



ISSN: 1886-8452
Vol. 4 Número 13, diciembre 2012
<http://www.eumed.net/rev/tecsistecat1/index.htm>

**DECISIÓN DE COMPRA BAJO LA CONTINGENCIA DE ELECCIÓN BINARIA
UTILIZANDO LA BASE DE SEGMENTACIÓN DEMOGRÁFICA DEL
CONSUMIDOR. CASO: ACEITES COMESTIBLES.**

Dr. Julio César Ceniceros Angulo.

Profesor en la Universidad de Occidente, en Sinaloa, México.

Departamento de Ciencias Económico-Administrativas.

Miembro del Sistema Sinaloense de Investigadores y Tecnólogos de Sinaloa.

julioceniceros@gmail.com

Resumen.

El propósito de este trabajo consiste en describir en primera instancia las variables de segmentación demográfica consideradas en el estudio de como las personas tienden a discriminar entre dos criterios básicos de compra: el de diferenciación de productos vs. precios. En una segunda etapa de la investigación se progresa en las pruebas de hipótesis con fundamento en Chi-cuadrado y estadísticos Wald para mostrar la significancia de estos criterios de segmentación en la explicación del criterio de compra seleccionado por el consumidor. Esta investigación se realiza en la ciudad de Culiacán, Sinaloa México a finales de 2012, a través de una muestra de 349 compradores. Los resultados obtenidos nos muestran una relación significativa para un nivel $\alpha=0.05$, para solo 2 de 8 variables; el nivel

educativo y el nivel de ingresos del consumidor. Por tanto, se cuenta con una relación concluyente y estadísticamente significativa.

Palabras Clave: Segmentación demográfica, Decisión de compra binaria, Estadístico de prueba de Wald. **JEL:** C25 , M31

Abstract.

The purpose of this paper is to describe demographic segmentation variables considered in the study in the first instance of as people tend to discriminate between two basic buying criteria: that of differentiating products vs. prices. In a second stage of research progress in testing of hypotheses based on Chi-square and statistical Wald to show the significance of these segmentation criteria in the explanation of the buying criteria selected by the consumer. This research is conducted in the city of Culiacan, Sinaloa, Mexico in late 2012, through a sample of 349 buyers. The results obtained show a significant relationship to a level $\alpha = 0.05$, for only 2 of 8 variables; the educational level and the level of income of the consumer. Therefore, it boasts a conclusive and statistically significant relationship.

Keywords: Demographic segmentation, binary purchase Decision, Wald test.

JEL: C25, M31

1.- Introducción.

El antecedente inmediato de la presente investigación se deriva directamente de la investigación precedente *Diferenciación de Productos a través de Regresión Logística Múltiple. El caso de los Alimentos Básicos*. Que el autor presento como requisito en la investigación para la consecución del grado de Doctor en Administración, derivando de ello una serie de ventanas de oportunidad para futuras indagaciones, siendo en esta ocasión, el planteamiento relativo a la importancia significativa de las características de segmentación en los compradores de productos básicos que optan por el requisito para la adquisición de satisfactores el criterio de diferenciación de productos en contrapartida al criterio prevalente de los precios. De esta forma se establece una liga natural con la temática tratada por el autor de la cual se retoman y trabajan algunos aspectos íntimamente relacionados.

Por consiguiente, el propósito de trabajar en el diseño de modelos para el área de negocios y estudios de mercados concentrando fuertemente la atención, particularmente en el campo de la diferenciación de productos y, las variables de segmentación en las decisiones de los consumidores como elementos concluyentes en la discriminación que se hace entre el criterio de precios contra el basado en la diferenciación de atributos en el producto, tiene correspondencia con el reto que se plantean las organizaciones de distinguirse del resto de los competidores entre empresas y productos y, en conocer las variables estratégicas en la segmentación de su mercado y, con ello lograr un posicionamiento efectivo teniendo como corolario deseable en consecuencia, un cierto grado de control monopólico sobre el precio de la mercancía que ofrece en el mercado.

De esta manera, es de trascendencia mayor la actividad de identificar, seleccionar y especificar las principales características (de los consumidores), que deben ser predichos y analizados en el proceso para la incorporación de éstos elementos en la explicación de la importancia que revisten en la toma de decisiones por parte del comprador de productos básicos.

Paralelamente, los esfuerzos y la tendencia hacia la definición del tipo de modelos a utilizar para propósitos de pronósticos y de diagnósticos comerciales direccionan hacia los prototipos multimétodos, así, según Padgett, citado por Kerlinger (2001), éstos, consisten en una mezcla de elementos cuantitativos y cualitativos en la explicación de un fenómeno de interés, por ejemplo en las áreas de la administración, economía, mercadotecnia, entre otras disciplinas sociales. Lo que motiva fuertemente la investigación a abreviar en el camino de los modelos de elección discreta.

De acuerdo a Ceniceros (2001), hasta aquí es clara la necesidad de establecer lineamientos para encauzar correctamente el esfuerzo en la consecución de estos propósitos a saber el diseño de modelos; sin embargo, esta etapa plantea los siguientes aspectos, primero; la inserción de los negocios en un ámbito de competencia internacional antes no vista, lo que traza el problema de reingeniería de procesos y la consideración de escenarios cambiantes para la instrumentación de estrategias de competitividad en los negocios, ante la vorágine de acontecimientos en el corto plazo que cambia rápidamente la posición competitiva de los países y sus mapas tecnológicos. En segundo lugar, el cambio vertiginoso en las tendencias económicas, financieras, tecnológicas, políticas, condiciona a las organizaciones comerciales

a un estado de la naturaleza de incertidumbre en donde el factor información adquiere una dimensión vital.

Un ambiente de esta naturaleza dificulta el proceso de toma de decisiones en la empresa, de allí la importancia capital de trabajar arduamente en el diseño de estrategias (diferenciación de productos, segmentación, *v.*, *gr.*) para lograr posicionar efectivamente a la empresa y sus productos, distinguiéndola de los competidores. Una vez, que el administrador se da cuenta de la complejidad del macro ambiente de los negocios, desecha generalmente los métodos poco robustos y, se compromete seriamente con análisis que utilizan una mayor cantidad de información y planeación de sus actividades, buscando establecer o distinguir su ventaja competitiva.

Sin duda, es manifiesta la importancia de diferenciar los productos alimenticios básicos ya que trae entre otros beneficios para la empresa, el contar con un poder de mercado y cierto control sobre el precio que cobra al consumidor, dándole cierto grado monopólico y, adicionalmente un importante posicionamiento del producto en la mente del consumidor, lo que permite a la empresa buscar en el más largo plazo la fidelidad de los clientes hacia el producto, generando de paso confianza sobre el mismo y prolongando su horizonte de vida en el mercado; situación que transfiere como consecuencia deseada un importante flujo económico en pago de remuneraciones y desde el punto de vista social, quizás lo más importante, la conservación de las plazas de trabajo, entre otros beneficios. Sin embargo, para que ello fructifique, antes es necesario segmentar o estratificar a los grupos de consumidores por lo que reviste de especial importancia el correlacionar y determinar la contribución individual de cada variable de segmentación en las personas que opten o se decidan por el criterio de la diferenciación de atributos en la toma de decisiones con relación a los productos básicos a adquirir frente a la alternativa decisoria del criterio de los precios, de allí, que en esta propuesta de investigación se busque explorar por la estrategia de elecciones discretas, a través de un modelo pues, que discrimine la variable dependiente en forma binaria (precios vs atributos).

