



ISSN: 1886-8452

Vol.6 Número 16, junio 2014

<http://www.eumed.net/rev/tecsistecat1/index.htm>

¿TIENDEN LAS PERSONAS CASADAS A DISCRIMINAR PREFERENTEMENTE A FAVOR DEL CRITERIO DE LOS PRECIOS EN LA COMPRA DE PRODUCTOS BÁSICOS? EVIDENCIA EMPÍRICA PARA LA COMPRA DE ACEITES COMESTIBLES EN LA CIUDAD DE CULIACÁN, SINALOA, MÉXICO.

Julio César Ceniceros Angulo.

Resumen.

En el presente trabajo se trata de probar la relación existente entre el estado civil manifiesto del comprador de aceites comestibles, y el criterio de compra expresada bajo la dicotomía: Precio del producto vs. Diferenciación del mismo. Investigación realizada en la ciudad de Culiacán, Sinaloa México durante el Verano de 2011 en la etapa de levantamiento de la información y, finales de 2012 en su etapa de análisis, a través de una muestra de 349 compradores; en donde se presentan los resultados, validados a través de Análisis de Regresión Logística Binaria, potencializando la prueba con Bootstrapping; en forma concluyente se tiene que, no existe una relación significativa a un nivel $\alpha=0.05$, entre la variable demográfica de segmentación basada en el estado civil del comprador con relación al criterio de compra binario del consumidor. Por esta

razón es de gran trascendencia para el especialista en mercadotecnia atender de manera especial los aspectos relacionados con la diferenciación de productos.

PALABRAS CLAVES: Diferenciación de Productos, Variables de Segmentación, Análisis de Regresión Logística Binaria.

Summary.

This current work aims to prove the existing relation between the declared marital status of the buyer of cooking oils, and the purchasing criterion expressed in the dichotomy: Product price vs. its differentiation. Research conducted in the city of Culiacan, Sinaloa, Mexico during the summer of 2011 for the survey stage, and the end of 2012 during its analysis phase, through a sample of 349 shoppers; the results are presented, validated through a binary logistic regression analysis. Conclusively, it follows that there is no significant relation at a level $\alpha=0.05$ between the demographic variable of segmentation based on the marital status of the buyer with relation to the binary shopping criterion of the consumer. For this reason, it is of great significance for the specialist in marketing to pay special attention to the aspects related to product differentiation

Keywords.

Product differentiation, segmentation variables, binary logistic regression analysis.

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 Administracion, 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 decisorio 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 enfatizado 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 múltiple (RLM), 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 encuestadas en esos sitios:

Cuadro 1. Negociaciones y número de encuestas aplicadas.

	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 fé	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 método de entrevista personal de intercepción en centros comerciales,

seleccionando en forma aleatoria a cada $k = 3$ sujetos haciéndole la pregunta clave, sobre cuales

es el principal criterio utilizado en la compra de aceites comestibles. Siendo las categorías de respuestas:

1.- Precio.

2.- Características del producto.

Cuadro 2. Resultados de la prueba piloto realizada en el 2008.

Hipermercado	Precio	Características del producto	Total
MZ Centro	5	7	12
Ley Centro	5	5	10
MZ de las Américas	5	10	15
Soriana Universitari	9	6	15
Wal-Mart 68	4	11	15
Sams Montebello	6	9	15
TOTAL	34	48	82
Proporción	.41	.59	1

Fuente: Elaboración propia.

Finalmente, para realizar la encuesta se procederá a determinar el tamaño de muestra de la siguiente manera:

Niveles de error (ϵ) y confianza en la determinación del tamaño de muestra probabilística con proporciones (n_p) cuando se desconoce la población (N).

