UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Facultad de Estadística
Trabajo de Grado
Diciembre 2015

Estimación de las preferencias del consumidor a través de un modelo jerárquico bayesiano logístico multinomial

Estimation consumer's preferences using a Bayesian hierarchical multinomial logit model

José Melo Fuquene.[a]
josemelo@usantotomas.edu.co

José Babativa Marquez[b], Dagoberto Bermúdez Rubio[c]
[b−c] Directores Trabajo de Grado

Resumen

El uso de las técnicas estadísticas ha cobrado gran importancia en las investigaciones de mercados, es usual que se diseñen estudios donde el objetivo esté orientado a identificar los factores que afectan las decisiones de compra de los consumidores. Existen diversas metodologías en ese sentido, entre las más conocidas están: evaluaciones comparativas, análisis conjunto, maxdiff y trade off. Sin embargo las más simples como la evaluación comparativa carece de rigor estadístico y otras como el análisis conjunto suelen ser dispendiosas de llevar a cabo especialmente durante el proceso de levantamiento de información, debido a que en ocasiones a pesar de trabajar con diseños ortogonales el número de ofertas que debe evaluar cada entrevistado sigue siendo tan alto que podría ocasionar fatiga y por ende sesgo en los resultados.

En este trabajo se propone una metodología de tipo trade off, la misma consiste en empezar con dos ofertas al azar e identificar cuál de las dos es más atractiva para el consumidor, posteriormente la oferta ganadora se compara con otra oferta y también se identifica la más atractiva, la cual a su vez se compara con otra oferta que no haya sido comparada y así sucesivamente hasta terminar. Si el número de ofertas es k en total, el encuestado tendrá que comparar $k-1$ pares de ofertas, de esta forma se logra reducir los tiempos de la encuesta evitando los sesgos ocasionados por el cansancio, además de disminuir los costos del proyecto de forma importante.

Con el fin de estimar la probabilidad de que el sujeto i $(i = 1, ..., n)$ elija la oferta j $(j = 1, ..., k)$ se propone realizar un Modelo Jerárquico Bayesiano Logístico Multinomial usando métodos MCMC (Markov Chain Monte Carlo), de esta forma se puede obtener las probabilidades de que cada individuo seleccione una oferta sin la necesidad de comparar todos los pares. Una ventaja de usar el enfoque bayesiano es debido a la gran cantidad de parámetros, que en ocasiones resulta complejo o imposible hacerlo de forma clásica; además con el fin de mejorar las estimaciones de los parámetros se involucra en la distribución de los mismos la información sociodemográfica de los sujetos. El ajuste del modelo considera un proceso Dirichlet, apropiada para distribuciones multinomiales.

Una vez obtenidas las probabilidades para cada sujeto en cada oferta, se aplica un análisis de componentes principales y usando las primeras componentes se realiza una clasificación de los de los sujetos con el fin de identificar los nichos del mercado para las diferentes ofertas.

[a]Estudiante de Pregrado en Estadística, U. Santo Tomas

Palabras clave: *Análisis Conjunto, Trade Off, Modelo Jerárquico, Modelo Logistico Multinomial, Estimación Bayesiana, Análisis de componentes principales, Métodos de clasificación*

Abstract

The statistics techniques have won importance in market research, this techniques usually use when the objective wants to know the consumer's factors that affects decisions of consumer. Many methodologies are used how conjoint analysi, maxdiff and trade off. But the simples techniques don't have optimal results and other techniques have a long duration, for this reason this techniques can give a bad results.

In that job, we propose a different trade off methodology, that methodology compares two supplies and the subject selects one, again the last selection compares with other supply and finally the subject compares all supplies. If the numbers deals are k, the subject have to compare $k-1$ deals, for that reason that methodology can reduce the times and the costs.

For estimate the probability that the subjet $i(i = 1, ..., n)$ select the supply $j(j = 1, ..., k$, we propose a Bayesian hierarchical multinomial logit model using MCMC (Markov Chain Monte Carlo), with that model we can estimate the probability that the subject i select each supply when the subject couldn't compare all supplies.

When we are found the probabilities of each subject, we can apply a principal component analysis and in this manner we can do a classification of the subjects where finally we can identify markets niches for the differents supplies.

Keywords: *Conjoint analysis, Trade Off, Hierarchical Model, Multinomial logistic model, Bayesian estimation, Principal component analysis, classification analysis methods*

1. Introducción

En investigación de mercados, en la mayoría de los casos es fundamental encontrar el enlace entre consumidor y producto, por esta razón es necesario obtener resultados verídicos y precisos, estos resultados afectan en la toma de decisiones acerca del objetivo de la investigación, por esto se debe recolectar toda la información posible y diseñar un método de recolección el cual debe realizarse mediante un diseño científico y así de esta forma se está cumpliendo con las condiciones necesarias para encontrar la relación entre consumidor y producto, como resultado final se podrá tomar las decisiones necesarias del mercado (establecimiento de precios, promociones, distribuciones, etc.).

Las diversas herramientas estadísticas proporcionan de una forma ordenada y clara herramientas que dan solución en encontrar la relación entre consumidor y producto, optimizando las estrategias de marketing y midiendo el comportamiento real del consumidor frente al producto que se quiere medir.

En este trabajo se quiere estimar las preferencias de los consumidores cuando se le dan a elegir diferentes ofertas, existen numerosas metodologías para medir la importancia de los ofertas, poder calcular el valor relativo de cada oferta y así determinar cual tendrá mayor nivel de ventas, las metodologías más comunes son un análisis conjunto, trade off y un maxdiff.