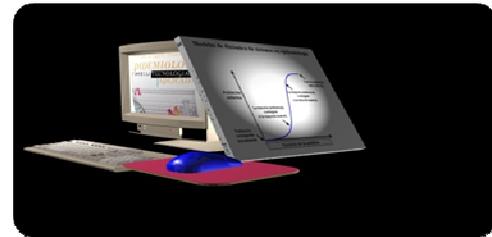
Observemos también que, con la necesidad descrita hasta ahora, es de suma trascendencia el desarrollo de prototipos para la diferenciación de productos en base a criterios de segmentación a través de modelos de elección discreta ya que, en la disertación teórica de la mercadotecnia en este tema se ha

enfaticado en aspectos propiamente generales al reconocer solo orientaciones en el tratamiento del tema de interés presentando un área de procedencia para la investigación y la práctica ante la ausencia de modelos que vengan a contribuir en forma práctica y concreta a resolver falencias en este campo de conocimiento.

2.-Estrategia Metodológica.

2.1. Enfoque de investigación.

En cuanto al tratamiento y cumplimiento de los objetivos planteados en la investigación se guiarán por el acatamiento a la metodología tradicional econométrica, concretizado a través del modelo de regresión logística Binaria (RLB), en donde en primera instancia se parte del planteamiento de la teoría y de las hipótesis, para posteriormente especificar el modelo de acuerdo a la aportación teórica, obtener la información y estimar los parámetros del prototipo econométrico.



2.2. Contexto de la investigación.

La presente investigación, se desarrolla en la ciudad de Culiacán, Sinaloa con el levantamiento de la información. La estratificación del área se realizará dividiendo la ciudad en 4 regiones (Norte, Sur, Noreste y Noroeste) abarcando prácticamente la totalidad de los centros comerciales (hipermercados), en donde los consumidores ordinariamente realizan la compra de los alimentos básicos. A continuación, se presenta la relación de hipermercados y su correspondiente número de personas encuetadas en esos sitios:

Tabla 1. Negociaciones y número de encuestas aplicadas.

Negociaciones	Frecuencia	Porcentaje	Porcentaje acumulado
Ley Juan José Ríos	20	5.7	5.7
Ley Rubí	20	5.7	11.5
Ley los Ángeles	15	4.3	15.8
Mz Centro	20	5.7	21.5
Mz Las Américas	20	5.7	27.2
Bodega Aurrera Estadio	10	2.9	30.1
Mega Campiña	10	2.9	33.0
Ley Express Villa bonita	10	2.9	35.8
Mz Calzada	16	4.6	40.4
Mz Guadalupe	13	3.7	44.1
Wal-Mart Montebello	31	8.9	53.0
Soriana Abastos	21	6.0	59.0
Ley Plaza Fiesta	15	4.3	63.3
Ley Del Valle	15	4.3	67.6
Mz Lomalinda	16	4.6	72.2
Ley Palmito	16	4.6	76.8
Soriana Zapata	16	4.6	81.4
Soriana Barrancos	12	3.4	84.8
Ley Humaya	20	5.7	90.5
Mz plaza norte	10	2.9	93.4
Mz santa fe	10	2.9	96.3
Wal-Mart Humaya	13	3.7	100.0
Total	349	100.0	

Fuente: Elaboración Propia.

2.3. Universo y procedimiento muestral.

El universo de la presente investigación se encuentra conformado por todas aquellas personas que realizan la compra de productos básicos en la ciudad de Culiacán, no pudiéndose determinar numéricamente ya que, en algunas ocasiones se trata de hogares con solo un decisor, en otros casos con hogares de 2 o más personas que realizan las compras. Con relación al tipo y procedimiento en la determinación y selección de los participantes en la muestra tenemos lo siguiente:

Se realiza una nueva prueba piloto durante el verano del 2011, para determinar el tamaño de muestra aplicada a compradores, para contrastarla con la realizada durante los días 12 y 15 de Junio de 2008, en hipermercados de la localidad con el propósito de determinar los nuevos valores de p (proporción de compradores que privilegian el precio en la compra de productos básicos) y $q = 1 - p$, a través del

error de 5.0% y un nivel de confianza del 95%, así como las proporciones que se obtendrán con anterioridad en la prueba piloto, en la determinación del tamaño mínimo de muestra. Es importante recordar que en el trabajo realizado durante el año 2008 se aplicaron 310 cuestionarios. Ahora se aplicarán con los valores actualizados de p y q un total de 349 cuestionarios.

2.4. Diseño de investigación.

El presente trabajo se corresponde con un diseño de investigación de tipo no experimental en donde se relacionan y explican a través de una función de tipo discreta la variable dependiente binaria (diferenciación de producto frente a precios), por medio de covariables (segmentación), explicativas seleccionadas a través de la aportación teórica sobre segmentación de consumidores. Así mismo, se trata de un estudio de tipo transversal.

2.5 Proceso de recolección de datos.

La encuesta se realiza preferentemente en el horario de 10 a.m. a 12 p.m., seleccionando en forma aleatoria sistemática en intercepción en centro comercial a cada $k = 3$ sujetos y, buscando que cada establecimiento comercial cubriese geográficamente la región seleccionada (cuadrante) y determinando el número de entrevistas a realizar.

2.6. Procedimiento general y procesamiento de datos.

Para llegar a estimar definitivamente el modelo con fines de diagnóstico y predicción, se procede en términos generales de la forma siguiente:

- _ Revisión teórica de los factores de segmentación, diferenciación y modelos de elección discreta.
- _ Análisis de la información secundaria, lo que permitirá comprender de mejor manera el problema de investigación.
- _ Selección del método de recogida de datos en escala métrica y no métricas a través de la técnica de la encuesta.
- _ Preparación y redacción del cuestionario.

- _ Diseño del plan de muestreo y determinación del tamaño de la muestra.
- _ Plan para análisis de la información.
- _ Estimación los parámetros del modelo seleccionado que permita entender la contribución y explicación de cada variable de segmentación seleccionado en la diferenciación del producto.

En cuanto al procesamiento y análisis los datos, se utilizará el paquete estadístico SPSS V.19.

3.- Los paradigmas de dependencia: regresión logística simple y múltiple en el estudio de la segmentación y la diferenciación de productos.