Cuadro 3. Determinación del tamaño muestral.

p	$q = 1 - p$	ϵ	Nivel de confianza	n_p^*
.41	.59	5%	95%	371
.41	.59	5.5%	95%	307
.41	.59	6%	95%	258
.41	.59	5%	94.12%	346
.41	.59	5.5%	94.12%	286

Fuente: Elaboración propia.

$$* n_p \frac{Z^2(p)(q)}{\epsilon^2}$$

Tratándose de estimación de proporciones entre los que deciden la compra preponderantemente a través de la variable precio y los que se deciden por las características (atributos) del producto, se empleará un 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_n , *parámetros o coeficientes de la ecuación.*
- 3.) x_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 k y las x_n Hair., *et. al.*, (1999), nos presentan la siguiente representación sigmoide:

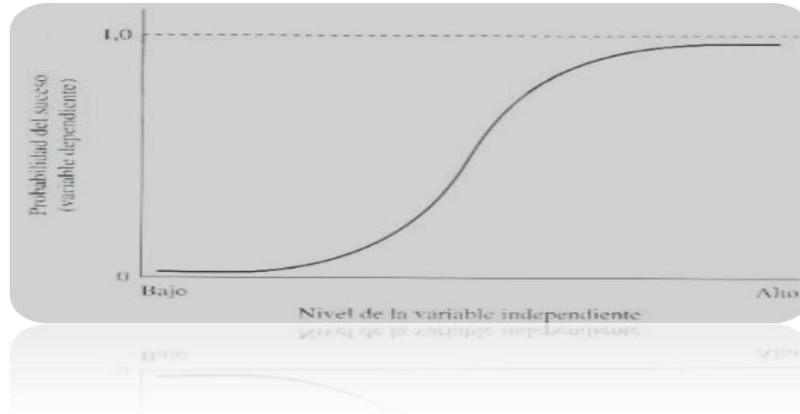


Figura 1. Representación sigmoidea de la función logística.

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:

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

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

Para Figueroa (2009), los problemas a abordar del modelo precedente quedan resueltos, si p_i toma

valores de 0 y 1 de la siguiente manera:

$$\text{Si, } z \rightarrow \infty \Rightarrow p_i = \frac{1}{1 + \frac{1}{e^z}} = \frac{1}{1 + \frac{1}{e^\infty}} = \frac{1}{1+0} = 1$$

$$\text{De la misma forma, Si, } z \rightarrow -\infty \Rightarrow p_i = \frac{1}{1+e^\infty} = 0$$

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:

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

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,

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

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:

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

Autores principales en la Regresión Logística son Hosmer y Lemeshow (1989), que en su trabajo clásico Applied Logistic Regression, 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:

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

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

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

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:

$$(7) \quad \left(\frac{\delta}{\lambda} \right) = e^{\beta_0} (e^{\beta_1})^{x_1} \dots (e^{\beta_p})^{x_p}$$

Segundo, con relación a ω_η , 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

$\beta_2 x_1$, 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:

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

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:

Cuadro 4. Signos e interpretación de los odds ratio.

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

Complementando lo anterior, de acuerdo a Álvarez (1995), con respecto al significado de los coeficientes y signos 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).

Tercero, en lo que se refiere a las variables independientes o explicativas en el modelo (RLS), un

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

modelo simple solo incluye una variable explicativa, pudiendo representarse de la siguiente manera:

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

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 ser codificadas en forma numérica con la regla conocida de que si se tienen K , número 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? Es necesario comprobar entonces 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 siguiente ecuación: (10) $p_i = \frac{1}{1 + e^{-z_i}}$, el exponente puede tener incluidos varias

χ^2 , por ejemplo si:

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

Siguiendo el razonamiento anterior, 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:

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

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

Cuadro 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.- Resultados y Discusión.

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.

La exacta observación del estado civil de las personas puede llevarnos a cuestionar por ejemplo. ¿Es el consumidor soltero quien con mayor frecuencia tiende a seleccionar el criterio de diferenciación o atributos del producto en la adquisición de aceites comestibles o en general de artículos básicos? ¿Tienden las personas casadas a optar por los precios de los productos como criterio de discriminación en la compra de este tipo de productos? Interesante conocer las respuestas a estas interrogantes por ello debemos darle especial énfasis a esta variable de segmentación del consumidor. Dada la fuerte preeminencia de las personas con el status declarado como casados, es conveniente intentar reconvertir esta variable a dicotómica. Pudiendo quedar de la forma siguiente: Casados (207 personas) que representan el 61.1% de la muestra total y no casado (132 sujetos), que se constituyen en el 38.9% de la muestra. Para mayor detalle de la forma a en cómo se estructura la variable estado civil de los respondientes, *vid., Infra.*, cuadro 6 y figura 2.