El análisis conjunto consiste en medir el valor de cada oferta y encontrar la combinación necesaria de estos ofertas para poder maximizar la probabilidad de elección del consumidor,uno de los análisis conjuntos que se pueden realizar es el choice based conjoint, el cual consiste en que el investigador organiza unos perfiles que contiene un conjunto de atributos de forma sistemática, donde el individuo es el encargado de seleccionar el perfil que considera mejor dependiendo de los conjuntos de atributos que considere mas atractivos (Raghavarao et al. 2011) , esta metodología contiene un diseño ortogonal donde el número de opciones que debe evaluar cada sujeto puede ser demasiado extenso y ocasionar fatiga generando que los resultados no sean verídicos y otro factor determinante es que los costos para realizar esta metodología son muy altos.

Por el contrario el maxdiff es un método donde el sistema de recolección de la información es demasiado rápido y es más económico, este procedimiento consiste en donde se contiene una lista de ofertas y solo se selecciona cual es la mejor y cuál es la peor (Bartomowicz & Bak 2014), el problema radica a medida que el numero de ofertas aumenta puede generar de igual forma un desgaste en el momento de seleccionar la mejor y la peor oferta.

Por ultimo el trade off es el método muy usado donde se compara todas las combinaciones de las ofertas a evaluar, se empieza con la comparación de dos ofertas al azar y el sujeto identifica cuál de las dos es la más atractivo, después procede a comparar otras dos ofertas diferentes a las primeras y selecciona la más atractiva, así sucesivamente hasta realizar las comparaciones faltantes (Wyner 1995), este procedimiento realiza $\binom{k}{2}$ comparaciones en total, esta metodología también puede ser muy extensa cuando el número de comparaciones que tiene que realizar el individuo es muy grande.

En este trabajo se propone realizar una variación en la metodología de trade off, donde la oferta que el sujeto selecciono como la más atractiva, se compara con otra oferta y la oferta ganadora en esta nueva comparación se vuelve a comparar con otra oferta y así sucesivamente hasta acabar con el total de ofertas que se van a evaluar,donde el total de comparaciones que el individuo tiene que realizar es $k-1$. De esta forma se está logrando una mayor cantidad de información , se está reduciendo los tiempos en el levantamiento de la información donde no se estaría generando sesgos por la fatiga de la metodología y al mismo tiempo se está reduciendo costos.

La herramienta estadística que se propone es un Modelo Jerárquico Bayesiano Logístico Multinomial usando métodos MCMC (Markov Chain Monte Carlo) con el fin de poder estimar la probabilidad de que el sujeto $i(i = 1, ..., n)$ seleccione la oferta $j(j = 1, ..., k)$, con este modelo se estará encontrando las probabilidades de que cada sujeto seleccione una oferta sin la necesidad de haber realizado todas las comparaciones. Cada sujeto tendrá un vector de parámetros, donde los parámetros de los sujetos dependen de otro modelo donde se está involucrando la información sociodemográfica del sujeto; por esta razón la estimación del vector de parámetros de cada sujeto se realiza por medio de métodos bayesianos con el fin de encontrar la distribución a posteriori del parámetro ya que por métodos clásicos puede resultar muy complejo por la gran cantidad de parámetros que se quieren estimar.

Finalmente ya con las probabilidades calculadas, se procede a realizar un análisis de componentes principales para determinar cuántos factores se deben retener y así proceder a utilizar un método de clasificación para segmentar a los sujetos por sus variables sociodemográficas con sus respectivas ofertas y así se está identificando los nicho de mercado para las diferentes ofertas.

2. Marco Teórico

2.1. Inferencia Bayesiana

Alguna de las finalidades de la estadística bayesiana es poder llegar a encontrar una estimación inicial para θ (donde θ es un vector parametrico), este valor se encuentra asociado a una función de probabilidad (caso contrario en la estadística frecuentista, donde las observaciones iniciales están asociadas a una función de probabilidad). Esta función de probabilidad está condicionada a unos valores iniciales de una variable y, donde $y_1, ..., y_n$ es una muestra aleatoria (Gelman et al. 2004).

El teorema de Bayes empieza con una función de probabilidad conjunta para θ y Y. Esta función puede ser escrita de la siguiente manera:

$$p(\theta, y) = p(\theta)p(y/\theta),$$

donde $p(\theta)$ es la función de probabilidad previa, esta probabilidad puede estar condicionada a unos hiperparámetros $(p(\theta/\eta))$ y $p(y/\theta)$ es la distribución de los datos (función de verosimilitud). Utilizando

la propiedad básica de probabilidad condicional (teorema de Bayes) se puede encontrar la función de probabilidad a posteriori.

$$p(\theta/y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y/\theta)}{p(y)},$$

donde $p(y)$ en el caso discreto es:

$$p(y) = p(\theta_1, y) + p(\theta_2, y) + p(\theta_3, y) + ... + p(\theta_k, y),$$
$$p(y) = p(y/\theta_1) * p(\theta_1) + p(y/\theta_2) * p(\theta_2) + p(y/\theta_3) * p(\theta_3) + ... + p(y/\theta_k) * p(\theta_k),$$
$$p(y) = \sum_{i=1}^{k} p(y/\theta_i)p(\theta_i),$$

y en el caso continuo de θ:

$$p(y) = \int p(\theta, y)d\theta = \int p(y/\theta_i)p(\theta)d\theta,$$

$p(y)$ es considerada una constante ya que no depende de θ, es un valor fijo, entonces la distribución a posteriori seria igual a:

