | 83            | 6          | 235    |
|       |                                | 21,7%   | 4,3%             | 2,6%                | 33,6%          | 35,3%         | 2,6%       | 100,0% |
|       |                                | 9,0%  | 14,3%            | 16,7%               | 11,0%          | 21,8%         | 2,4%       | 11,6%  |
|       | Coloso-Colosio                 | 53  | 6                | 1                   | 103            | 14            | 3          | 180    |
|       |                                | 29,4%   | 3,3%             | ,6%                 | 57,2%          | 7,8%          | 1,7%       | 100,0% |
|       |                                | 9,4%  | 8,6%             | 2,8%                | 14,3%          | 3,7%          | 1,2%       | 8,9%   |
| Total |                                | 565   | 70               | 36                  | 718            | 381           | 249        | 2019   |
|       |                                | 28,0%   | 3,5%             | 1,8%                | 35,6%          | 18,9%         | 12,3%      | 100,0% |
|       |                                | 100,0%  | 100,0%           | 100,0%              | 100,0%         | 100,0%        | 100,0%     | 100,0% |

La prueba de independencia para la intención de cambio en el consumo de agua de botella y zonas se muestra en la tabla 3.18 y en ella se muestra que “existe evidencia al nivel de significación  $\alpha = 0.05$  para rechazar la hipótesis de que “La intención de cambio es independiente de la zona donde vive el cliente potencial”, toda vez que el P-valor nuevamente es cero que es menor que 0.05. Por lo anterior podemos concluir que la intención de cambio en el consumo de agua de botella en las distintas presentaciones depende de la zona donde vive el cliente potencial.

Tabla 3.18 Prueba de Homogeneidad de la razón de posible cambio de marca por zona para botella

## Pruebas de chi-cuadrado

|                              | Valor                | gl | Sig. asintótica<br>(bilateral) |
|------------------------------|----------------------|----|--------------------------------|
| Chi-cuadrado de Pearson      | 232,564 <sup>a</sup> | 35 | ,000                           |
| Razón de verosimilitud       | 244,725              | 35 | ,000                           |
| Asociación lineal por lineal | 5,615                | 1  | ,018                           |
| N de casos válidos           | 2019                 |    |                                |

a. 7 casillas (14,6%) tienen una frecuencia esperada inferior a 5.  
La frecuencia mínima esperada es 2,78.

En resumen, se tiene que en el 39.8% de las viviendas particulares, sus habitantes han manifestado que pudieran cambiar de marca del agua de garrafón por su calidad, lo que representa 55,782 viviendas del total de viviendas cuyos habitantes han manifestado consumir agua de garrafón. Otra razón por lo que los consumidores estarían dispuestos a cambiar de marca es por el sabor del agua con el 26.8%.

Tabla 3.19 Estructura de la razón de cambio de marca para garrafón

| ¿Por qué cambiar?   | Frecuencia | %      | % válido | Intervalo de confianza al 95% |        | E. Estándar |
|---------------------|------------|--------|----------|-------------------------------|--------|-------------|
|                     |            |        |          | Li                            | Ls     |             |
| Por su sabor        | 594        | 23,789 | 26,781   | 23,220                        | 30,342 | 1,817       |
| Por su comodidad    | 80         | 3,204  | 3,607    | -0,479                        | 7,693  | 2,085       |
| Por su presentación | 17         | 0,681  | 0,766    | -3,379                        | 4,912  | 2,115       |
| Por su calidad      | 882        | 35,322 | 39,766   | 36,536                        | 42,996 | 1,648       |
| Por su precio       | 444        | 17,781 | 20,018   | 16,296                        | 23,740 | 1,899       |
| Otra razón          | 201        | 8,050  | 9,062    | 5,094                         | 13,031 | 2,025       |
| No contesta         | 279        | 11,173 | 100      |                               |        |             |

En la Tabla 3.20 se tiene que el 35.6% de los consumidores, que respondieron, estarían dispuestos a cambiar de marca de agua embotellada por una mejor calidad, el 28% cambiaría si se mejora el sabor, el 18.9% si el precio al consumidor se mejora, la razón por comodidad con 3.5%, por su presentación 1.8% y otra razón con 12.3%. Observe que el 82.42% de la población cambiaría a una nueva agua purificada si se mejora la calidad el sabor y el precio. Esto representa 340615 clientes potenciales de los 413268 posibles.

Tabal 3.20 Estructura de la razón de cambio de marca para botella

| ¿Por qué cambiar?   | Frecuencia | Porcentaje | Porcentaje válido | Intervalo de confianza al 95% |        | E. Estándar |
|---------------------|------------|------------|-------------------|-------------------------------|--------|-------------|
|                     |            |            |                   | Li                            | Ls     |             |
| Por su sabor        | 565        | 22,627     | 27,984            | 24,282                        | 31,686 | 1,889       |
| Por su comodidad    | 70         | 2,803      | 3,467             | -0,819                        | 7,753  | 2,187       |
| Por su presentación | 36         | 1,442      | 1,783             | -2,540                        | 6,106  | 2,206       |

|                |     |        |        |        |        |       |
|----------------|-----|--------|--------|--------|--------|-------|
| Por su calidad | 718 | 28,755 | 35,562 | 32,061 | 39,064 | 1,786 |
| Por su precio  | 381 | 15,258 | 18,871 | 14,942 | 22,800 | 2,005 |
| Otra razón     | 249 | 9,972  | 12,333 | 8,249  | 16,417 | 2,084 |
| No contestó    | 478 | 19,143 | 100    |        |        |       |

#### 4. Estimaciones de mercado

La definición del mercado objetivo está dada en términos del número de compradores posibles. O sea aquel al que está dirigida esta nueva marca de agua purificada, es el paso más importante en el análisis de la factibilidad de la instalación de la planta. No será efectiva ninguna prospección del mercado sin conocer las actitudes del consumidor actual y cuantos son consumidores potenciales. Los objetivos de ventas determinaran la factibilidad de la nueva empresa de agua purificada. El plan de ventas de la empresa debe ser alcanzable en un breve plazo. Estos se basarán en una estimación realista de las oportunidades del mercado. Al desarrollar las metas de ventas no hay que olvidar la existencia de factores cuantitativos y cualitativos. Los primeros son aquellos que pueden ayudar numéricamente a calcular las ventas, siendo bueno comenzar con los datos de los años anteriores. Las tendencias del mercado y de la empresa son los aspectos mas importantes a considerar cuando se proyecta para el futuro. Esto es señalado por Jiménez Moreno y García Villaverde (1999) y Klapper (2005).

A partir de los resultados de la encuesta se hicieron las estimaciones de los consumos de cada tipo de envase de agua y del mercado potencial. Llamaremos mercado disponible a todos aquellos consumidores que tienen la necesidad de consumir agua de garrafón y agua embotellada en sus distintas presentaciones, llamaremos mercado real a aquellos consumidores del mercado disponible que compran agua de garrafón y agua embotellada en sus distintas presentaciones y, finalmente, llamaremos mercado potencial al conjunto de consumidores que no consumen el nuevo producto, dado que no existe en el mercado, y que forman parte del mercado real. Las estimaciones del mercado real y potencial se realizan mediante la estimación de los totales respecto al mercado disponible, es decir,  $\hat{T} = Np$  o  $\hat{T} = N\bar{x}$ .

##### 4.1 La demanda del Garrafón de 20L

Como se estableció antes, las purificadoras pequeñas representan el 20.4% del mercado, electropura y diana cubren el 75.6%. Si consideramos las características por las cuales los clientes potenciales cambiarían a una marca nueva agua de garrafón tenemos que 39.8% lo harían por la calidad, 26.8% por su sabor y el 20.02% por su precio. Usando estas proporciones se tiene que las estimaciones del mercado potencial son:

Tabla 4.1 Estimaciones de los potenciales para garrafón

|         | Mercado de agua de garrafón de 20 litros |         |            |            |            |            |            |
|---------|--|---------|------------|------------|------------|------------|------------|
|         | Disponible                               | Real    | Potencial1 | Potencial2 | Potencial3 | Potencial4 | Potencial5 |
| Hogares | 413268                                   | 372768  | 84307      | 312431     | 148362     | 99902      | 74628      |
| Demanda | 1285263                                  | 1159308 | 262194     | 971659     | 461404     | 310694     | 232093     |

Para las estimaciones de mercado potencial se considera como unidad de muestreo a los hogares y para la demanda la unidad de medida es el garrafón. El mercado potencial<sup>1</sup> es el total de hogares que actualmente son atendidos por las purificadoras pequeñas y para estimar la demanda de garrafones habrá que multiplicar esta cantidad por el consumo medio de garrafones, que para este caso es de 3.11 garrafones por hogar. De la misma forma se procede con las restantes estimaciones.