Cuadro 6. Estado civil de las personas encuestadas en los supermercados.

Estado Civil					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Soltero	70	20.1	20.6	20.6
	Casado	207	59.3	61.1	81.7
	Divorciado	20	5.7	5.9	87.6
	Viudo	22	6.3	6.5	94.1
	Unión Libre	20	5.7	5.9	100.0
	Total	339	97.1	100.0	
Perdidos	Sistema	10	2.9		
Total		349	100.0		

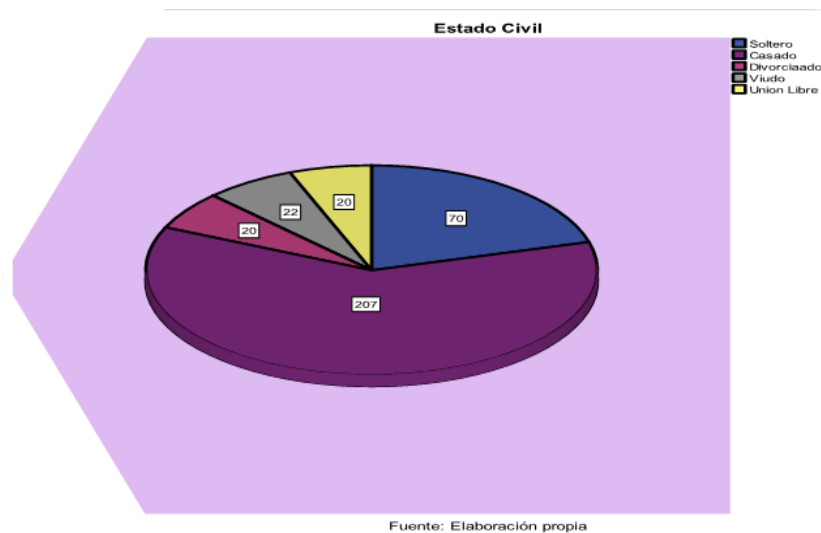


Figura 2. Estado civil de las personas encuestadas en los supermercados.

4.1. 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:

Cuadro 7. 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 % del total	10 2.9%	10 2.9%	20 5.9%
Total		Recuento % del total	111 32.7%	228 67.3%	339 100.0%

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 al cuadro 8, 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 gráfico. Bowerman, O'Connell & Murphree (2009 : 878)

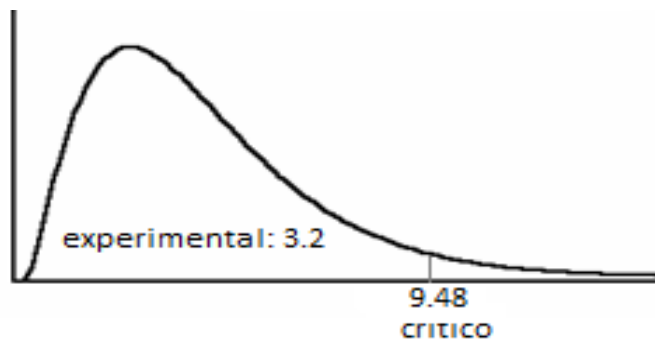


Figura 3. Valor crítico y calculado en la hipótesis nula de independencia de variables.

Cuadro 8. Prueba de hipótesis sobre independencia: estado civil
Criterio de compra.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3.212^a	4	.523 NS
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.

Sin embargo, lo que realmente interesa en el propósito central de la investigación es comparar si el estado civil del respondiente (casado), tienden a discriminar en favor del criterio de los precios contra cualquier otra categoría independientemente de lo manifestado en términos de reconocimiento de su estatus legal o no. Por lo anterior, es que se procede a recategorizar esta variables y llevarla solo a nivel dicotómico. Por tanto, los resultados ahora se presentan bajo la etiqueta de Estaci_re.