$$p(\theta/y) \quad \alpha \quad p(\theta)p(y/\theta).$$

Donde la distribución a posteriori es proporcional a la información a priori multiplicada por la función de verosimilitud; esta forma de escribir el teorema de Bayes nos sera útil si las estimaciones de nuestros parámetros presentan problemas. Si se desea recuperar la constante se debe realizar la siguiente integral (Carioca 2011):

$$k^{-1} = \int p(y/\theta_i)p(\theta)d\theta.$$

2.1.1. Distribuciones a priori

Las distribuciones a priori del parámetro de interés θ se pueden especificar de diferentes maneras. Una forma es cuando el parámetro de interés es discreto, se le asigna para cada posible valor del parámetro θ una probabilidad y si el parámetro es continuo se asignan unos valores a los percentiles de la distribución (esta metodología es muy subjetiva). La otra manera es asociar θ a una familia de distribuciones paramétricas (Distribución normal, exponencial, Bernoulli, etc), estas distribuciones paramétricas dependen de unos hiperparámetros (μ, σ^2 en el caso de una distribución normal) (Migon et al. 2008).

2.1.2. A priori no informativa

Muchas veces no se puede encontrar información a priori de θ, lo que se sugiere en atribuirle una distribución a priori no informativa que tenga un poco de influencia en la distribución a posteriori. Una de las distribuciones a priori no informativas mas usuales es la propuesta de Jeffreys (Gelman et al. 2004). Esta distribución es proporcional a la información de Fisher.

$$p(\theta) \quad \alpha \quad (I(\theta))^{\frac{1}{2}},$$

$$p(\theta) \ \alpha \ E\left[\left(\frac{d}{d\theta} \ Log \ p(y/\theta)\right)^2\right],$$

$$p(\theta) \ \alpha \ -E\left(\frac{d^2}{d\theta^2} \ Log \ p(y/\theta)\right).$$

2.1.3. Distribuciones conjugadas

Según Migon et al. (2008), se asigna P a una clase de distribución a priori, esta distribución P puede ser conjugada de una familia de distribuciones de probabilidad F.

$$F = p(y/\theta).$$

Donde el valor del parámetro θ pertenece a un espacio paramétrico Θ. Si para toda función de verosimilitud $p(y/\theta)$ pertenece a una familia de distribuciones de probabilidad F y para toda distribución a priori $p(\theta)$ pertenece a una clase de distribuciones a priori P, se dice que la distribución a posteriori $p(\theta/y)$ pertenece a una clase de distribuciones a priori P.

En el caso de la distribución Bernoulli, se dice que una familia de distribuciones beta es conjugada de una distribución Bernoulli, si $p(y/\theta) \sim Bernoulli(\theta)$ y $p(\theta) \sim Beta(\alpha, \beta)$ se concluye que $p(\theta/y) \sim Beta(\alpha + \sum_{i=1}^{n} y, n - \sum_{i=1}^{n} y + \beta)$

2.2. Métodos de Monte Carlo vía cadenas de Markov

De acuerdo con Ntzoufras (2009), las técnicas MCMC son construidas basadas en una cadena de Markov que converge a una distribución de una población objetivo, en bayesiana esta técnica converge a una distribución a posteriori $p(\theta/y)$.

El algoritmo de una cadena de Markov es un proceso estocastico $\theta_1, ..., \theta_t$ de forma que la distribución de θ tiene una secuencia de una observación futura dada por todas observaciones anteriores de θ.

$$f(\theta_{t+1}/\theta_t, ..., \theta_1) = f(\theta_{t+1}/\theta_t).$$

Para construir una cadena de Markov se debe realizar los siguientes pasos:

- Seleccionar unos valores iniciales para θ_0
- Realizar el número de simulaciones necesarias ($t = 1, ..., T$) para encontrar la distribución a posteriori.
- Realizar los respectivos diagnósticos de convergencia para ir supervisando el algoritmo, si es necesario aumentar el número de simulaciones.
- Eliminar las primeras observaciones B de nuestro parámetro de interés.
- Obtenemos $\theta^{B+1}, \theta^{B+2}, ..., \theta^T$.
- Se realiza un gráfico para poder observar la convergencia de nuestro parámetro de interés.
- Se realiza los cálculos descriptivos de la distribución a posteriori encontrada.

Los Algoritmos más utilizados son el algoritmo de Metropolis Hastings y El muestro de Gibbs, los cuales se describirán a continuación.

2.2.1. Algoritmo Metropolis Hastings

El algoritmo Metropolis Hastings consiste primero en asumir una distribución del parámetro de interés con el cual se podrá generar una muestra de tamaño T, después se realizan los siguientes pasos:

- Se establece unos valores iniciales de θ^0.
- Se genera un nuevo parámetro $\theta^{'}$ el cual esta asociado a la distribución propuesta inicialmente la cual se denotara $q(\theta^{'}/\theta)$.
- Se acepta el valor con probabilidad α

$$\alpha = min\left(1, \frac{f(y/\theta^{'})f(\theta^{'})q(\theta/\theta^{'})}{f(y/\theta)f(\theta)q(\theta^{'}/\theta)}\right).$$

Después con una distribución uniforme de realiza una regla de decisión para determinar si se acepta el nuevo valor del parámetro o se mantiene la observación anterior.

Teóricamente este algoritmo converge a una distribución equilibrada independiente sin importar cual sea la distribución propuesta inicialmente pero prácticamente es importante elegir con cuidado esta distribución propuesta para no presentar problemas con la convergencia de la distribución a posteriori (Ntzoufras 2009).

2.2.2. Muestreo de Gibbs

Según Migon et al. (2008), se asume un vector de probabilidades aleatorias θ y $p(\theta)$ es la función de densidad conjunta y $p(\theta_i/\theta_{-i})$ es la distribución condicional de θ dado los otros parámetros.