En el presente tratado consideremos una primera aproximación a los modelos para validar empíricamente el problema de la diferenciación de productos alimenticios básicos, entonces se parte de la ecuación (1) con la incorporación de un término de error estocástico (ε):

$$(1) \quad k = \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \dots + \omega_n x_n + \varepsilon.$$

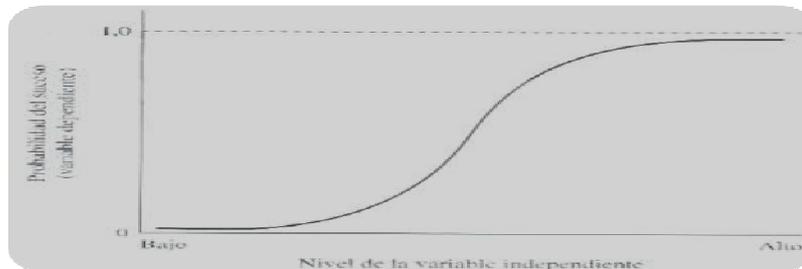
Por lo tanto, de la expresión anterior, se pueden identificar los siguientes elementos estructurales, *vid.*, Gujarati (2010), que permitirán en lo subsiguiente ir revisando y evaluando la factibilidad de los modelos propuestos en la consecución de los objetivos planteados en la investigación:

- 1.) κ , *valor teórico (variable respuesta).*
- 2.) ω_p , *parámetros o coeficientes de la ecuación.*
- 3.) x_{p_n} , *variables independientes.*
- 4.) ε , *residual o término de error estocástico.*

Analicemos, pues, el modelo de Regresión Logística (RL), a la luz de su estructura funcional. Primero, con relación a los valores κ , tenemos que para Hair, Anderson, Tatham y Black (1999), en una primera versión de este modelo la consideran como una variable dicotómica (binaria), es decir, se refieren a una variable respuesta de dos grupos, a diferencia de la Regresión Múltiple, (RM) que predicen las probabilidades de ocurrencia del fenómeno a analizar. Por lo que los valores respuesta se encuentran acotados entre los valores 0 y 1.

Para modelar la relación funcional entre κ y las x_n Hair., *et., al.*, (1999), nos presentan la siguiente representación sigmoide:

Figura 1. Representación sigmoidea de la función logística.



Fuente: Tomado de Hair., *et., al.*, (1999: P. 281)

Especificando, la parte generalizada funcional de κ , en su forma operativa, según Gujarati (2010), tenemos que:

Si p_i = probabilidad de éxito de un evento determinado.

Una forma de modelar un problema con variable dependiente dicótoma, puede ser:

$$(2) \quad p_i = \frac{1}{1 + e^{-z_i}} \quad \text{Función de Distribución Logística.}$$

$$\text{Donde } z_i = \beta_1 + \beta_2 x_i$$

∴ La probabilidad de no ocurrencia del evento, se puede establecer como:

$$1 - p_i = \frac{1}{1 + e^{-z_i}} \Rightarrow \text{La variable respuesta puede quedar expresada como la siguiente razón}$$

$$\text{de probabilidades (odds ratio): } \frac{p_i}{1 - p_i} = \frac{1 + e^{z_i}}{1 + e^{-z_i}}$$

Ahora bien, este modelo también puede ser presentado de la siguiente manera en relación a su variable respuesta, así para Pyndyck y Rubinfeld (2001), el modelo se basa en la siguiente expresión de probabilidad logística acumulativa:

$$(3) p_i = f(z_i) = f(\alpha + \beta x_i) = \frac{1}{1 + e^{-z_i}} = \frac{1}{1 + e^{-(\alpha + \beta x_i)}}$$

Donde e , base de logaritmos naturales ≈ 2.718 , el autor retoma (3) y multiplica ambos lados de la ecuación por $1 + e^{-z_i}$ y se obtiene $(1 + e^{-z_i})p_i = 1$, para luego dividirlo entre p_i y restándole 1,

tenemos: $e^{-z_i} = \frac{1}{p_i} - 1 = \frac{1 - p_i}{p_i}$, como $e^{-z_i} = \frac{1}{e^{z_i}} \therefore e^{z_i} = \frac{p_i}{1 - p_i}$ aplicando el logaritmo natural

en ambos lados, tenemos que:

$$Z_i = \log \frac{p_i}{1 - p_i}$$

Por tanto, retomando (3), finalmente se puede expresar la variable respuesta como:

$$(4) \log \frac{p_i}{1 - p_i} = Z_i = \alpha + \beta x_i$$

Autores clásicos en la Regresión Logística son Hosmer y Lemeshow (1989), que en su trabajo clásico Applied Logistic Regresión, razonan de la siguiente manera en relación al valor esperado de la variable respuesta en una función lineal como:

$$E(y/x) = \beta_0 + \beta_1 x$$

De donde se establece que x se mueve en rangos de $-\infty$ y ∞ . Pero con variables de respuesta de tipo

dicotómica los rangos se establecen en $[0 \leq E(y/x) \leq 1]$ Si, $\pi(x) = E(Y/x)$, Por lo tanto el

modelo Logístico se especifica como:

$$(5) \pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Finalmente, efectúan una transformación logística definiéndola en términos de: $\pi(x)$, así,

$$(6) \varphi(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x$$

Para seguir a Ferrán (2001), digamos que: $\frac{\pi(x)}{1 - \pi(x)} = \frac{\delta}{\lambda}$ y $\varphi = \beta_1 x_1 + \dots + \beta_P x_P + \beta_0$, entonces:

$\lg \left(\frac{\delta}{\lambda} \right) = \varphi$ \therefore una forma adicional de presentar el modelo es:

$$(7) \left(\frac{\delta}{\lambda} \right) = e^{\beta_0} (e^{\beta_1})^{x_1} \dots (e^{\beta_P})^{x_P}$$

Segundo, con relación a ω_p , parámetros o coeficientes de la ecuación (en términos generales), se tiene

que para Gujarati (2000), esos parámetros quedan expresados en términos de las siguientes literales: β_1

y β_2 , así tenemos que, si: $\ln \left(\frac{P_i}{1 - P_i} \right) = Z_i$, \therefore linealizando la expresión, $Z_i = \beta_1$ y $\beta_2 x_i$, ahora

bien, ¿qué interpretación hace el autor de estos coeficientes o parámetros?. En el contexto de un problema que relaciona los niveles de ingreso con las probabilidades de adquirir una casa, Gujarati (2010), comenta:

La interpretación del modelo logit es la siguiente: β_2 , es la pendiente, mide el cambio en ℓ ocasionado por un cambio unitario en x , es decir, dice cómo el logaritmo de las probabilidades a favor de poseer

una casa cambia a medida que el ingreso cambia en una unidad, por ejemplo US \$ 1000. El intercepto

β_1 es el valor del logaritmo de las probabilidades a favor de poseer una casa si el ingreso es cero.

(P.544).

Ahora bien, si la variable dependiente queda expresada como un odds ratio, según Hair, et., al. (1999),

entonces, los coeficientes quedan expresados como exponentes en la siguiente expresión:

Tabla 4. Signos e interpretación de los odds ratio.

Fuente: Elaboración propia con información de Hair, et., al. (1999:P.283).

Por lo que es necesario, volver a transformarlos al aplicarles el *anti log*, los signos de los coeficientes, entonces se interpretarían de la siguiente manera:

$$(8) \frac{prob_{(evento)}}{prob_{(noevento)}} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \Lambda}$$

Complementando lo anterior, de acuerdo a Álvarez (1995), con respecto al significado de los coeficientes en la explicación o contribución de la variable de respuesta binaria tenemos que:

El signo de los coeficientes tiene un significado importante. Si los coeficientes de las variables son positivos, eso significa que la variable aumenta la probabilidad del suceso que estamos estudiando. Si este fuera una enfermedad, el factor cuyo coeficiente es positivo aumentaría la probabilidad de padecer la enfermedad y, por lo tanto, dicho factor sería un factor de riesgo. Si el coeficiente es negativo, el factor cuyo coeficiente es negativo disminuye la probabilidad del suceso que estamos estudiando; en caso de que dicho suceso fuera una enfermedad, estaríamos ante un factor de protección. (P.158).