#### **4.2 La demanda de agua embotellada en presentaciones de 2L, 1.5L, 1L, 0.6L y 0.5L**

Para las distintas presentaciones de agua embotellada las estimaciones del mercado potencial se obtienen sabiendo que la proporción de personas que consumen agua embotellada es del 83.02%, la proporción del mercado que cubren las purificadoras pequeñas es de 6.44%, las purificadoras grandes que manejan las marcas Manantial, Bonafón y E-pura representan el 93,56%. Si atendemos a las razones por las cuales cambiarían a una nueva marca, el 35.6% lo haría por la calidad, el 27.98% por el sabor y el 18.9% por el precio. El consumo promedio semanal es de 4.7 botellas.

Tabla 4.2 Estimaciones de los potenciales para garrafón

|         | Mercado del agua embotellada en sus distintas presentaciones |         |            |            |            |            |            |
|---------|--|---------|------------|------------|------------|------------|------------|
|         | Disponible   | Real    | Potencial1 | Potencial2 | Potencial3 | Potencial4 | Potencial5 |
| Hogares | 413268   | 343095  | 26614      | 386654     | 122011     | 96012      | 64745      |
| Demanda | 1942360  | 1612547 | 125088     | 1817272    | 573454     | 451255     | 304304     |

En términos de ventas significa que del mercado real, que en este caso representa un total de 343095 viviendas, la demanda de botellas de agua es de 1 612 547 unidades en sus distintas presentaciones, si competimos con el mercado de las purificadoras pequeñas (26614 viviendas), la demanda de producto será de 125 088 unidades, si competimos con las purificadoras grandes, la demanda es de 1 817 272 unidades. Ahora, si atendemos a las características por las cuales la gente cambiaría a otra marca se tiene que nos centramos en la calidad la demanda estimada es de 573 454 unidades, si se atiende al sabor se estiman una demanda de 451 255 unidades y si atendemos al precio la demanda sería de 304 304 unidades para una población de 64 745 hogares.

## **5. Comportamiento de la lealtad a los productos en el mercado.**

### **5.1 El árbol de clasificación y regresión**

Los modelos de árbol de clasificación y regresión (Classification & Regresión Tree, C&RT), fueron introducidos por Breiman et al. (1984) (ver también Kuncheva (2004) y Sutton (2005)). Estos han sido usados en estudios de mercado, vea por ejemplo Ranilla Y Bahamonde. (1995). Los dividen en modelos de árbol de regresión, cuando la variable de respuesta es cuantitativa y en modelos de árbol de clasificación, cuando la variable de respuesta es cuantitativa. Los modelos C&RT son una técnica que permite explorar la estructura de los datos, también se utiliza como un mecanismo de predicción, entre otros usos. En el caso del presente estudio, la técnica fue utilizada para analizar la fidelidad, por parte de los consumidores, a la marcas de agua embotellada en sus dos presentaciones, en garrafones de 20 litros y en recipientes con volumen menor a dos litros. Este análisis ha posibilitado conocer las características del consumidor para proponer algunos elementos que conduzcan al diseño de una estrategia de penetración en el mercado de agua embotellada, por parte de los ejidatarios del poblado de Azinyahualco.

Para aplicar la técnica de los modelos C&RT, se ha utilizado el paquete estadística S-Plus, puesto que ofrece facilidades para la modelación tanto a través de sus menús, como a través de su código, facilidades que no ofrecen otros paquetes comerciales como el SPSS. Vea Venables y Ripley (2002) y Clark y Pregibon (1992) para obtener detalles sobre el paquete. A pesar de todo, diversos paquetes particulares han sido desarrollados, vea Djukova, E. V. N. V. Peskov (2007), dado que los comerciales no satisfacen las necesidades de las aplicaciones.

Se supone que se tiene una variable de respuesta  $Y$  cualitativa con  $K$  categorías  $C = \{C_1, C_2, \dots, C_K\}$  y  $P$  variables explicativas o independientes  $\{X_1, X_2, \dots, X_P\}$ , donde cada  $X_i$  puede ser cualitativa o cuantitativa. Para aplicar el método se requiere de  $n$  observaciones tanto de la variable de respuesta como de las variables independientes.

La estructura de árbol es utilizada para analizar el particionamiento de la población. Un concepto clave es la de árbol binario. Este está constituido por nodos internos o de división y nodos terminales, al primer nodo interno se le conoce como nodo raíz. Cada nodo interno se divide en nodo hijo izquierdo y nodo hijo derecho. Existen distintos criterios para dividir los nodos internos. La regla de división establece que se selecciona una variable  $X_i$  y un valor  $r$  de tal forma que aquellos datos que satisfacen  $X_i \leq r$ , son asignados al nodo hijo izquierdo, y aquellos que satisfacen  $X_i > r$  son asignados al nodo hijo derecho. Esto es cuando la variable  $X_i$  es cuantitativa. Cuando la variable es cualitativa, se selecciona un subconjunto  $A$  de las categorías de  $X_i$ , tal que, aquellos datos que satisfacen  $X_i \in A$  son asignados al nodo hijo izquierdo y los que no la cumplen son asignados al nodo hijo derecho. Existen diversos criterios para seleccionar la regla de división, uno de ellos utiliza el concepto de deviance (Para un enfoque diferente de la elección de la variable  $X_i$  y el valor de  $r$  ver Loh y Shih (1997)). Bajo el supuesto de normalidad de la variable de respuesta, la deviance es equivalente a la suma de los cuadrados de los residuales (para el caso de modelos de árbol clasificación ver Venables y Ripley (2002)). Un criterio para seleccionar la regla de división, esto es la variable  $X_i$  y el valor  $r$  o en su defecto el subconjunto  $A$ , será aquella regla de división que haga mínima la deviance en el nodo  $t$ . La deviance asociada al árbol de clasificación  $T$  es la suma de la deviance obtenida en cada nodo terminal y la Deviance Media Residual (DMR) será la deviance total dividida por los grados de libertad ( $g.l. = n - b$ ), esto es, la diferencia entre el total de observaciones y el número de nodos terminales. Para construir un modelo de árbol binario, se le aplica la regla de división al nodo raíz,

generando los nodos hijos izquierdo y derecho. A cada uno de estos nodos se le vuelve aplicar la regla de división hasta que se satisface la regla de parada. En el proceso de crecimiento de un árbol, para que un nodo  $t$  sea declarado como nodo terminal, debe satisfacer una de las siguientes condiciones: i) antes de dividir el nodo, que existan en el nodo  $t$  al menos 10 observaciones; ii) después de dividirse el nodo  $t$ , que el número de observaciones de la variable de respuesta en los nodos hijos, al menos sean cinco en cada uno; iii) que el valor de la deviance  $D_t$  en el nodo  $t$ , sea menor que el 1% de la deviance en el nodo raíz. Un subárbol  $T'$ , es aquel árbol obtenido de un árbol  $T$  eliminando algunos nodos. El proceso de podado se realiza sin pérdida significativa de bondad en el ajuste. La forma de medir la importancia de un subárbol  $T'$ , está dado por la llamada medida del costo de complejidad, la cual se define como:

$$D_k(T') = D(T') + \alpha * \text{tamaño}(T')$$

donde  $D(T')$  denota la deviance del subárbol  $T'$ ,  $\alpha$  es el parámetro del costo de complejidad y  $\text{tamaño}(T')$  representa el número de nodos terminales del árbol  $T'$ .