Donde θ_{-j} es:

$$\theta_{-j} = (\theta_1, \theta_2, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_p).$$

El muestreo de Gibbs es un caso particular del algoritmo de Metropolis Hastings, donde la distribución propuesta $q(\theta^{'}/\theta)$ son las distribuciones condicionales $p(\theta_i/\theta_{-i}$ del parámetro θ

- Se establecen unos valores iniciales $\theta^0 = \theta_1^0, ..., \theta_p^0$.
- Para cada iteración se realiza:

  a) se genera θ_1^j de $p(\theta_1/\theta_2^{j-1}, ..., \theta_p^{j-1})$.

  b) se genera θ_2^j de $p(\theta_2/\theta_1^{j-1}, \theta_3^{j-1}, ..., \theta_p^{j-1})$.

  c) se genera θ_3^j de $p(\theta_1/\theta_2^{j-1}, \theta_4^{j-1}, ..., \theta_p^{j-1})$.

  d) Por ultimo se genera θ_p^j de $p(\theta_1/\theta_2^{j-1}, ..., \theta_{p-1}^{j-1})$.

Finalmente se esta encontrando la distribución a posteriori del parámetro θ por medio de las distribuciones condicionales.

2.3. Modelos jerárquicos

De acuerdo con Migon et al. (2008), los modelos jerárquicos consisten en una construcción de un modelo en diferentes niveles, esta construcción se realiza mediante una combinación de una información estructural con una información subjetiva.

Donde la información estructural consiste en la división de los niveles y la información subjetiva en la especificación de cada nivel.

Una explicación sencilla del concepto de jerarquía es un modelo de dos niveles. El primer nivel hace referencia al comportamiento de la variable respuesta Y (observación general) asociado a unas covariables (observaciones individuales) X.

$$Y_i = X_i\beta_i + \varepsilon_i, \quad \varepsilon_i \sim iidN(0, \sigma^2 I_{ni}).$$

El segundo nivel consiste en estimar el valor de β mediante un modelo o bajo un supuesto de una distribución con unos parámetros conocidos.

$$\beta_i = \Delta' Z_i + \upsilon_i \quad \upsilon_i \sim iidN(0, V_\beta).$$

Dependiendo la complejidad del problema se pueden realizar modelos con K niveles, cuanto mas alto es el nivel, mas difícil es la especificación de las distribuciones de cada nivel (Migon et al. 2008).

Volviendo al modelo del primer nivel, la matriz de varianzas y covarianzas σ^2 de los errores tiene la siguiente distribución:

$$\sigma^2 \sim \frac{\upsilon_i S_{0i}^2}{X_{\upsilon i}^2},$$

$$\upsilon_i \sim iidN(0, V_\beta).$$

Una forma resumida del proceso que se lleva acabo es el siguiente:

$$Y|X_i, \beta_i, \sigma_i^2,$$
$$\beta_i|Z_i, \Delta, V_\beta,$$
$$\sigma_i^2|\upsilon_i, S_{0i}^2.$$

3. Descripción de los datos

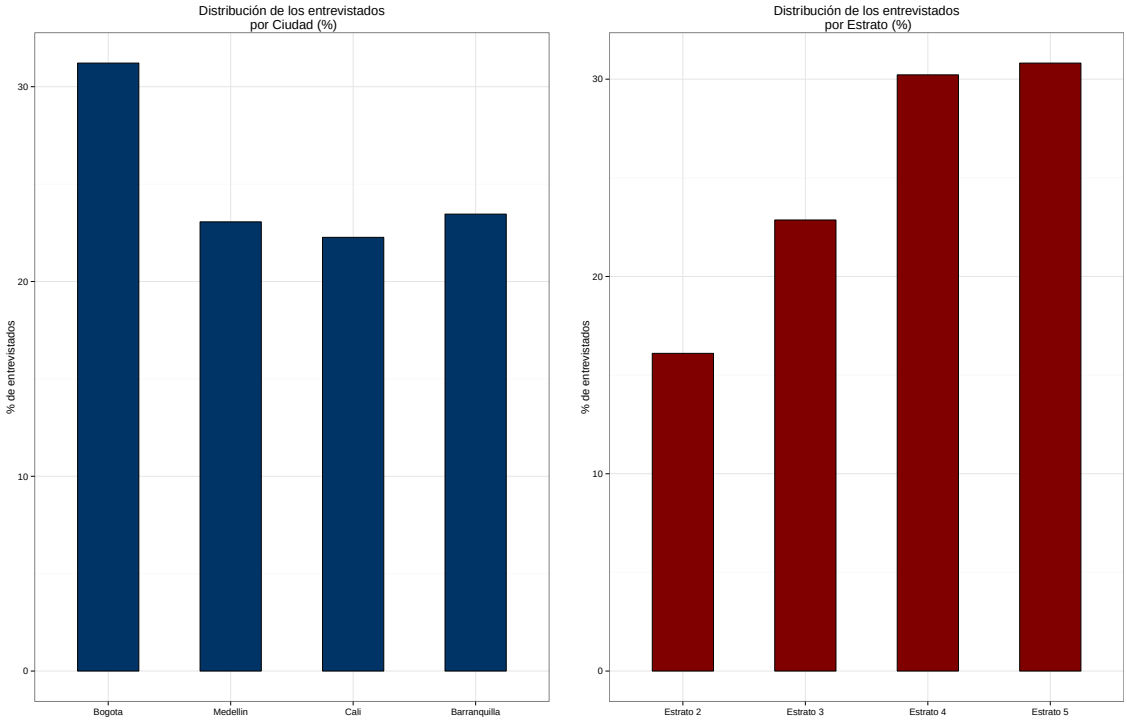
El conjunto de datos consiste en un total de 503 individuos, la información fue recogida usando un muestreo no probabilistico por cuotas en las ciudades de Bogotá, Cali, Medellín y Barranquilla en los estratos 2, 3, 4, y 5.