Signos de los β_i	Transformación anti log	Odds ratio
positivo	> que 1	Aumentara
negativo	< que 1	Disminuirá
cero	= 1	No produce cambios

Tercero, en lo que se refiere a las variables independientes o explicativas en el modelo (RLS), un modelo simple solo incluye una variable explicativa, pudiendo representarse de la siguiente manera:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1)}} , \text{ pero así mismo, un modelo se puede especificar como un modelo múltiple de la}$$

$$\text{siguiente forma: } (9) p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}} ,$$

Es importante destacar que las variables explicativas pueden ser tanto cuantitativas como cualitativas. En el caso de las cualitativas es necesario convertirlas en dummy. Ésta es una variable cualitativa, siempre y cuando tengan la propiedad de de ser codificadas en forma numérica con la regla conocida de que si se tienen K , numero de categorías, entonces habrá que crear $K-1$ variables dummy. Álvarez (1995).

Pasemos al examen, ¿Qué sucede cuando en un modelo RL, se tienen más de una variable explicativa? Para Álvarez (1995), es necesario comprobar si existe efecto interacción entre las variables consideradas en el modelo, de tal forma que si retomamos la ecuación (9) y la simplificamos, obtenemos la ecuación

(10):
$$p_i = \frac{1}{1 + e^{-z_i}}$$
 el exponente puede tener incluidos varias χ^s , por ejemplo si:

(11)
$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_3 + \beta_5 x_1 x_2 x_3$$

Entonces, se habla de un modelo con interacción binaria ($\beta_4 x_1 x_3$) en la primer parte y de una interacción terciaria ($\beta_5 x_1 x_2 x_3$), en la última parte. Para obtener al final un modelo de regresión logística múltiple (RLM).

Cuarto, como parte estructural del modelo tenemos el término de error, residuales o perturbación aleatoria (ε). Así, Pyndyck y Rubinfeld (2001) relacionan la probabilidad de éxito con las variables explicativas, suponiendo que la media del residual es cero. Luego, púes:

$$E(\varepsilon_i) = (1 - \alpha - \beta x_i) p_i + (-\alpha - \beta x_i) (1 - p_i) = 0$$
, por lo que en términos de $p_i = \alpha + \beta x_i$,

$$\therefore 1 - p_i = 1 - \alpha - \beta x_i$$

Tabla 5. Distribución de probabilidades de ε_i .

y_i	ε_i	Probabilidad
1	$1 - \alpha - \beta x_i$	p_i
0	$-\alpha - \beta x_i$	$1 - p_i$

Fuente: Tomado de Pyndyck y Rubinfeld (2001: P. 314)

Otra forma de formular lo anterior lo tenemos en Hosmer y Lemeshow (1989), cuando fija el valor de salida como: $y = \pi(x) + \varepsilon$, donde π , es la probabilidad de éxito del evento considerado, \therefore si $y = 1$, entonces, $\varepsilon = 1 - \pi(x)$ con probabilidad $\pi(x)$ y el caso complementario, si $y = 0$, entonces $\varepsilon = -\pi(x)$, con probabilidad $1 - \pi(x)$, por lo que el residual se distribuye de acuerdo a $\sim (0, \pi(x) [1 - \pi(x)])$

En donde la media de una distribución binomial, se obtiene de $\mu = n\pi$, en cambio la varianza, se obtiene de $\sigma^2 = n\pi(1 - \pi)$. Según Lind, Marchal y Wathen (2005).

Para concluir, Gujarati (2010), plantea que la distribución del error (u_i), cuando el número de casos es elevado (N), sigue una distribución normal (N) como:

$$(12) \quad u_i \sim N\left[0, \frac{1}{N_i p_i (1 - p_i)}\right]$$

4.- Segmentación Demográfica.

En este apartado es importante destacar en este bloque un total de 8 variables dentro de la base o factores de segmentación demográfica. Estas son; Genero, edad, número de personas que contribuyen en el gasto familiar, número de habitantes por hogar, educación, ocupación, estado civil e ingreso mensual. De la misma forma importante es, prevenir sobre la dificultad en la presentación de la información en forma desplegada por la gran cantidad de categorías de respuesta en total. En este sentido se procederá a relacionar aquellas variables con fuerte implicancia teórico mercadológica y, con nivel de medición categórica en función de la variable dependiente. Para ello, cada una de las variables independientes sucesivas (covariables) por si solas (manteniendo constantes las demás), se someterán a la prueba de hipótesis nula inicial de que no se encuentra significativamente relacionada con la variable dependiente. Esto con el propósito de ir detectando en primera instancia aquellas variables que son sospechosas de ser buenas candidatas en la conformación del modelo y, por otro lado que haga posible detectar dentro del bloque estudiado el peso específico de cada variable de segmentación. Para comprender mejor el proceso enseguida se procederá a realizar la corrida con aquellas variables candidatas a ser incluidas en el modelo.

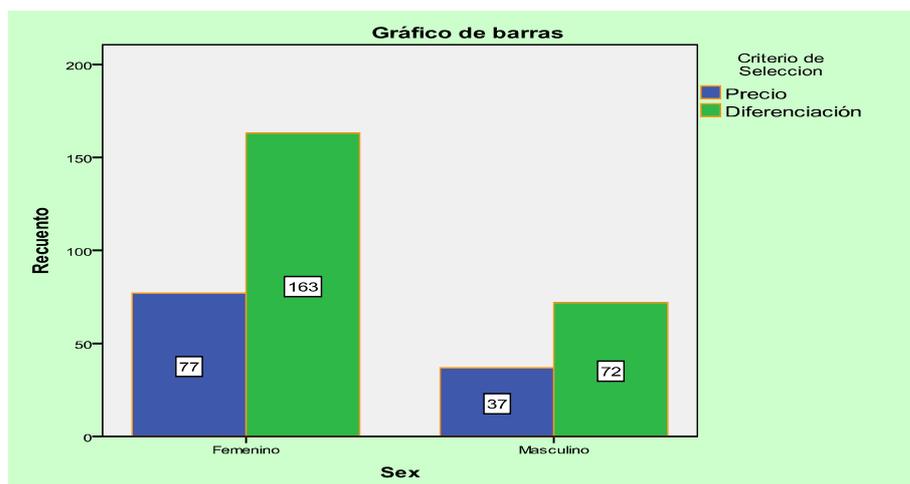
4.1 Cruce de variables categóricas: Criterio de selección (compra) y género del comprador.

Evidentemente como ya se ha señalado por cada 7 compradores del sexo femenino 3 se corresponden con el género masculino y, en cuanto a la preferencia total en el criterio de compra se puede concluir que los porcentajes son parecidos pero en el sentido de que del total de los entrevistados el 67% aproximadamente prefieren el criterio de compra basado en la diferenciación. Sin embargo, si analizamos la información de la tabla 6, dentro de los grupos observaremos que por ejemplo en el caso de las féminas que el 22.1% prefiere el criterio de los precios contra el 46.7% que opta por la diferenciación. En cambio en el caso de los hombres se corresponde a 10.6% y 20.6% respectivamente.

Tabla 6. Tabla de contingencia Sex * Criterio de Selección.