El criterio a seguir es seleccionar aquel subárbol que minimiza la deviance, dado el parámetro de complejidad,  $\alpha$ . Para valores grandes de  $\alpha$  se obtienen subárboles  $T'$ , con un número menor de nodos terminales y viceversa. La categoría  $C_k$  asignada a cada nodo terminal  $t$ , será aquella que tenga la mayor frecuencia de observaciones en el nodo  $t$ . Este valor es considerado como el valor estimado de la variable de respuesta  $Y$  en el nodo  $t$ . Las observaciones de la variable de respuesta  $Y$  que no coinciden con la categoría asignada al nodo terminal se les llaman observaciones mal clasificadas o subclasificadas. La suma de las observaciones subclasificadas de cada uno de los nodos terminales dividida por el total de las observaciones  $n$ , es la tasa de error de subclasificación.

En el paquete estadístico S-Plus existe una opción para ajustar un modelo de árbol de clasificación a los datos, proporcionando la gráfica del árbol y un reporte general donde se incluye las variables del modelo, el número de nodos terminales, deviance media residual y la tasa del error de subclasificación. También se reportan los datos correspondientes a cada uno de los nodos como son: el número de nodo, regla de división, el número de observaciones, la deviance, el valor estimado de la variable de respuesta y las probabilidades estimadas para cada categoría. Este paquete estadístico, en el caso de modelos de árbol de clasificación, en el proceso de podado se puede elegir el criterio de minimizar la deviance o también el criterio de minimizar la tasa del error de subclasificación.

## **5.2 Estudio del mercado del garrafón de 20 litros.**



La pregunta número 11 del cuestionario, principal razón de cambio de marca por parte de la familia, se tomó como variable de respuesta. Para analizar aquellas variables que tienen mayor influencia sobre la variable de respuesta, se realizó la prueba de independencia Ji-Cuadrada, en la Tabla 5.1 se presentan los resultados.

**Tabla 5.1.** Prueba de independencia Ji-Cuadrada de la variable: razón de cambio de marca de agua embotellada (garrafón de 20 litros) contra otras variables.

| Variables              | Descripción           | G. L.(*) | Ji-Cuadrada | P-valor |
|------------------------|-----------------------|----------|-------------|---------|
| ZONA                   | Zona geográfica       | 35       | 192.115     | 0.0     |
| V3                     | Ocupación             | 40       | 44.254      | 0.2968  |
| V4                     | Número de focos       | 95       | 128.671     | 0.0122  |
| V1                     | Sexo                  | 5        | 5.549       | 0.3525  |
| V7                     | Marca de agua consumo | 35       | 87.861      | 0.0     |
| V8                     | Número de garrafones  | 35       | 127.896     | 0.0     |
| V9                     | Razón para el consumo | 25       | 330.789     | 0.0     |
| V10                    | Lugar de compra       | 20       | 68.508      | 0.0     |
| (*) Grados de Libertad |                       |          |             |         |

Las variables V9, V8, V7 y ZONA son las variables de mayor influencia sobre la variable de respuesta, por lo que se ajustaron diversos modelos a los datos, siguiendo una estrategia de ajuste hacia adelante. En la Tabla 5.2 se proporcionan las variables que intervienen en la construcción del modelo de árbol de clasificación y tres características del modelo: los nodos terminales, la Deviance Media Residual (DMR) y la tasa de error de clasificación.

**Tabla 5.2.** Modelos ajustados para el cambio de marca de garrafón.

| No.                         | Nodos terminales | D.M.R. (*) | Tasa de error de Subclasificación | Variables en el modelo ajustado |
|-----------------------------|------------------|------------|-----------------------------------|---------------------------------|
| 1                           | 8                | 2.753      | 0.596                             | V11~ZONA                        |
| 2                           | 33               | 2.693      | 0.575                             | V11~ZONA+V7                     |
| 3                           | 97               | 2.591      | 0.550                             | V11~ZONA+V7+V8                  |
| 4                           | 128              | 2.397      | 0.492                             | V11~ZONA+V7+V8+V9               |
| 5                           | 145              | 2.319      | 0.479                             | V11~ZONA+V7+V8+V9+V10           |
| 6                           | 149              | 2.225      | 0.469                             | V11~ZONA+V4+V7+V8+V9+V10        |
| 7                           | 138              | 2.148      | 0.448                             | V11~ZONA+V2+V4+V7+V8+V9+V10     |
| 8                           | 148              | 2.089      | 0.437                             | V11~ZONA+V2+V3+V4+V7+V8+V9+V10  |
| 9                           | 139              | 2.190      | 0.457                             | V11~V2+V3+V4+V7+V8+V9+V10       |
| 10                          | 143              | 2.110      | 0.439                             | V11~ZONA+V2+V3+V7+V8+V9+V10     |
| 11                          | 143              | 2.135      | 0.446                             | V11~ZONA+V2+V3+V7+V9+V10        |
| 12                          | 147              | 2.157      | 0.451                             | V11~ZONA+V2+V7+V8+V9+V10        |
| 13                          | 144              | 2.131      | 0.444                             | V11~ZONA+V2+V3+V7+V8+V9         |
| 14                          | 147              | 2.214      | 0.462                             | V11~ZONA+V3+V7+V8+V9+V10        |
| 15                          | 146              | 2.194      | 0.459                             | V11~ZONA+V2+V7+V9+V10           |
| (*) Deviance Media Residual |                  |            |                                   |                                 |

Se seleccionan los modelos correspondientes a las líneas número 8, 10 y 11 como los mejores modelos para realizar proceso de podado del árbol, con un parámetro de costo de complejidad igual a 20, mediante el criterio de minimizar la deviance. Las tres características de los modelos de árbol de clasificación y las variables que intervienen en su construcción, se observan en la Tabla 5.3. Los dos primeros modelos tienen la misma tasa de error de subclasificación, además, al eliminar las variables regresoras que no intervienen en la construcción del árbol, las variables restantes, en ambos casos, coinciden y son las variables que se presentan en el último renglón.

Con estas variables se ajusta el modelo de árbol de clasificación y sus características resultantes, se presentan en la última línea.

**Tabla 5.3.** Los 3 mejores modelos para describir la razón de cambio de marca de agua en garrafón.

| No.                         | Nodos Terminales | D.M.R. (*) | Tasa de error de Subclasificación | Variables en el modelo ajustado |
|-----------------------------|------------------|------------|-----------------------------------|---------------------------------|
| 10                          | 14               | 2.609      | 0.535                             | V11~ZONA+V2+V3+V7+V8+V9+V10     |
| 11                          | 14               | 2.609      | 0.535                             | V11~ZONA+V2+V3+V7+V9+V10        |
| 8                           | 17               | 2.588      | 0.538                             | V11~ZONA+V2+V3+V4+V7+V8+V9+V10  |
|                             |                  |            |                                   |                                 |
|                             | 14               | 2.608      | 0.535                             | V11~ZONA+V2+V7+V9+V10           |
| (*) Deviance Media Residual |                  |            |                                   |                                 |

### 5.2.3 Análisis del modelo.

Para determinar las características de los consumidores de agua embotellada que estarían dispuestos a cambiar de marca, se realizan los siguientes ejercicios considerando como base las variables de la última línea de la Tabla 5.3.

i) Se consideran las variables señaladas para construir el modelo de árbol de clasificación y se le aplica el proceso de podado utilizando el criterio de la minimización de la tasa de error de clasificación. El modelo que resulta de este proceso es el subárbol, del árbol que se muestra en la Figura No. 1, cuyos nodos terminales son 2,6 y 7. Este modelo es muy simple puesto que cuenta con sólo tres nodos terminales.

La única variable que se utiliza en la construcción del árbol es la correspondiente a la principal razón para consumir agua embotellada (V9). Las características de los consumidores en los nodos terminales son:

**Nodo terminal 2:** En este nodo se encuentran 21.7% del total de las observaciones, y que corresponden a los que consumen de agua de garrafón por la razón del sabor (V9:a). De este conjunto de consumidores 48.7% estarían dispuestos a cambiar de marca por la razón del sabor del agua (V11:a).