También el conjunto de datos esta compuesta por un grupo de variables sociodemográficas y las comparaciones de 5 ofertas propuestas por el investigador. La información sociodemográfica consiste en la ciudad de residencia, la edad exacta, estrato con el que llega el recibo de energía, la ocupación, la cantidad de dinero que gasta mensualmente en el operador móvil, el genero, el estado civil y a que operador pertenece.

Las comparaciones que realizaba cada individuo consistía en empezar con la comparación de dos ofertas al azar, él escogia la oferta más atractiva; posteriormente la oferta ganadora se comparaba con otra de las ofertas restantes y de igual forma él seleccionaba la oferta mas atractiva y así sucesivamente hasta terminar la comparación de las ofertas restantes, Como se indico en la introducción, cada individuo realizaba $k-1$ comparaciones, es decir que en este caso cada individuo realizó 4 comparaciones.

3.1. Variables sociodemográficas

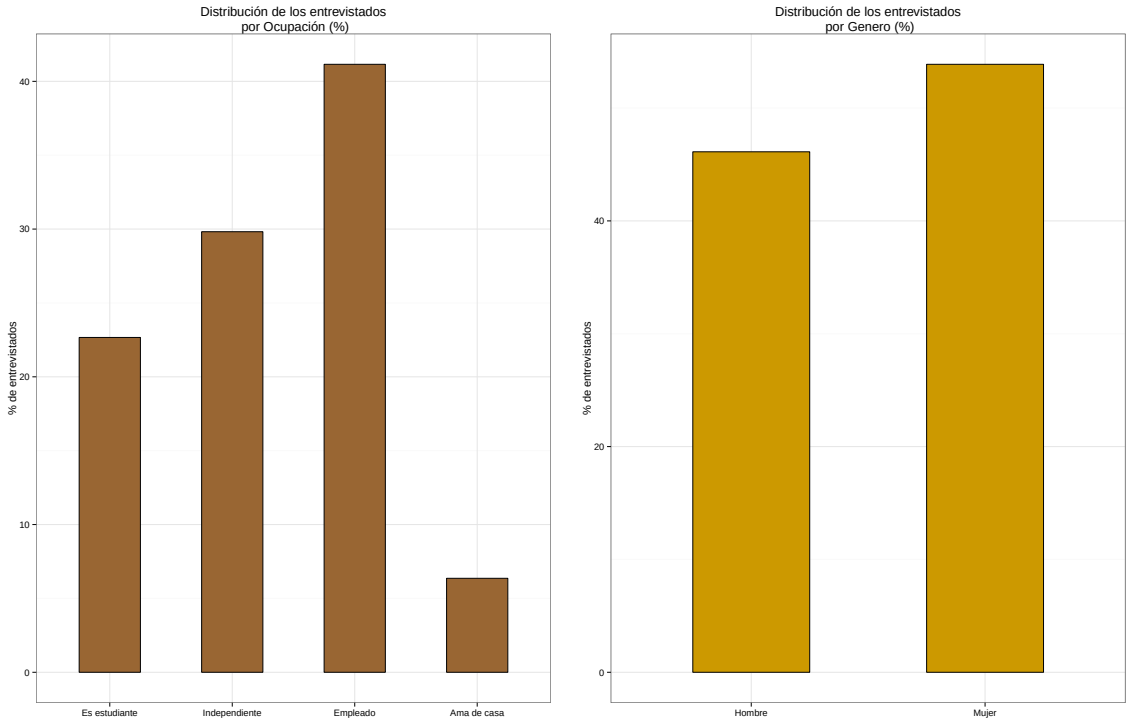
Por medio de un diagrama de barras se puede observar la distribución de las variables sociodemográficas, en la figura 1 se puede ilustrar la distribución para las variables de la ciudad y de los estratos. Se puede observar que el 31 % de los individuos residen en Bogotá, el 23 % en la ciudad de Medellín, en Cali residen el 22 % de los individuos y el 24 % en la ciudad de Barranquilla. En los estratos 4 y 5 se encuentra aproximadamente el 60 % de los individuos y el restante 40 % entre los estratos 2 y 3.



(a) Ciudad (b) Estrato

Figura 1: Distribución de la ciudad y estrato

Las variables relacionadas con la ocupación y el genero se encuentran en la figura 2, donde el 42 % de los individuos son empleados, el 29 % son independientes, el 23 % son estudiantes y el 6 % amas de casa. Las mujeres corresponden al 54 % y los hombres el restante 46 %.