			Criterio de Selección.		Total
			Precio	Diferenciación	
Sex	Femenino	Recuento	77	163	240
		% del total	22.1%	46.7%	68.8%
	Masculino	Recuento	37	72	109
		% del total	10.6%	20.6%	31.2%
Total		Recuento	114	235	349
		% del total	32.7%	67.3%	100.0%

Figura 2. Criterio de selección con relación al género del comprador.



¿Es suficiente la evidencia presentada para aseverar que se relacionan significativamente ambas variables? Pasamos al análisis para probar la hipótesis de no relacionamiento ya comentadas.

Tabla 7. Prueba de significancia de relación Chi- Cuadrado.

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	.118^a	1	.731		
Corrección por continuidad ^b	.049	1	.825		
Razón de verosimilitudes Estadístico exacto de Fisher	.118	1	.732	.806	.411
Asociación lineal por lineal	.118	1	.731		
N de casos válidos	349				

Efectivamente el género no se encuentra significativamente relacionado con el criterio de compra.

Sin embargo, aún cabe la duda razonada de que esa muestra en particular allá arrojado esos resultados producto todavía del azar. Existe una herramienta muy poderosa (*bootstrapping*) para generar en un instante un reemuestreo a través de poderoso software que hace más robustos los resultados y genera intervalos de confianza que nos ayuden a aumentar la fiabilidad de la prueba de hipótesis de significancia asimismo de los coeficientes estimados. Hair, Anderson y Tatham (2001). En lo sucesivo cuando una de las pruebas de significancia estadística se considere central en el desarrollo y análisis de la información, se señalara que se someterán a la estrategia del reemuestreo bootstrap definiendo para ello un nivel de confianza del 95% con muestreo simple con un total de 1,000 muestras.

5.- CrossTab y Prueba de significancia: Edad vs. Criterio de decisión.

Por lo que respecta a la caracterización y estructura de la información relativa a los rangos de edad de los consumidores, podemos destacar que, en la captura de la información se cuenta con un total de 340 sujetos y 11 casos perdidos. Cfr., tabla 8.

Tabla 8. Resumen del Procesamiento de los Datos.

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Edad * Criterio de Selección	340	96.9%	11	3.1%	351	100.0%

De ese total tenemos la presencia de la categoría de edad balanceada en el sentido de que se encuentra constituida por estratos que se describen a continuación: compradores en la categoría de 18-28 años 22.6%, compradores en el rango de 29-39 años de edad 28.5%, la siguiente categoría de respuesta lo constituyen quienes declararon en el rango de 40- 50 años con un 28.8%. Siendo los anteriores categorías las más representativas de la población. En términos marginales se cuenta con los rangos entre 51-61 años con el 13.2% y, quienes declararon tener más de 61 años con el 6.8%. Tengamos en cuenta que más de la mitad corresponden a individuos que declararon contar con una edad entre los 18 y 39 años lo que habla evidentemente de una conformación poblacional todavía joven en la ciudad. Para mayor detalle consulte la tabla 9.

Tabla 9. Tabla de contingencia Edad * Criterio de Selección.

			Criterio de Selección		Total
			Precio	Diferenciación	
Edad	18-28	Recuento	22	55	77
		% del total	6.5%	16.2%	22.6%
	29-39	Recuento	36	61	97
		% del total	10.6%	17.9%	28.5%
	40-50	Recuento	35	63	98
		% del total	10.3%	18.5%	28.8%
	51-61	Recuento	11	34	45
		% del total	3.2%	10.0%	13.2%
	62 En adelante	Recuento	7	16	23
		% del total	2.1%	4.7%	6.8%
Total		Recuento	111	229	340
		% del total	32.6%	67.4%	100.0%

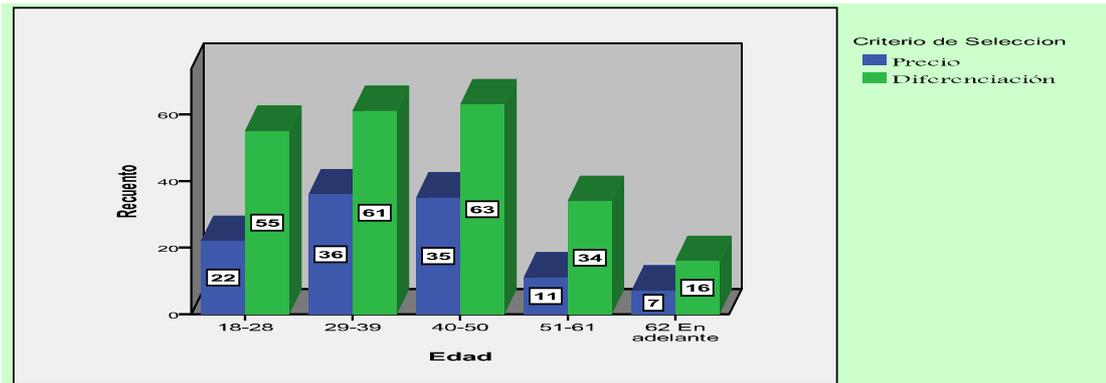
Tome en cuenta que en términos de la muestra, los rangos de edad en donde en mayor medida se opta por el criterio de la diferenciación del producto es el que va de 40-50 años con un 18.5% del total de toda la muestra que prefiere utilizar ese criterio de compra.

¿Están relacionados significativamente las variables edad del comprador y el criterio de selección de compra en la población? Veamos. Como se puede apreciar en la tabla siguiente el valor del estadístico Cui-cuadrado es de 3.309 y al verificar su significancia vemos que es superior al α .05. Por lo que no se rechaza la hipótesis nula. Esto es, no existe suficiente evidencia empírica para suponer que en la población exista una relación entre el criterio de compra utilizado y la edad del comprador.

Tabla 10. Pruebas de Chi-cuadrado.

Criterios de evaluación	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3.309 ^a	4	.507
Razón de verosimilitudes	3.370	4	.498
Asociación lineal por lineal	.074	1	.785
N de casos válidos	340		

Figura 3. Categorías de edades y criterio de compra.



6.- La variable nivel educativo de los compradores.

Pasemos al examen ahora en la forma como se relacionan el nivel educativo de los respondientes con el criterio utilizado en la compra de aceites comestibles. Para tal propósito, la variable de interés se ha categorizado en seis opciones que van desde primaria inconclusa hasta el nivel máximo de posgrado. La exacta observación de la estructura de este cruce de variables la podemos representar por medio de la siguiente tabla.

Tabla 11. Tabla de contingencia Edu * Criterio de Selección.

			Criterio de Selección.		Total
			Precio	Diferenciación	
Edu	Primaria inconclusa	Recuento	9 2.6%	11 3.2%	20 5.8%
	Primaria	Recuento	21 6.1%	18 5.3%	39 11.4%
	Secundaria	Recuento	24 7.0%	31 9.1%	55 16.1%
	Preparatoria	Recuento	33 9.6%	70 20.5%	103 30.1%
	Profesional	Recuento	22 6.4%	91 26.6%	113 33.0%
	Posgrado	Recuento	3 .9%	9 2.6%	12 3.5%
Total		Recuento % del total	112 32.7%	230 67.3%	342 100.0%

De la información anterior queda claro que a niveles superiores de estudio (licenciatura y posgrado), la tendencia favorece claramente al criterio de compra basado en la comparación y valoración de los atributos de los productos. Generalmente, a mayor nivel de estudios mayor ingreso percibido y por supuesto en esos niveles de estudios se busca mayor reconocimiento social. Hay un claro aprendizaje de los esfuerzos por parte de la publicidad por diferenciar y posicionar el producto hacia ciertos segmentos del mercado. Solo pensemos en dos posibles conectores con el nivel educativo de las personas y, por ende en su criterio de compra.