**Nodo terminal 6:** El número de observaciones existentes en este nodo son 185 que corresponde al 8.4% del total. Estas observaciones pertenecen a los consumidores de agua por las razones: por su presentación y por su precio (V9:c,e). El 45.9% de estos consumidores estarían dispuestos a cambiar de marca de agua embotellada por la razón de su precio (V11:e).

**Nodo terminal 7:** El 69.9% de las observaciones totales se encuentran en este nodo y pertenecen a los que consumen agua por las razones: de comodidad, por su calidad y por otra razón (V9: b,d,f). Considerando la calidad del agua embotellada, el 45% de los consumidores que se corresponden a las observaciones de este nodo, estarían dispuestos a cambiar de marca (V11:d).

En base al modelo de árbol obtenido, se puede señalar que más de 2/3 de los consumidores entrevistados, estarían dispuestos a cambiar de marca de agua embotellada por la principal razón de la calidad del agua (V11:d), con una probabilidad de 0.450. Además, estas observaciones corresponden a consumidores de agua embotellada por las razones: por comodidad, por su calidad y por otra razón (V9:b,d y f).

ii) En este caso, se ajusta el modelo con las mismas variables y el proceso de podado se realiza con el criterio de minimización de la deviance, considerando un valor igual a 30 para el parámetro de costo de complejidad. Tomando como referencia el modelo de árbol de la Figura No. 1, el modelo resultante de este proceso de podado es el subárbol que tiene seis nodos terminales los cuales son: 4, 5, 6, 15, 28 y 29.

En la Tabla 5.4, se proporcionan las estadísticas de cada nodo terminal del modelo ajustado. En particular, se proporciona la categoría de la variable de respuesta que es asignada por el algoritmo (cuarta columna) con la estimación de su probabilidad (quinta columna).

**Tabla 5.4.** Características de los nodos terminales del modelo de árbol resultante del proceso de podado considerando un valor del parámetro de costo de complejidad igual a 30.

| No. de nodo terminal | Observaciones |        | Categoría asignada (V11) | Probabilidad estimada de la categoría asignada |
|----------------------|---------------|--------|--------------------------|--|
|                      | Absolutas     | %      |                          |  |
| 4                    | 112           | 5.1    | Por su calidad           | 0.339  |
| 5                    | 368           | 16.6   | Por su sabor             | 0.544  |
| 6                    | 185           | 8.4    | Por su precio            | 0.459  |
| 15                   | 429           | 19.4   | Por su calidad           | 0.485  |
| 28                   | 894           | 40.4   | Por su calidad           | 0.447  |
| 29                   | 223           | 10.1   | Por su calidad           | 0.395  |
| Totales              | 2,211         | 100.0% |                          |  |

Con la finalidad de ilustrar la interpretación del modelo, se describen los nodos terminales 4 y 28.

**Nodo terminal 4:** La cantidad de observaciones en este nodo terminal corresponde al 5.1% del total y pertenecen a la categoría “por su sabor” de la variable V9 y a las categorías: Costa Azul-Icacos (a), Ciudad Renacimiento (d) y Col. Jardín (f) de la variable ZONA. El 33.9% de los consumidores que corresponden a este nodo terminal, opinan que estarían dispuestos a cambiar de marca de agua embotellada por la razón de la calidad del agua (V11:d).

**Nodo terminal 28:** En este nodo se concentra el 40.4% del total de entrevistados y en su opinión estarían dispuestos a cambiar de marca de agua embotellada por la razón de su calidad (V11: d), con una probabilidad de 0.447. Siguiendo los nodos de división del árbol en la Figura No. 1, las características de los consumidores son: tienen alguna de las categorías b, d o f de la variable V9; pertenecen a las categorías c,d,e,f o g de la variable ZONA; y deben ser menores de 54.5 en la variable V2.

iii) Se ajustó el modelo que se ha determinado como el más adecuado, según el análisis de la Tabla 5.3. En este caso el proceso de podado se realiza con un valor del parámetro de costo de complejidad igual a 20 y bajo el criterio de minimización de la deviance. El modelo de árbol de clasificación resultante corresponde a la Figura No. 1.

En la Tabla 5.5 se proporcionan las características de los nodos terminales de árbol de la Figura No. 1.

Comentario acerca del modelo: En la Tabla 5.5, los nodos terminales 15 y 113, concentran más del 50% de las observaciones totales y en ambos nodos, los consumidores manifiestan estar dispuestos a cambiar de marca de agua embotellada por la principal razón de la calidad del producto, esto con probabilidades superiores a 0.45. Se analiza en detalle las características de los consumidores en estos nodos.

Nodo terminal 15: Por la localización de este nodo dentro del modelo de árbol, las características de los consumidores están definidas por sólo dos variables. Los consumidores tienen alguna de las categorías b, d o f de la variable V9; y fueron entrevistados en el lugar b o h, categorías correspondientes a la variable ZONA.

Nodo terminal 113: Las observaciones, 35.1% del total, en este nodo terminal corresponden a los consumidores del agua embotellada que tienen algunas de las categorías b, d o f de la variable V9; tienen una edad (V2) menor que 54.5 años, radican en alguno de los lugares: a, c, d, e o f, los cuales son categorías de la variable ZONA; y tienen alguna de las categorías: a, b, d, e, g o h de la variable V7.

**Tabla 5.5.** Características de los nodos terminales del modelo de árbol resultante del proceso de podado, considerando un valor del parámetro de costo de complejidad igual a 20.

| No. de nodo terminal | Observaciones |        | Categoría asignada (V11) | Probabilidad estimada | Zona Categoría |
|----------------------|---------------|--------|--------------------------|-----------------------|----------------|
|                      | Número        | %      |                          |                       |                |
| 8                    | 54            | 2.4    | Por su calidad           | 0.482                 | a, d, f        |
| 11                   | 155           | 7.0    | Por su sabor             | 0.516                 | e, g, h        |
| 12                   | 26            | 1.2    | Por su precio            | 0.654                 | a, b, ...,h    |
| 13                   | 159           | 7.2    | Por su precio            | 0.428                 | a, b, ...,h    |
| 15                   | 429           | 19.4   | Por su calidad           | 0.485                 | b, h           |
| 18                   | 12            | 0.5    | Por su comodidad         | 0.333                 | a, d, f        |
| 19                   | 46            | 2.1    | Por su sabor             | 0.348                 | a, d, f        |
| 20                   | 199           | 9.0    | Por su sabor             | 0.578                 | b, c           |
| 21                   | 14            | 0.6    | Por su sabor             | 0.357                 | b, c           |
| 57                   | 105           | 4.7    | Por su calidad           | 0.371                 | g              |
| 58                   | 72            | 3.3    | Por su calidad           | 0.458                 | e, f           |
| 59                   | 151           | 6.8    | Por su calidad           | 0.364                 | a, c, d, g     |
| 112                  | 16            | 0.7    | Por su calidad           | 0.688                 | a, c, d, e, f  |
| 113                  | 773           | 35.1   | Por su calidad           | 0.453                 | a, c, d, e, f  |
| Totales              | 2,211         | 100.0% |                          |                       |                |

### 5.3 Caso del mercado de agua embotellada en recipientes de distintos tamaños.

Para realizar el análisis de la fidelidad a la marca, en el caso del mercado de agua embotellada de distintos tamaños, se considera como variable de respuesta: razón para cambiar de marca, pregunta número 18 del cuestionario (Anexo 1). Para indagar qué variables tienen influencia sobre la variable de respuesta, se realizó una prueba de independencia Ji-Cuadrada, cuyos resultados se muestran en la Tabla 5.6.