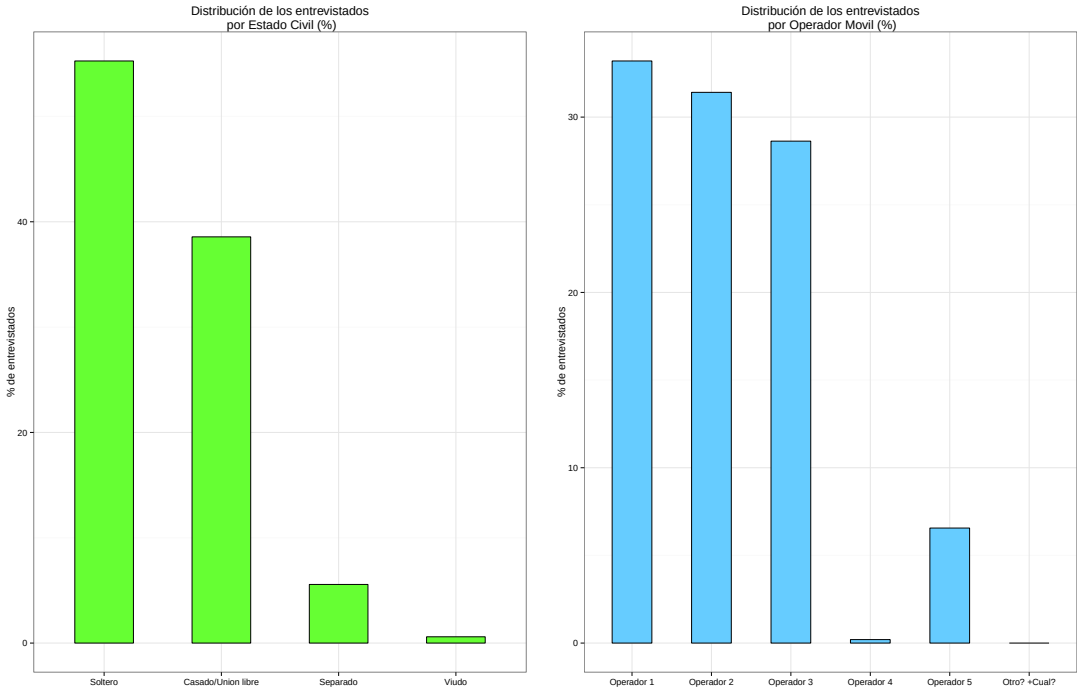


(a) Ocupación (b) Genero

Figura 2: Distribución de la ocupación y genero

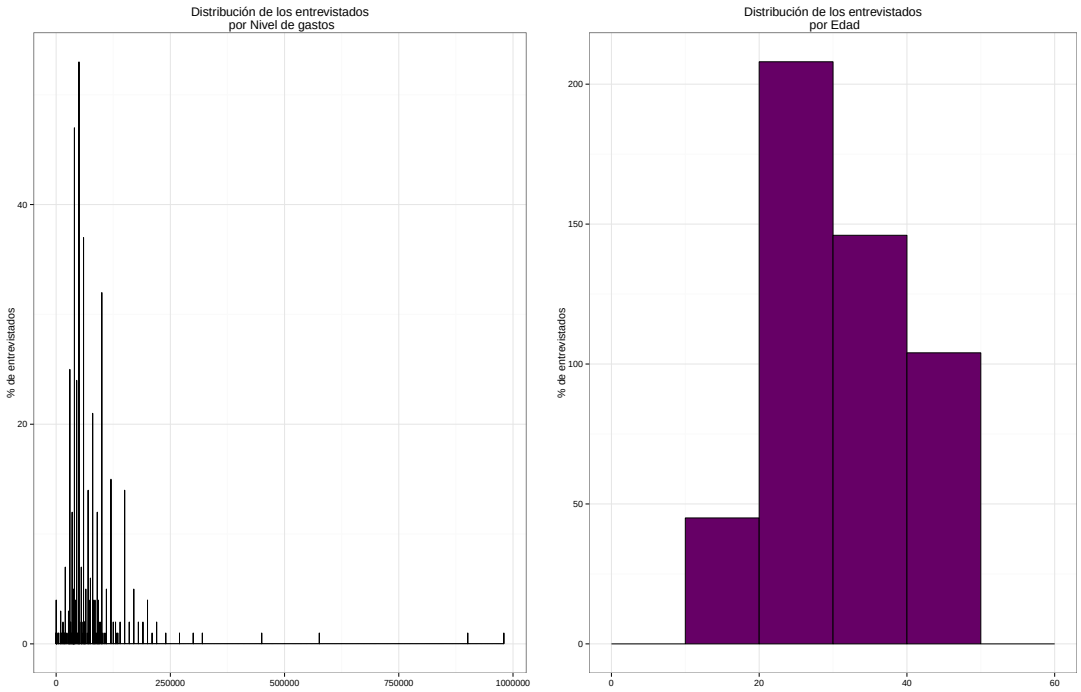
En la figura 3 se encuentran las variables de estados civil y el operador que pertenece cada individuo, donde el 55 % de los individuos son solteros, el 39 % son individuos casados o que viven en unión libre, el 5 % son separados y el 1 % son viudos. Entre los operadores móviles, los operadores 1 y 2 corresponden aproximadamente al 65 % de los individuos, el 28 % con individuos con operador 3, el operador 5 corresponde al 7 % de los individuos y el 0,2 % son individuos con operador 4.

Por ultimo, se puede observar en la figura 4 las distribuciones de los gastos en telefonía móvil y la edad de los individuos, se puede concluir que el promedio de los gastos de los individuos es de $ 73.130 con unas pocas observaciones demasiadas altas y el promedio de edad de los individuos es de 31 años aproximadamente.