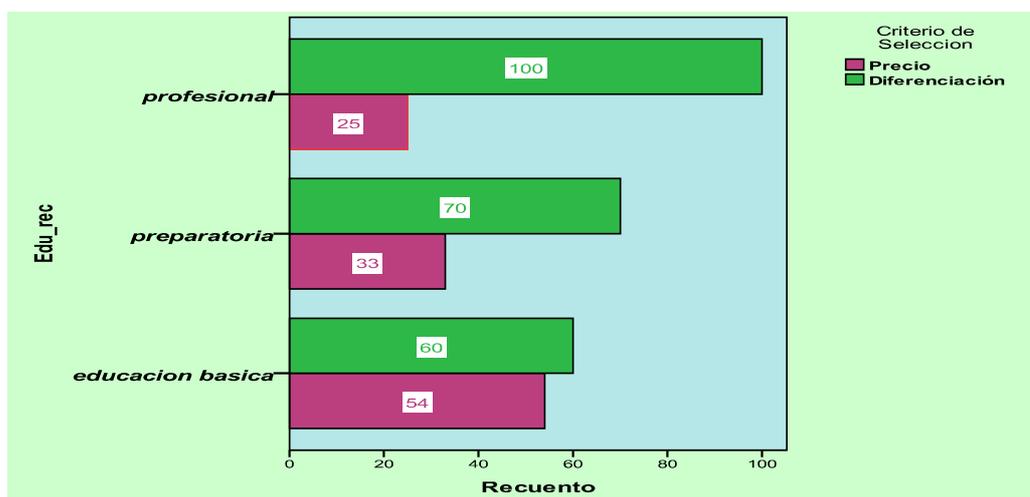
Factores importantes son, primero, el relativo a la discriminación de estímulos relacionado con el aprendizaje del comportamiento del consumidor, definido como la capacidad del individuo para percibir diferencias importantes en los estímulos. Segundo, la clase social siendo esta la misma se comparten actitudes, estilos de vida y comportamiento de compras similares. Kerin, Hartley y Redelius (2009).

Por la argumentación anterior, es al parecer interesante recategorizar las variables de suerte tal que, reconfiguremos un cluster constituido por aquellas personas que solo tienen como nivel de estudios hasta la secundaria y, otro conformado por aquello que van desde el nivel preparatoria hasta posgrado. Para aclarar esta propuesta, consideremos la información siguiente (recategorizada).

Tabla 12. Cross Tab. edu_rec * Criterio de Selección.

			Criterio de Selección		Total
			Precio	Diferenciación	
Edu_rec	educación básica	Recuento	54	60	114
		% del total	15.8%	17.5%	33.3%
	preparatoria	Recuento	33	70	103
		% del total	9.6%	20.5%	30.1%
	profesional	Recuento	25	100	125
		% del total	7.3%	29.2%	36.5%
Total		Recuento	112	230	342
		% del total	32.7%	67.3%	100.0%

Figura 4. Nivel educativo de los respondientes.



Observemos como de seis categorías se han reducido a tres bajo la nueva etiqueta Edu_rec: Educación básica, preparatoria y, por último educación profesional. Con lo que se ha alcanzado un balance representativo en cuanto al tamaño de los nuevos segmentos con fuerte significación social y profesional. Ahora, formalicemos la prueba de significancia a través de plantear la siguiente hipótesis.

H_0 : “La variable nivel educativo de los compradores no tiene relación con el criterio de compra utilizado bien sea este el precio o la diferenciación por atributos.”

Tabla 13. Pruebas de Chi-cuadrado.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	20.312 ^a	2	.000
Razón de verosimilitudes	20.533	2	.000
Asociación lineal por lineal	20.164	1	.000
N de casos válidos	342		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 33.73.

Como se comprueba, el ensayo resulta ser altamente significativo. Por lo que se rechaza la hipótesis nula y con ello procedemos a seleccionar a tan importante variable de segmentación demográfica en la configuración del modelo de regresión logística, ingresando ahora en la modalidad de covariable. ¿Es esta candidata una variable apropiada en el modelo? Pasemos a analizar esta situación.

6.1.- Prueba de Hipótesis con estadístico Wald y procedimiento bootstrap para el nivel educativo.

Antes de probar la hipótesis de significancia individual de la variable nivel educativo, procedemos a la codificación de la misma tal y como se muestra en la tabla 14.

Tabla 14. Codificaciones de variables categóricas.

		Frecuencia	Codificación de parámetros	
			(1)	(2)
edu_rec	educación básica	114	.000	.000
	preparatoria	103	1.000	.000
	profesional	125	.000	1.000

Como se puede deducir de la tabla anterior, la categoría educación básica es la categoría de referencia o comparación. ¿Por qué se selecciono esa categoría como referente?

Por qué lo que queremos indagar es el papel de la educación en el criterio de discriminación y, se parte de suponer que quienes tienen mayor nivel de educación tienden hacia la discriminación por atributos en el criterio de compra. En cambio al nivel de estudios de preparatoria se le asigna la categoría (1). Finalmente a nivel profesional el código (2).

Ahora bien, ¿Es el coeficiente que acompaña a la covariable nivel de estudios igual a cero en la población? Dicho de otra manera, ¿se encuentran relacionadas significativamente las variables dependiente e independiente? Expresemos formalmente la hipótesis nula.

H_0 : “La variable educación no se encuentra asociada en la explicación en la forma en cómo los consumidores deciden utilizar el criterio de compra binario.”

Para contar con mayores elementos de fiabilidad y potenciar la veracidad de los resultados se realiza la prueba de hipótesis utilizando el procedimiento *bootstrapping* con un total de 1,000 muestras. Los resultados se presentan a continuación.

Tabla 15. Especificaciones de Bootstrap.

Método de muestreo	Simple
Número de muestras	1000
Nivel de intervalo de confianza	95.0%
Tipo de intervalo de confianza	Percentil

Tabla 16. Variable y categorías en la ecuación.

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a Educareca			19.473	2	.000	
Educareca(1)	.647	.282	5.241	1	.022	1.909
Educareca(2)	1.281	.292	19.261	1	.000	3.600
Constante	.105	.188	.315	1	.574	1.111

a. Variable(s) introducida(s) en el paso 1: Educareca.

Tabla 17. Bootstrap para Variables en la ecuación.