**Tabla 5. 6.** Prueba de independencia Ji-Cuadrada de la variable: razón de cambio de marca de agua embotellada de diversos tamaños contra otras variables.

| Var.                   | Descripción       | G. L.(*) | Ji-Cuadrada | P-valor |
|------------------------|-------------------|----------|-------------|---------|
| ZONA                   | Zona geográfica   | 35       | 232.564     | 0.0     |
| V3                     | Ocupación         | 40       | 78.455      | 0.0003  |
| V1                     | Sexo              | 5        | 4.350       | 0.5002  |
| V14                    | Marca de agua     | 30       | 67.165      | 0.0001  |
| V15                    | Tamaño del envase | 20       | 37.879      | 0.0092  |
| V17                    | Lugar de compra   | 20       | 0.002       |         |
| (*) Grados de Libertad |                   |          |             |         |

La variable correspondiente a la zona geográfica es la que tiene mayor influencia sobre la variable de respuesta, le sigue en orden de importancia la ocupación (V3). Con estas variables se inicia el proceso de ajuste del modelo de árbol de clasificación, para incluir una a una las variables regresoras. Las características y las variables que se consideran en el ajuste del modelo se muestran en la Tabla 5.7.

Se seleccionan los modelos correspondientes a las líneas número 8, 12 y 14 de la Tabla 5.7, como los mejores modelos de árbol de clasificación, para realizar el proceso de podado con un valor del parámetro de costo de complejidad igual a 20, dando los siguientes resultados. Con el criterio de la tasa de error de clasificación los modelos marcados con los números 8 y 12 son los mejores, después del proceso de podado. En el caso del modelo número ocho, en la construcción del árbol correspondiente, las variables V2 y V4, no son utilizadas, en tanto las cuatro variables que realmente son utilizadas en la construcción del árbol son las que tienen mayor grado de dependencia con la variable de respuesta, según la prueba de independencia Ji-Cuadrada (Tabla 5.6), esta es la razón para utilizar estas variables para la construcción del mejor modelo, como se muestra en la última línea de la Tabla 5.8.

**Tabla 5.7.** Modelos ajustados para el cambio de marca de agua embotellada de diversos tamaños.

| No.                         | Nodos terminales | D.M.R. (*) | Tasa de error de Subclasificación | Variables en el modelo ajustado   |
|-----------------------------|------------------|------------|-----------------------------------|-----------------------------------|
| 1                           | 56               | 2.749      | 0.527                             | V18~ZONA+V3                       |
| 2                           | 32               | 2.790      | 0.601                             | V18~ZONA+V14                      |
| 3                           | 109              | 2.616      | 0.542                             | V18~ZONA+V3+V14                   |
| 4                           | 92               | 2.634      | 0.562                             | V18~ZONA+V3+V17                   |
| 5                           | 123              | 2.592      | 0.551                             | V18~ZONA+V3+V15                   |
| 6                           | 142              | 2.340      | 0.491                             | V18~ZONA+V3+V14+V2                |
| 7                           | 136              | 2.310      | 0.489                             | V18~ZONA+V2+V3+V4+V14             |
| 8                           | 145              | 2.242      | 0.467                             | V18~ZONA+V2+V3+V4+V14+V15         |
| 9                           | 135              | 2.229      | 0.481                             | V18~ZONA+V2+V3+V4+V14+V15+V16     |
| 10                          | 136              | 2.194      | 0.477                             | V18~ZONA+V2+V3+V4+V14+V15+V16+V17 |
| 11                          | 140              | 2.196      | 0.473                             | (Modelo 10) – V3                  |
| 12                          | 133              | 2.236      | 0.469                             | (Modelo 10) – V3 – V14            |
| 13                          | 137              | 2.244      | 0.479                             | (Modelo 10) – V3 – V17            |
| 14                          | 137              | 2.277      | 0.470                             | (Modelo 8) – V3                   |
| (*) Deviance Media Residual |                  |            |                                   |                                   |

Analizamos el modelo para determinar algunas características de los consumidores que estarían dispuestos a cambiar de marca de agua embotellada en presentaciones menores o iguales a dos litros. Ajustamos el modelo con las variables que se encuentran en la última línea de la Tabla 5.6 y aplicando el proceso de podado con los distintos criterios.

**Tabla 5.8.** Mejores modelos ajustados para el cambio de marca de agua embotellada de diversos tamaños

| No.                         | Nodos Terminales | D.M.R. (*) | Tasa de error de subclasificación | Variables en el modelo ajustado |
|-----------------------------|------------------|------------|-----------------------------------|---------------------------------|
| 8                           | 13               | 2.792      | 0.601                             | V18~ZONA+V2+V3+V4+V14+V15       |
| 12                          | 12               | 2.799      | 0.601                             | V18~ZONA+V2+V4+V15+V16+V17      |
| 14                          | 11               | 2.814      | 0.610                             | V18~ZONA+V2+V4+V14+V15          |
|                             |                  |            |                                   |                                 |
|                             | 13               | 2.790      | 0.600                             | V18~ZONA+V3+V14+V15             |
| (*) Deviance Media Residual |                  |            |                                   |                                 |

i) En este caso el proceso de podado se realiza bajo el criterio de minimización de la deviance y con un valor igual a 30 del parámetro de costo de complejidad. El modelo de árbol de clasificación resultante es un subárbol del modelo que se muestra en la Figura No. 2, pero podado hasta los nodos terminales 2, 6 y 7. Este subárbol utiliza solamente la variable de la zona geográfica (ZONA) en su construcción. Las características de los consumidores en los nodos terminales son:

**Nodo terminal 2:** En este nodo se encuentra el 58.5% del total de observaciones (2002) y corresponden a las categorías de la variable de zona geográfica: Costa Azul-Icacos (a), Barrios Históricos (c), Ciudad Renacimiento (d), Col. Emiliano Zapata (e) y Col. Jardín (f). Considerando este subconjunto de observaciones, el 32.5 % de consumidores de agua embotellada estarían dispuestos a cambiar de marca por la razón de su calidad (V18:d).

**Nodo terminal 6:** El 11.5% de las observaciones totales se encuentran en este nodo y corresponden a la categoría de la variable predictora de la zona: La Sabana-La Máquina. Sobre este subconjunto, el 35.9% de consumidores señalan que estarían dispuestos a cambiar de agua embotellada por la razón de su precio (V18: e).

**Nodo terminal 7.** En este nodo terminal se encuentra el 30.0% de las observaciones totales y corresponden a las categorías de la variable de zona geográfica: Centro (b) y El Coloso-Colosio (h). El 42.2% de las observaciones que pertenecen a este nodo terminal, manifiestan que estarían de acuerdo en cambiar de marca de agua embotellada por la razón de su calidad (V18:d).

ii) En este caso también se utilizó las variables de la última línea de la Tabla 5.8, se ajusta el modelo con un valor del parámetro de costo de complejidad igual a 25 y con el criterio de minimización de la deviance. El modelo de árbol obtenido después del proceso de podado tiene 8 nodos terminales, cuyas características se muestran en la Tabla 5.9. Este árbol es realmente un subárbol del modelo mostrado en la Figura No. 2.

**Tabla 5. 9.** Características de los nodos terminales del modelo de árbol de clasificación obtenido mediante el proceso de podado con un valor del parámetro de costo de complejidad igual a 25.

| No. de nodo Terminal | Observaciones |        | Categoría asignada (V18) | Probabilidad estimada de la categoría asignada |
|----------------------|---------------|--------|--------------------------|--|
|                      | número        | %      |                          |  |
| 6                    | 231           | 11.5   | Por su precio            | 0.359  |
| 8                    | 721           | 36.0   | Por su calidad           | 0.300  |
| 9                    | 27            | 1.4    | Otra razón               | 0.556  |
| 10                   | 246           | 12.3   | Por su calidad           | 0.423  |
| 11                   | 177           | 8.8    | Por su calidad           | 0.311  |
| 15                   | 178           | 8.9    | Por su calidad           | 0.573  |
| 28                   | 300           | 15.0   | Por su sabor             | 0.437  |
| 29                   | 122           | 6.1    | Por su calidad           | 0.385  |
| Totales              | 2002          | 100.0% |                          |  |

Análisis del modelo: Se analiza el nodo terminal número 8, por tener el más alto porcentaje de observaciones y los nodos 6 y 15 por zonificar las razones por las cuales el consumidor estaría dispuesto a cambiar de marca de agua embotellada.