(a) Estado civil (b) Operador Móvil

Figura 3: Distribución del estado civil y el operador móvil



(a) Gastos en telefonía móvil (b) Edad

Figura 4: Distribución de los gastos y edad

4. Aplicación

Para identificar los nichos del mercado para las diferentes ofertas, el primer paso es encontrar las probabilidades de que cada uno de los sujetos seleccione cada una de las 5 ofertas; mediante el modelo jerárquico logístico multinomial bayesiano que se propone se pueden calcular dichas probabilidades.

$$P(y_i = j) = \frac{e^{\beta_{ij}}}{1 + \sum_{j=2}^{5} e^{\beta_{ij}}}.$$

Se construyó el modelo según Agresti (2002) donde i representa el sujeto ($i = 1, ..., 503$), j la oferta ($j = 2, ...5$), entonces $P(y_i = 3)$ significa la probabilidad de que el sujeto i seleccione la oferta numero 3. Donde se tomo como referencia la primera oferta y su calculo se hizo mediante el siguiente modelo:

$$P(y_i = 1) = \frac{1}{1 + \sum_{j=2}^{5} e^{\beta_{ij}}},$$

de esta forma se necesita estimar el parámetro β para cada individuo, este parámetro β se pueden estimar mediante el siguiente modelo:

$$\beta_i = \Delta' Z_i + \upsilon_i \quad \upsilon_i \sim iidN(0, V_\beta),$$

donde la distribución de cada β_i es:

$$\beta_i \sim N(\Delta' Z_i, V_\beta).$$

Para poder encontrar la distribución a posteriori del parámetro β_i y de igual forma la distribución a posteriori del parámetro Δ se utiliza la función rhierMnlDP del paquete bayesm del programa estadístico R tomando como base los ejercicios de Rossi et al. (2005).

Se comienza utilizando una distribución inversa Wishart a priori para el parámetro V_β y una distribución normal a priori para el parámetro Δ condicionado al parámetro V_β.

$$vec(\Delta / V_\beta) \sim N(0, A^{-1} \otimes V_\beta),$$

.

$$V_\beta \sim IW(\upsilon, V_0).$$

.

Donde la distribución inversa Wishart en la aplicación bayesiana es muy frecuente utilizarla como una información a priori de la matriz de varianzas y covarianzas de una distribución normal multivariada. Para más detalles consultar Nydick (2012).

Con la información a priori que se propone, se procede a construir una función de densidad normal inversa whishart (H).

$$H(\Delta, V_\beta).$$

Donde esta nueva función de densidad H sirve como uno de los parámetros para realizar un proceso Dirichlet, donde se esta obteniendo un vector de probabilidades X asociadas al parámetro de interés que se quiere obtener.

$$X \sim DP(H, \alpha).$$

Este vector de probabilidades X asociadas al parámetro de interés tiene una distribución Dirichlet, para mas detalles consultar Lid et al. (2010).

$$(X(\beta_1), X(\beta_2), ..., X(\beta_5)) \sim Dir(\alpha H(\beta_1), \alpha H(\beta_2), ..., \alpha H(\beta_5)).$$

Después de obtener ese vector de probabilidades se realiza el algoritmo de Metropolis Hastings para encontrar la primera iteración del parámetro de interés. En la regla de decisión para obtener la primera iteración se esta involucrando la variable respuesta y la matriz diseño de las comparaciones realizadas de cada individuo.

En la tabla 1 se puede observar como se construyo la variable respuesta y la matriz diseño, donde la variable respuesta es la oferta ganadora de la comparación que se realizo, se elaboro la matriz diseño de forma de que la oferta ganadora se colocaba 1 y la oferta que perdió en la comparación se colocaba -1.

No Ind.	**Variable Respuesta Oferta Ganadora**	**Matriz Diseño**				
		Of. 1	**Of. 2**	**Of. 3**	**Of. 4**	**Of. 5**
1	3	0	-1	1	0	0
1	4	0	0	-1	1	0
1	1	1	0	0	-1	0
1	1	1	0	0	0	-1

Tabla 1: Variable respuesta y matriz diseño

Después de encontrar la primera iteración se procede a realizar un muestreo de Gibbs para encontrar la primera iteración del valor de $\Delta' Z_i$ y de V_β, con esa primera iteración se vuelve a construir la función de densidad H y otra vez por medio del proceso Dirichlet se encuentra un vector de probabilidades con distribución Dirichelt y de igual forma se procede a realizar el algoritmo de Metropolis Hastings y se encuentra la segunda iteración y se vuelve a realizar el muestreo de Gibbs para encontrar la segunda iteración de $\Delta' Z_i$ y de V_β y así el procedimiento de vuelve repetitivo hasta terminar con el numero de iteraciones que se propongan.

Se propuso realizar 1.000.000 de iteraciones donde cada 10 observaciones se iban descartando, de esta forma se esta eliminando la correlación producida por el método de convergencia y finalmente se obtiene un total de 100.000 simulaciones.

La figura 5 muestra las simulaciones de la distribución a posteriori del parámetro Δ, donde se puede observar que existe una convergencia a medida que aumenta la cantidad de iteraciones.

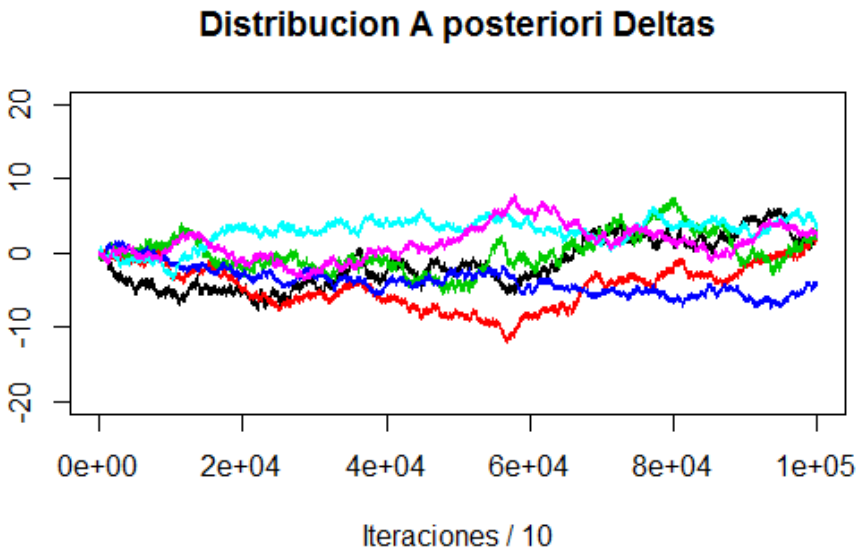


Figura 5

La tabla 2 muestra los resultados del promedio de la distribución a posteriori del parámetro Δ