	B	Bootstrap ^a			Intervalo de confianza al 95%	
		Sesgo	Típ. Error	Sig. (bilateral)	Inferior	Superior
Paso 1 Educareca(1)	.647	.012	.291	.023	.077	1.251
Educareca(2)	1.281	.012	.295	.001	.735	1.903
Constante	.105	.005	.186	.555	-.247	.484

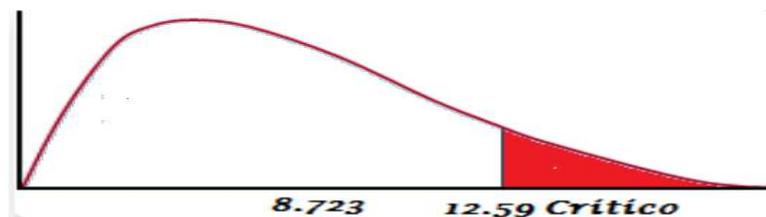
a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

De la tabla anterior se deduce la alta significancia de la variable educación en la discriminación en el criterio de compra del consumidor y alternativamente el procedimiento bootstrap nos permite construir intervalos de confianza al 95% para probar la H_0 de independencia entre la covariable y la variable dependiente. Si observa con atención podrá notar que los coeficientes par las variables Educareca (1) y Educareca (2), caen dentro de los intervalos de confianza respectivos, por lo que de igual manera se rechaza la H_0 llegando a la misma conclusión. Por lo que habremos de retomar más adelante tan significativa covariable en la construcción del modelo de elección discreta.

6.2.- La ocupación del comprador como covariable candidata en el modelo de elección discreta.

Cuando analizamos la variable ocupacional de los compradores entrevistados encontramos que no es una candidata significativa a ingresar al modelo ya que, como se puede apreciar en la tabla 47, el valor calculado del estadístico (8.723) no es significativo dado que el valor teórico o crítico es de 12.592 (*extraído de tablas*), para 6 grados de libertad, es decir $(r-1)(c-1)$. En forma complementaria se muestra el modelo grafico en la prueba de hipótesis en la figura 35.

Figura 5. Modelo grafico de la no significancia de la variable ocupación.

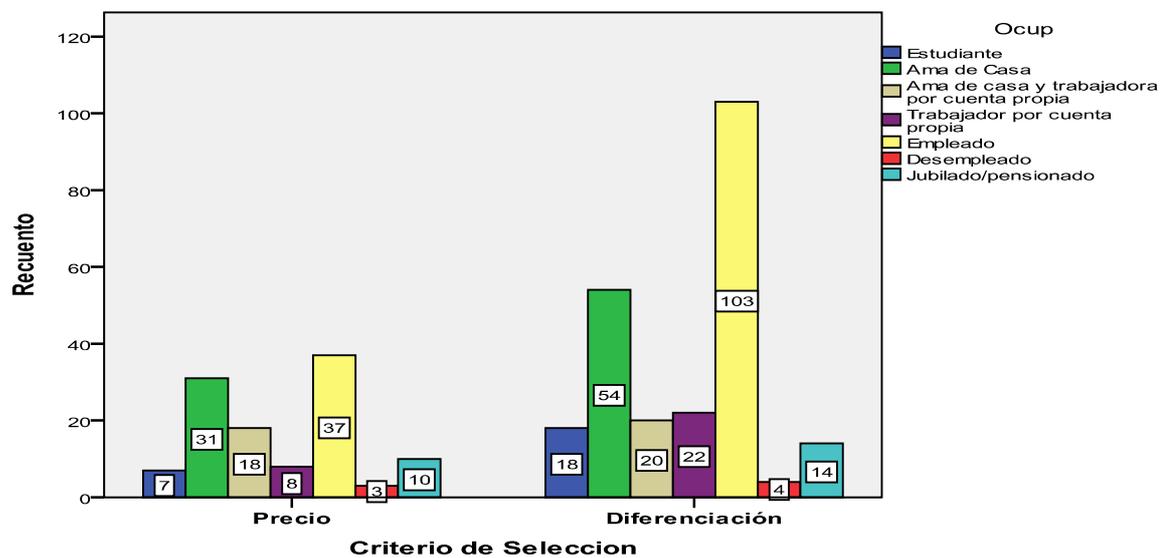


Adicionalmente, hay que hacer notar que como parte de la restricción en el uso de la técnica de tabulación cruzada tenemos 2 casillas con frecuencias esperadas menor a 5, lo cual se constituye en un problema estructural en la prueba. Ciertamente, se pudieran compactar las categorías de respuesta siempre y cuando tuviesen una consideración teórico-práctica lo cual no es el caso, dada la diversidad de categorías de respuesta utilizados en esta variable. Dada la contundencia de la prueba estadística utilizando el procedimiento bootstrap, la variable se descarta.

Tabla 18. Prueba Chi- cuadrado. ¿Se relacionan la ocupación de los compradores y el criterio de compra utilizado?

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	8.723^a	6	.190
Razón de verosimilitudes	8.572	6	.199
Asociación lineal por lineal	.408	1	.523
N de casos válidos	349		

Figura 6. Ocupación y criterio de compra.



6.3.- La variable estado civil de los compradores en el criterio de compra.

Por lo que respecta a la posible correlación entre la variable estado civil de los respondientes y el criterio binario de compra los resultados son:

Tabla 19. Tabla de contingencia Edo. Civil * Criterio de Selección.

			Criterio de Selección		Total
			Precio	Diferenciación	
Edo.Civi	Soltero	Recuento	21	49	70
		% del total	6.2%	14.5%	20.6%
	Casado	Recuento	66	141	207
		% del total	19.5%	41.6%	61.1%
	Divorciado	Recuento	6	14	20
		% del total	1.8%	4.1%	5.9%
	Viudo	Recuento	8	14	22
		% del total	2.4%	4.1%	6.5%
	Unión Libre	Recuento	10	10	20
		% del total	2.9%	2.9%	5.9%
Total		Recuento	111	228	339
		% del total	32.7%	67.3%	100.0%

Aún y cuando se compacten las categorías de respuesta a casados que representan el 61.1% de la muestra y el resto de las categorías los resultados seguirán siendo no significativos.

Por tanto al realizar la corrida correspondiente a la tabla 20, se puede concluir sobre la independencia de las variables. Para apreciar de mejor manera lo poco significativa de esta prueba de hipótesis se complementa con el modelo grafico.

Figura 7. Valor crítico y calculado en la hipótesis nula de independencia de variables.

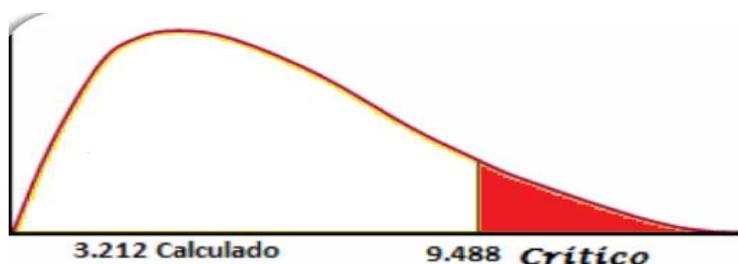


Tabla 20. Prueba de hipótesis sobre independencia: estado civil vs Criterio de compra.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3.212^a	4	.523
Razón de verosimilitudes	3.047	4	.550
Asociación lineal por lineal	2.299	1	.129
N de casos válidos	339		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5.

6.4. - Prueba de hipótesis de independencia entre las variables nivel de ingreso y criterio de compra.

Para analizar la significancia o no de esta variable estratégica desde la teoría económica y mercadológica dentro del bloque de segmentación demográfica, procedemos a la utilización de la técnica bootstrap.