**Nodo terminal 6:** En este nodo se encuentran el 11.5% de las observaciones totales y pertenecen a consumidores que estarían dispuestos a cambiar de marca de agua embotellada por la razón de su precio (V18: e) y pertenecen a la zona g (La Sabana-Máquina) de la variable ZONA.

**Nodo terminal 8:** El 36.0% de las observaciones totales se encuentran en este nodo Terminal, que corresponden a los consumidores que estarían dispuestos a cambiar de marca de agua embotellada por la principal razón de la calidad del producto con una probabilidad de 0.30. Estos consumidores fueron entrevistados en algunos de los lugares: c, d o e que corresponden a las categorías de la variable ZONA.

**Nodo Terminal 15:** En este nodo Terminal se encuentran 8.9% de los entrevistados del total y pertenecen al área geográfica denominada El Coloso-Colosio que corresponde a la categoría h de la variable ZONA. De este conjunto de entrevistados 57.3% manifestaron que estarían dispuestos a cambiar de marca de agua embotellada por la razón de la calidad del producto (V18:d).

iii) En la Tabla 5.8 en la última línea se presentan las variables que contiene el modelo que mejor explica el comportamiento de los datos. Se ajusta el modelo con estas variables y se realiza el proceso de podado con un valor del parámetro de costo de complejidad igual a 20, utilizando el criterio de minimización de la deviance. El modelo de árbol de clasificación que se obtiene se muestra en la Figura No. 2.

En la Tabla 5.10, se proporcionan las características de los nodos terminales del modelo de árbol de clasificación para el mercado de agua embotellada en recipientes menores a dos litros. En esta tabla se observa la distribución del total de las observaciones en cada uno de los nodos terminales; además, la categoría de la variable de respuesta (V18) asignada y su correspondiente probabilidad estimada. En la última columna se proporciona las áreas a las cuales pertenecen las observaciones que se encuentran en cada nodo terminal.

**Tabla 5.10.** Características de los nodos terminales del modelo de árbol de clasificación obtenido mediante el proceso de podado con un valor del parámetro de costo de complejidad igual a 20.

| No. de nodo terminal | Observaciones |       | Categoría asignada (V18) | Probabilidad Estimada | Zona Categoría |
|----------------------|---------------|-------|--------------------------|-----------------------|----------------|
|                      | Número        | %     |                          |                       |                |
| 6                    | 231           | 11.5  | Por su precio            | 0.359                 | g              |
| 9                    | 27            | 1.4   | Otra razón               | 0.555                 | c, d, e        |
| 11                   | 177           | 8.8   | Por su calidad           | 0.311                 | a, f           |
| 15                   | 178           | 8.9   | Por su calidad           | 0.573                 | h              |
| 16                   | 333           | 16.6  | Por su sabor             | 0.372                 | c, d, e        |
| 17                   | 388           | 19.4  | Por su calidad           | 0.327                 | c, d, e        |
| 29                   | 122           | 6.1   | Por su calidad           | 0.385                 | b              |
| 40                   | 47            | 2.4   | Por su calidad           | 0.426                 | a, f           |
| 41                   | 60            | 3.0   | Por su calidad           | 0.383                 | a, f           |
| 42                   | 108           | 5.4   | Por su calidad           | 0.463                 | a, f           |
| 43                   | 31            | 1.5   | Por su sabor             | 0.355                 | a, f           |
| 56                   | 158           | 7.9   | Por su calidad           | 0.449                 | b              |
| 57                   | 142           | 7.1   | Por su sabor             | 0.507                 | b              |
| Totales              | 2,002         | 100.0 |                          |                       |                |

### 5.3.1 Análisis del modelo:

La variable regresora más relevante en la construcción del modelo es la variable ZONA, puesto que el algoritmo la selecciona como variable de división en los tres primeros nodos de división. Se hacen algunos comentarios por áreas aprovechando la zonificación que presenta el modelo:

**Áreas a y f:** En esta zona se encuentran los nodos 11, 40, 41, 42 y 43, los cuales contienen en conjunto 21.1% de las observaciones totales. En el nodo Terminal 43 la variable de respuesta toma la categoría “Por su sabor”, sin embargo mantiene una situación de empate con la categoría “Por su calidad”, por lo cual perfectamente se puede optar por esta última. En consecuencia, en esta zona, los consumidores estarían dispuestos a cambiar de marca por la razón de la calidad del agua embotellada, aún cuando en la definición de los nodos terminales intervienen las variables regresoras V3, V14 y V15.

**Área c, d y e:** En esta zona se encuentran los nodos terminales 9, 16 y 17; y concentran el 37.4% del total de observaciones. En esta zona a diferencia de la anterior, en cada nodo terminal la variable de respuesta asume una categoría distinta.

**Área b:** En esta zona se encuentran los nodos terminales 29, 56 y 57 y contienen 21.1% del total de observaciones. En los nodos 29 y 56, la categoría asignada a la variable de respuesta es “Por su calidad” y en el nodo 57 la categoría asignada es “Por su sabor”

**Área g:** En este caso el único nodo terminal que contiene es el número 6, con un 11.5% de las observaciones totales. En este nodo se encuentran los consumidores que estarían dispuestos a cambiar de marca por la principal razón del precio del producto, esto con una probabilidad estimada de 0.359.

**Área h:** El único nodo terminal contenido en esta zona es el número 15, con el 8.9% de las observaciones totales. Estas observaciones corresponden a consumidores que manifiestan la posibilidad de cambiar de marca por la principal razón de la calidad del producto, con una probabilidad de 0.573.

## 6. Conclusiones.

Se ha logrado establecer que las marcas de agua de garrafón con las cuales se tendrá que competir son Electropura y diana, la primera representa el 66.45% del mercado, y la segunda solo participa con 9.21%. Para el caso del agua embotellada en sus distintas presentaciones, las marcas líderes en el mercado son Manantial, Bonafón y e-pura, que representan el 44%, 34,5% y 15%, respectivamente.

Se ha identificado también que el agua de garrafón es adquirido principalmente en misceláneas por aproximadamente cuatro de cada diez consumidores (39.7%) y en los camiones repartidores la adquieren aproximadamente tres de cada diez (34.8%), por lo que en la estrategia de distribución habrá que tomarse en cuenta este hecho. El agua de botella, en sus distintas presentaciones, es adquirida por aproximadamente siete de cada diez consumidores en misceláneas (70.44%), aproximadamente dos de cada diez en tiendas de abarrotes (17.81%) y, aproximadamente, uno década diez la adquiere en el supermercado (7.6%). Esta información permitirá establecer la logística en términos de los lugares en que este se ofertará a la venta.

Se ha confirmado un segmento del mercado para el cual se puede producir y comercializar el nuevo producto si destacamos sus características de calidad, sabor y precio, es decir, poseemos la información apropiada para fijar los precios y las propiedades del nuevo producto.

A partir de la estimación de la demanda potencial (ver tabla 4.1), es posible desarrollar criterios sobre el número de botellones y botellas de agua requeridos para satisfacer la relación oferta-demanda del municipio de Acapulco. Se ha precisado el comportamiento de los clientes respecto a



las marcas existentes en el mercado, la frecuencia de consumo y el lugar de compra de los distintos tipos de envase.

Las dos marcas, Electropura y Diana, predominantes en el mercado de agua en garrafón tienen una participación en conjunto de 75.7%, pero, individualmente tienen una participación menor al 10% en tres lugares: La Colonia Jardín (f), Sabana-Máquina (g) y Coloso-Colosio (h). Para el caso del mercado de agua embotellada, las marcas predominantes, Manantial y Bonafón, participan en conjunto, en el mercado con 78.6%. En el caso de Manantial su presencia está por debajo del 10% en la Colonia E. Zapata (9.7%), La Colonia Jardín (7.5%) y Coloso-Colosio (6.9%); para la marca Bonafón los lugares en los que tiene menor presencia son: Colonia Emiliano Zapata (9.9%), Colonia Jardín (6.0%), Sabana-Máquina (7.4%) y Coloso-Colosio (8.8%). Con estos elementos y conociendo las características de los consumidores mediante los modelos de árbol de clasificación, se pueden diseñar estrategias de penetración en los lugares de baja presencia de las marcas dominantes.