Cuadro 9. Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Criterio de Selección * Estaci_re	349	99.4%	2	.6%	351	100.0%

Cuadro 10. Tabla de contingencia Criterio de Seleccion * Estaci_re

			Estaci_re		Total
			otro	casado	
Criterio de elección	Precio	Recuento	48	66	114
		% del total	13.8%	18.9%	32.7%
	Diferenciación	Recuento	94	141	235
		% del total	26.9%	40.4%	67.3%
Total		Recuento	142	207	349
		% del total	40.7%	59.3%	100.0%

Se trata por tanto de probar ante esta nueva configuración la siguiente hipótesis nula.

H₀: “Existe independencia entre las variables criterio de compra binario y el estado civil manifiesto del respondiente y por tanto, pueden ser incluidas como covariables en el modelo de regresión logística”.

Como puede deducirse fácilmente con la información del cuadro 11 no puede rechazarse la H₀. Por tanto, el estado civil de las personas no puede ser considerada como una covariable en el modelo predictivo. Esto es, no existe suficiente evidencia empírica que permite afirmar que quienes han declarado estar casados opten por el criterio de los precios. Para mayores detalles, Cfr., cuadro 11. Y figura 4.

Cuadro 11. Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	.141 ^a	1	.707 NS		
Corrección por continuidad ^b	.067	1	.795		
Razón de verosimilitudes	.141	1	.708		
Estadístico exacto de Fisher				.728	.397
Asociación lineal por lineal	.141	1	.708		
N de casos válidos	349				

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

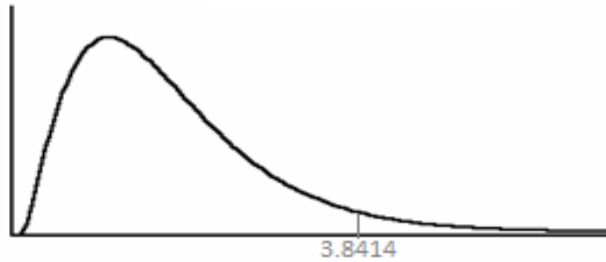


Figura 4. Prueba de hipótesis a través de expresión grafica.

Los resultados son consistentes y robustos una vez que se realiza la corrida para validar con el modelo de regresión logística binaria como se demuestra a continuación. No podemos omitir aquí que la incorporación de la variable de interés en el modelo no contribuye en forma significativa a su capacidad predictiva, como se puede observar en el cuadro 12.

Cuadro 12. Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	.141	1	.708
	Bloque	.141	1	.708
	Modelo	.141	1	.708 NS

Finalmente, se incluyen las variables en la ecuación de regresión logística simple y, se concluye que la variable Estaci_re con un valor Wald de .141 no es significativa. Vid., cuadro 13.

Cuadro 13. Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	Estaci_re	.087	.232	.141	1	.707 NS	1.091
	Constante	.672	.177	14.353	1	.000	1.958

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

Literatura citada.

Kerlinger, F. y Lee, H. (2001). *Investigación del Comportamiento*. México: Mc Graw Hill.

Ceniceros, J. (2001). *Modelo de Pronostico de Exportación de Calabaza Kabocha al Mercado Japonés*. Tesis de maestría no publicada, Universidad de Occidente, Culiacán, Sinaloa, México.

Gujarati, D. (2010). *Econometría*. (Quinta Edición). México: Mc Graw Hill.

Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis Multivariante*. (5^{ta} Edición). España: Prentice Hall Iberia.

Pindyck, R. y Rubinfeld, D. (2001). *Econometría: Modelos y Pronósticos*. (4ta Edición). M
Hosmer, D. y Lemeshow, S. (1989). *Applied Logistic Regression*. United States of America: Wiley
Interscience Publication.

Álvarez, R. (1995). *Estadística Multivariante y no Paramétrica con SPSS. Aplicación a las Ciencias de la Salud*. España: Díaz de Santos.

Lind, D., Marchal, W. y Wathen, S. (2005). *Estadística Aplicada a los Negocios y la Economía*. México: Mc Graw Hill.