Tabla 20. Tabla de contingencia Ing * Criterio de Selección.

			Criterio de Selección		Total
			Precio	Diferenciación	
Ing	Hasta 2 salarios mínimos	Recuento	45	46	91
		% del total	13.0%	13.3%	26.4%
	de 3 a 6 salarios mínimos	Recuento	53	119	172
		% del total	15.4%	34.5%	49.9%
	De 7 a 10 salarios mínimos	Recuento	14	44	58
		% del total	4.1%	12.8%	16.8%
	Más de 10 salarios mínimos	Recuento	2	22	24
		% del total	.6%	6.4%	7.0%
Total		Recuento	114	231	345
		% del total	33.0%	67.0%	100.0%

Como podemos apreciar en la tabla siguiente la prueba estadística resulta altamente significativa con un valor calculado Chi-cuadrado de Pearson de 20.161 con un valor de probabilidad bajísimo (.000).

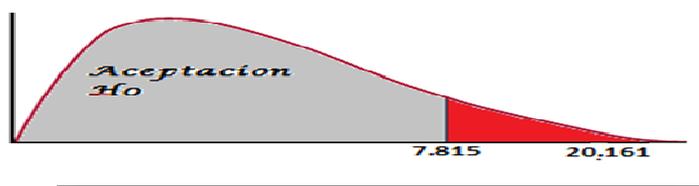
Tabla 21. Pruebas de Chi-cuadrado.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	20.161^a	3	.000
Razón de verosimilitudes	21.321	3	.000
Asociación lineal por lineal	18.835	1	.000
N de casos válidos	345		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5

Para mayor información se incorpora en la figura 8, el valor teórico para 3 grados de libertad y nivel de significancia $\alpha = 0.05$.

Figura 8. Modelo grafico de rechazo de H_0 .



En suma, contamos ahora con nuestra segunda variable dentro del bloque de segmentación demográfica a incorporar al modelo propuesto. Por lo que ahora le conoceremos en el ambiente de la regresión logística como covariable (explicativa).

6.5.- Variables métricas en la segmentación demográfica: Número de personas que contribuyen al gasto familiar y número de habitantes por hogar.

Una variable importante a considerar de acuerdo a la aportación teórica es sin duda el número de personas que contribuyen con el gasto familiar en el hogar. Hay que recordar que dicha variable tiene en principio un nivel de medición de tipo métrico. También es pertinente aclarar que del total de los 351 encuestados por lo referente a esta medición se cuenta con la ausencia de respuesta de tan solo 10 casos por lo que el tamaño de la muestra sigue siendo representativo. En la tabla 22 se presenta a detalle la información.

Tabla 22. Resumen del procesamiento de los casos.

Casos no ponderados ^a		N	Porcentaje
Casos seleccionados	Incluidos en el análisis	341	97.2
	Casos perdidos	10	2.8
	Total	351	100.0
Casos no seleccionados		0	.0
Total		351	100.0

En cuanto a la importancia relativa de esta variable métrica encontramos que efectivamente ésta no es significativa a un nivel de significancia del 5%. Por tanto, se procede a no rechazar la hipótesis nula y, con ello concluimos que midiendo en forma métrica esta variable se descarta para posterior análisis. Note en el círculo en rojo el valor asociado al estadístico Wald (.524) y su correspondiente significancia (.469).

Tabla 23. Variables en la ecuación.

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a No.Personas_Contribuyen	.081	.112	.524	1	.469	1.085
Constante	.564	.236	5.708	1	.017	1.758

a. Variable(s) introducida(s) en el paso 1: No.Personas_Contribuyen.

Aún y cuando se intenta recategorizar para formar cluster más consistentes tenemos la misma decisión de rechazar la H_0 . A continuación se presentan los detalles.

Tabla 24. Tabla de contingencia Contrica * Criterio de Selección.

		Criterio de Selección.			
		Precio	Diferenciación	Total	
Contrica	una persona	Recuento	49	92	141
		% del total	14.0%	26.4%	40.4%
	dos personas	Recuento	48	89	137
		% del total	13.8%	25.5%	39.3%
	más de dos personas	Recuento	17	54	71
		% del total	4.9%	15.5%	20.3%
Total		Recuento	114	235	349
		% del total	32.7%	67.3%	100.0%

Solo hay que puntualizar sobre el hecho de la recategorización en la nueva variable *Contrica* que queda reducida a tres categorías consistentes en; una persona contribuye con los gastos del hogar, dos personas y la última categoría más de dos personas.

Tabla 25. Pruebas de Chi-cuadrado.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3.085^a	2	.214
Razón de verosimilitudes	3.216	2	.200
Asociación lineal por lineal	1.911	1	.167

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5.

Por lo que respecta a la variable número de habitantes por hogar se tiene que al realizar la corrida correspondiente a la estimación del modelo de regresión logística simple encontramos los siguientes resultados.

Tabla 26. Resumen del procesamiento de los casos.

Casos no ponderados ^a	N	Porcentaje
Casos seleccionados	340	96.9
Incluidos en el análisis		
Casos perdidos	11	3.1
Total	351	100.0

a. Si está activada la ponderación, consulte la tabla de clasificación para ver el número total de casos.

Para un total de 340 casos validos computados los resultados de significancia individual medidos a través del estadístico de Wald son igualmente no relevantes como se muestra en la tabla siguiente.

Tabla 27. Variables en la ecuación.

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a No.Habitantes_Por_hogar	-.045	.078	.329	1	.566	.956
Constante	.881	.342	6.650	1	.010	2.414

a. Variable(s) introducida(s) en el paso 1: No.Habitantes_Por_hogar.

Si intentamos recategorizar y reducir el número de segmentos, los resultados no cambian e incluso aún ni reduciendo a dos categorías de respuesta, siendo estas de 1 a 4 habitantes y más de 4 habitantes. Por tanto, el número de habitantes por hogar no es factor de discriminación contrario a lo que el “sentido común” pudiese indicar.

Referencias Bibliográficas.

- Álvarez, R. (1995). *Estadística Multivariante y no Paramétrica con SPSS. Aplicación a las Ciencias de la Salud*. España: Díaz de Santos.
- Ceniceros, J. (2001). *Modelo de Pronóstico de Exportación de Calabaza Kabocha al Mercado Japonés*. Tesis de maestría no publicada, Universidad de Occidente, Culiacán, Sinaloa, México.
- Gujarati, D. (2010). *Econometría*. (5ra Edición). Mexico: Mc Graw Hill.
- Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis Multivariante*. (5^{ta} Edición). España: Prentice Hall Iberia.
- Hosmer, D. y Lemeshow, S. (1989). *Applied Logistic Regression*. United States of America: Wiley Interscience Publication.
- Kerin, R., Berkowitz, E., Hartley, S. y Redelius W. (2009). *Marketing*. México: Mc Graw Hill.
- Kerlinger, F. (2001). *Enfoque Conceptual de la Investigación del Comportamiento*. México: Interamericana.
- Lind, D., Marchal, W., y Wathen, S. (2005). *Estadística Aplicada a los Negocios y a la Economía*. (12 edición). México: Mc Graw Hill.
- Pindyck, R. y Rubinfeld, D. (2001). *Econometría: Modelos y Pronósticos*. (4ta Edición). México: Mac Graw Hill.