Las principales razones para el cambio de marca obtenidas en el presente estudio son: por la calidad (39.7%), por el precio (20.1%), y por su sabor (26.8%), además mediante el modelo de árbol de clasificación (Figura No.1 del anexo), por área geográfica, se ha podido determinar las características de los consumidores. Se ha logrado definir el perfil de la clientela en términos de sus necesidades, de los factores que caracterizan la calidad del agua purificada y los componentes de la fidelidad a las marcas.

Sobre el análisis y modelación de la disposición del consumidor a cambiar de marca de agua embotellada, se dan los elementos para el diseño de una estrategia de penetración del mercado por parte del nuevo producto.

## **Referencias**

Arcos Cebrián, Antonio y María del Mar Rueda García (2004): El problema del tamaño de la muestra en poblaciones finitas. (En Tamaño de muestra y precisión estadística coord. Antonio Pascual Acosta, Alfredo Martínez Almécija, Joaquín Muñoz García), 19-36. Universidad de Almería, Almería .

Berry, Michael J. A. y Gordon S. Linoff (2004) Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 2nd Edition  
Wiley, Chichester.

BreimanL., Friedman J.H., OlshenR.A. and Stone .J. (1984): Classification and regression Trees. WadsworthInternational Group, Belmont, California.

Clark, L. A. y Pregibon, D. (1992): Tree-based models. En Chambers, J. M. y Hastie, T. J. (Editores): Statistical Models in S. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA.

Coinbio (2008): Proyecto de Conservación de la Biodiversidad por Comunidades e Indígenas de los Estados de Oaxaca, Michoacán y Guerrero. (Consultado 4 de Julio del 2008)  
<http://www.coinbio.org>

Columna de Opinión (2008) Agua purificada es el producto estrella del verano, El Diario de Atacama, Chile, 6 de febrero de 2008

Dominick, Salvatore (2003): Estadística y Econometría, Mc Graw Hill, N. York.

Djukova, E. V. N. V. Peskov (2007): A classification algorithm based on the complete decision tree. MAIK Nauka/Interperiodica , 17, 363-367.

Ferran Aranaz, Magdalena. (1996): SPSS para Windows: Programación y análisis estadístico. Mc Graw Hill, Madrid.

Ferrarese, Laura (2006) Crece la venta de agua a domicilio, La Nación. Argentina, 12 de marzo de 2006

Jiménez Moreno, J.J. y García Villaverde (1999): Metodología para la investigación en marketing y dirección de empresas, Pirámide, Madrid,

Hildebrandt, Lutz y Udo Wagner Larose, Daniel T.(2004): Discovering Knowledge in Data: An Introduction to Data Mining. Wiley, Chichester

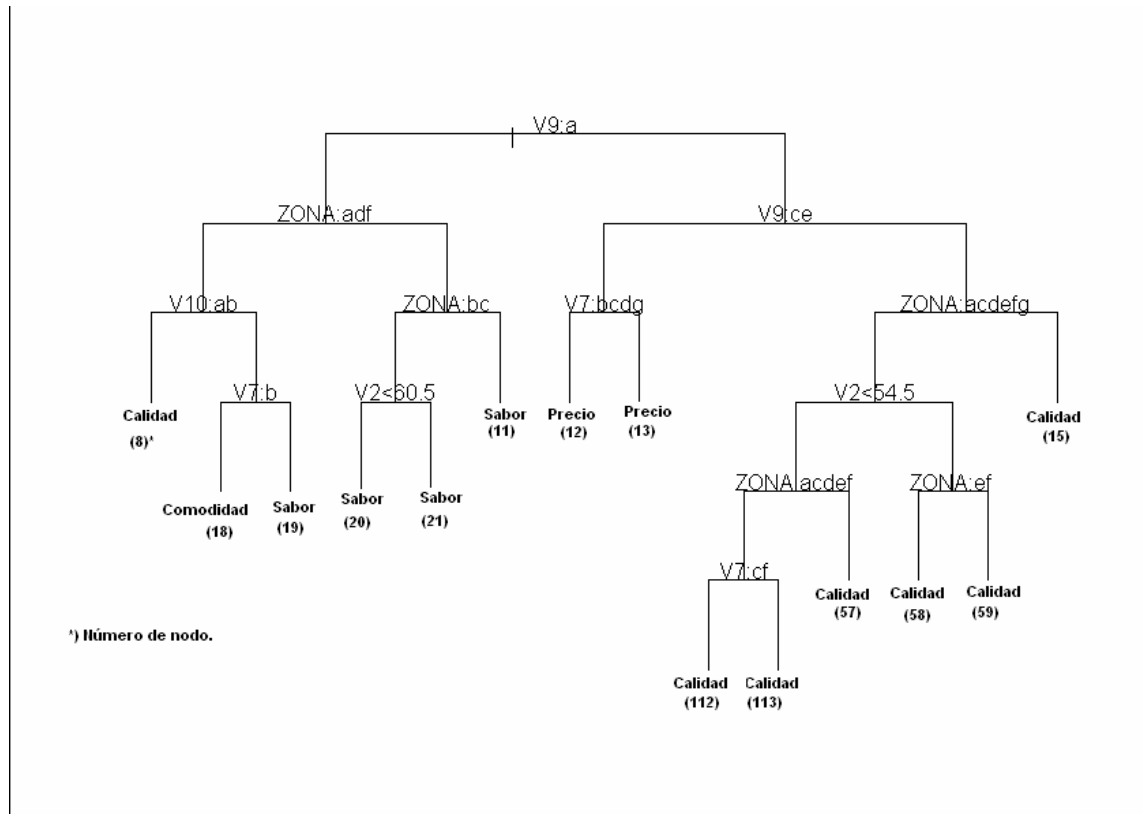
INEGI (2008): Generación e integración de información estadística y geográfica sobre el Censo de Población y Vivienda 2005., (Consultado 11 de Mayo 2008),  
[www.inegi.gob.mx/inegi/default.aspx](http://www.inegi.gob.mx/inegi/default.aspx).

Klapper, Daniel (2005): An econometric analysis of product variety impact on competitive market conduct in consumer goods markets. OR Spectrum. 27, 583-601.

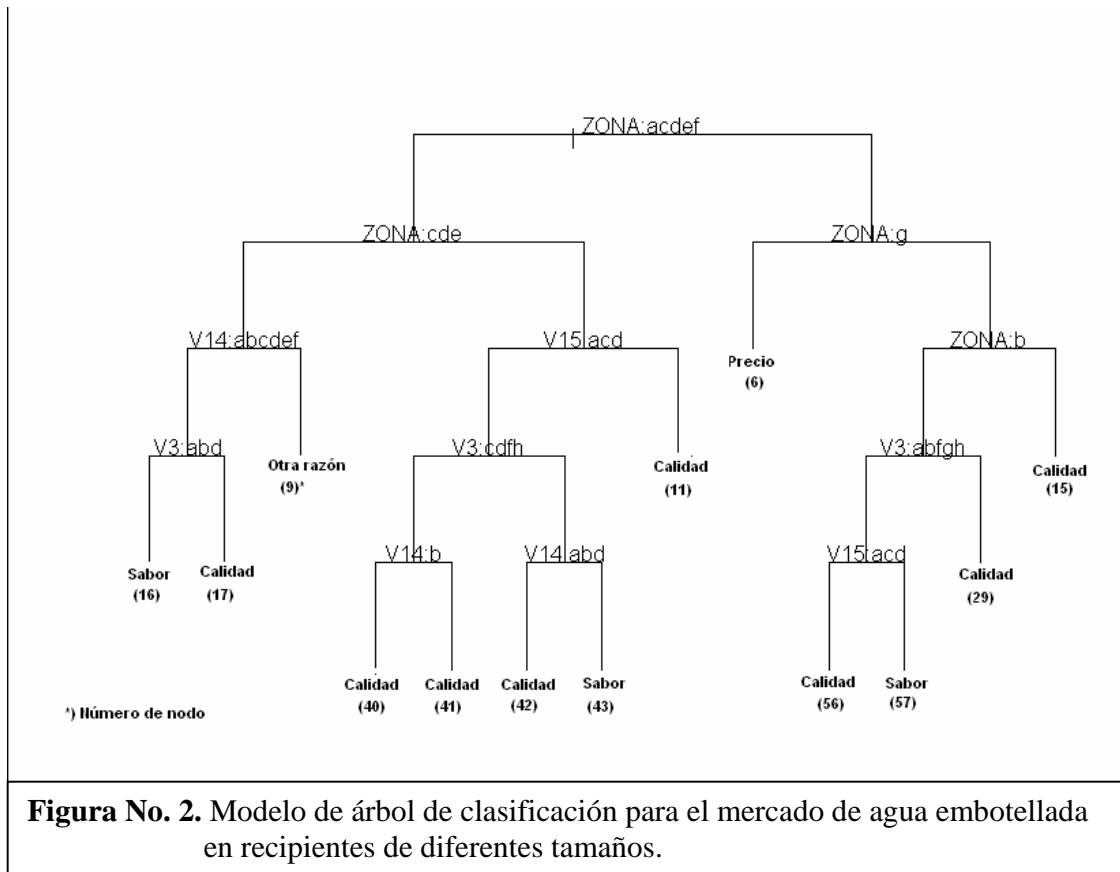
Kuncheva, Ludmila I. (2004): Combining Pattern Classifiers. Methods and Algorithms. John Wiley & Sons. Inc. Publication.USA.

- Levy, Paul S. y Stanley Lemeshow (2008): *Sampling of Populations: Methods and Applications, Solutions Manual, 4th Edition*. Wiley, Chichester.
- Loh, W. y Shih, Y. (1997). *Split Selection Methods for Classification Trees*. *Statistica Sinica*, Vol. 7, pp. 815-840.
- Mena Martínez, Alfredo (2008) *Eleva calor consumo de agua en Juárez*. *El Diario De Chihuahua*, México, 05 de Junio de 2008
- Ranilla, J. Y Bahamonde, A. (1995): "Segmentación de valores numéricos para el aprendizaje a partir de ejemplos". *Actas de la VI Conferencia de la Asociación Española para la Inteligencia Artificial, CAEPIA-95*, 225-234. Alicante.
- Rueda García, María del Mar , Antonio Arcos Cebrián y Andrés González Carmona (1992): *Métodos de muestreo en poblaciones finitas*. Facultad de Ciencias, Granada.
- Rueda García, María del Mar y Antonio Arcos Cebrián (1998): *Problemas de muestreo en poblaciones finitas* Grupo Editorial Universitario, Granada.
- Ruiz Talavera, José Armando (2008): *Estrategia anti calor. Se eleva la demanda de agua purificada y refrescos en Tekax*. *Yucatán Sur*, México 9 de abril de 2008
- Scheaffer Richard L., Lyman Ott R. , Mendenhall William (2006): *Elementos De Muestreo*. Paraninfo. Madrid
- Sutton, Clifton (2005): *Classification and Regression Trees, Bagging and Boosting*. *Handbook of Statistics*, Vol. 24 pp.303-329.
- Venables, W. N. y B.D. Ripley (2002): *Modern Applied Statistics with S-PLUS*. 4ta. Edicion. Springer, Berlin
- Webster, Allen L.(2000): *Estadística aplicada a los negocios y la Economía*. 3ed. Irwin-McGraw Hill, Bogotá.
- Zanutto, Elaine L. y Eric T. Bradlow (2006): *Data pruning in consumer choice models*. *Quantitative Marketing and Economics*, 4, 267-287.
- Zou,Guohua, Feng Shiyong y Qin Huaizhen (2002): *Sample rotation theory with missing data*, *Science in China Series A: Mathematics*, 45, 42-63

ANEXOS



**Figura No. 1.** Modelo de árbol de clasificación para el mercado de agua embotellada en recipientes de 20 litros (garrafrones).



**Figura No. 2.** Modelo de árbol de clasificación para el mercado de agua embotellada en recipientes de diferentes tamaños.

## Anexo 1.

Encuestador: Buenos días (tardes), soy estudiante de la Universidad Autónoma de Guerrero. Estamos realizando un estudio para los habitantes del poblado Azinyahualco ubicado en el municipio de Chilpancingo, los cuales desean poner en marcha una embotelladora de agua natural proveniente de sus manantiales. ¿Me permite hacerle algunas preguntas? Toda la información es general, no necesita dar su nombre. Muchas gracias.

**1.- Sexo.** (Anotar sin preguntar)

1) Masculino                      2) Femenino

**2.- ¿Cuál es su edad?** (Anotar con número)**3.- ¿Cuál es su ocupación?**

1.- Estudiante                      2.- Empleado                      3.- Obrero                      4.- Comerciante                      5.-  
Jornalero  
6.- Ama de casa                      7.- Profesionista                      8.- No trabaja.                      9.- Otro

**4.- ¿Cuántos focos o lámparas hay en su casa?** (anotar la cantidad)**5.- ¿Usualmente en esta vivienda se consume agua de garrafón de 20 litros?**

1.- Si (Pasar a la Pregunta No. 7)                      2.- No. (Pasar a la siguiente pregunta)

**6.- ¿Cuál es la razón principal por la que no se consume agua de garrafón en esta vivienda?**

1.- Por el precio                      2.- En casa hierven el agua                      3.- La consumen de la toma  
4.- Por costumbre  
5.- Desconfianza                      6. Se usa filtro en casa                      7.- Otra razón. (Pasar a la pregunta No. 12)

**7.- ¿Cuál es la marca del agua de garrafón que se consume en esta vivienda?**1.- Electopura    2.- Diana    3. Peyper    4. Luvi    5. Florida water    6. La sierra    7. Shanito  
8. Otra**8.- ¿Cuántos garrafones de 20 litros se consumen a la semana en esta vivienda?**

Anotar el número de respuesta

**9.- En su opinión, ¿Cuál es la principal razón para consumir el agua de garrafón?**

1.- Por su sabor    2.- Por comodidad    3.- Por su presentación    4.- Por su calidad    5.- Por su precio    6. Otra razón.

**10.- Usualmente, ¿Dónde compra el agua de Garrafón?**

1.- Miscelánea    2.- Tienda de abarrotes    3.- Supermercado    4.- Camión    5.- Otros

**11.- ¿Cuál sería la principal razón para que su familia pudiera cambiar de marca de agua de garrafón?**

1.- Por su sabor    2.- Por comodidad    3.- Por su presentación    4.- Por su calidad    5.- Por su precio    6.- Otra razón

**12.- ¿Usualmente Usted consume agua embotellada?**

1.- Si \_\_\_\_ (Pasar a la Pregunta No. 14).

2.- No \_\_\_\_ (Continuar con la siguiente Pregunta----)

**13.- ¿Cuál es la razón principal para no consumir agua embotellada?**1.- Por el precio    2.- En casa hierven el agua    3.- La consumen de la toma    4.- Por costumbre  
5.- Desconfianza en el agua embotellada    6.- Otra razón. ( Terminar la entrevista)**14.- ¿Qué marca de agua embotellada usualmente consume?**

1.- Manantial    2.- Bonafont    3.- E-Pura    4.- Diana    5.- Ciel                      6.- Agua fiel    7.- Otra

**15.- Usualmente, ¿Cuál es el tamaño de la botella de agua que consume más frecuentemente?**

1.- Medio litro    2.- 600 ml.    3.- un litro    4.- litro y medio.    5.- dos litros

**16.- En una semana ¿Cuántos veces consume el agua embotellada del tamaño señalado?**

(Anotar la cantidad)

**17.- Usualmente, ¿Dónde compra el agua embotellada?**

1.- Miscelánea    2.- Tienda de abarrotes    3.- Supermercado    4.- Camión    5.- Otros

**18.- ¿Cuál sería la principal razón para que pudiera cambiar de marca de agua embotellada?**

|   |
|---|
| 1. Por su sabor 2. Por comodidad 3. Por su presentación 4. Por su calidad 5. Por el precio<br>6. Otra razón |
|---|

**¡Muchas gracias por su atención!**