Ofertas	**Edad**	**Gastos**	**Cali**	**Bogotá**
Oferta 1	-0,049418	-0,052740	-0,059167	-0,065706
Oferta 2	-0,582532	-0,399636	-0,663750	-0,364849
Oferta 3	-0,604068	-0,188663	-0,342406	-0,531173
Oferta 4	-0,030959	-0,435306	0,092090	-0,172957
Oferta 5	-0,673713	-0,736541	-0,759631	-0,600463

Ofertas	**Estrato 2**	**Estrato 3**	**Estrato 4**	**Estrato 5**
Oferta 1	-0,055262	0,000002	0,000002	0,000002
Oferta 2	-0,629160	-0,516616	-0,796608	-0,935478
Oferta 3	-0,504723	-0,382534	-0,627165	1,243800
Oferta 4	-0,203665	-0,192476	-0,226792	-0,264132
Oferta 5	0,571137	0,401143	0,376414	0,363486

Ofertas	**Ocup. Estudiante**	**Ocup. Independiente**	**Hombre**	**Soltero**
Oferta 1	0,000001	0,000001	-0,414401	-0,568500
Oferta 2	-1,008722	-0,911443	-0,482927	-0,408362
Oferta 3	0,971537	1,012990	1,222220	-0,081111
Oferta 4	1,156579	1,307700	1,223100	1,301300
Oferta 5	0,337595	1,612900	1,569700	1,592400

Tabla 2: Promedio distribución a posteriori del parámetro Δ

donde las sujetos de mayor edad son mas sensibles en las ofertas 3 y 5, a un mayor nivel de gastos en el operador movil los sujetos son mas sensibles en las ofertas 4 y 5. En Cali los sujetos son mas sensibles en las ofertas 2 y 5, En la ciudad de Bogotá los sujetos son mas sensibles en las ofertas 3 y 5, los sujetos que viven en estratos 2, 3 y 4 son mas sensibles en la oferta 5 y en el estrato 5 los sujetos son mas sensibles

en la oferta 3. Para los estudiantes y las personas independientes son mas susceptibles en la oferta 4 y los hombres y los sujetos que sean solteros son mas sensibles en la oferta 5.

La figura 5 muestra la distribución log verosimil del parámetro β a posteriori.

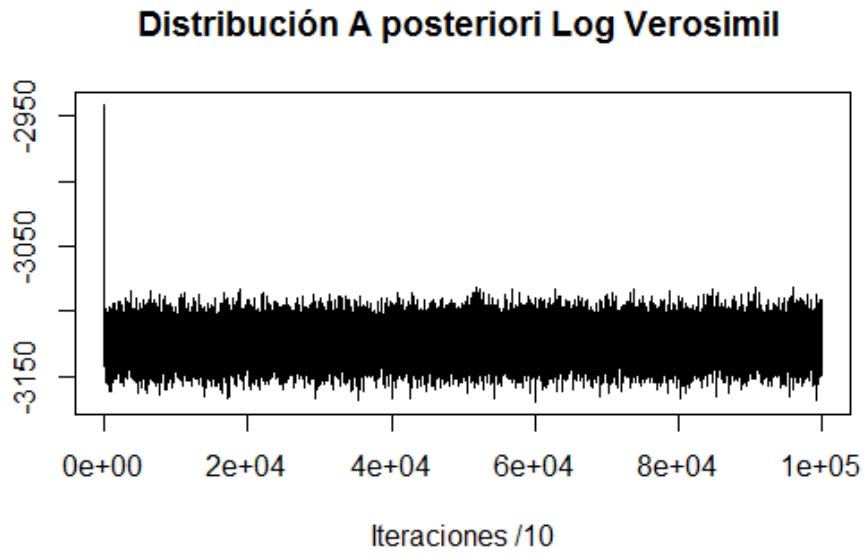


Figura 6

Para determinar la convergencia de las simulaciones de la distribución log verosimil del parámetro β y del mismo modo determinar cuantas simulaciones se deben quemar, se realizó un muestreo bootstrap el cual consiste primero en seleccionar una partición de las simulaciones realizadas y de esa partición se genera una muestra con reemplazo del mismo tamaño de la partición y se procede a calcular el promedio de esta muestra.

Este procedimiento se realizó 3.000 veces, después de finalizar esta simulación se ordenaron los promedios bootstrap y se generó un intervalo que contenía desde el 2,5 % hasta el 97,5 % del total de los promedios bootstrap para obtener un 5 % de significancia. De esta forma se obtuvo el primer intervalo de los promedios bootstrap de la primera partición, luego con una segunda partición (diferente de la primera) se realizó el mismo procedimiento para encontrar un segundo intervalo de promedios bootstrap con el mismo nivel de significancia.

Si los dos intervalos comparten limites se puede concluir que si existe una convergencia en las simulaciones de la distribución log verosimil del parámetro β y dependiendo de las particiones que se seleccionen se puede determinar cuantas simulaciones se deben quemar.

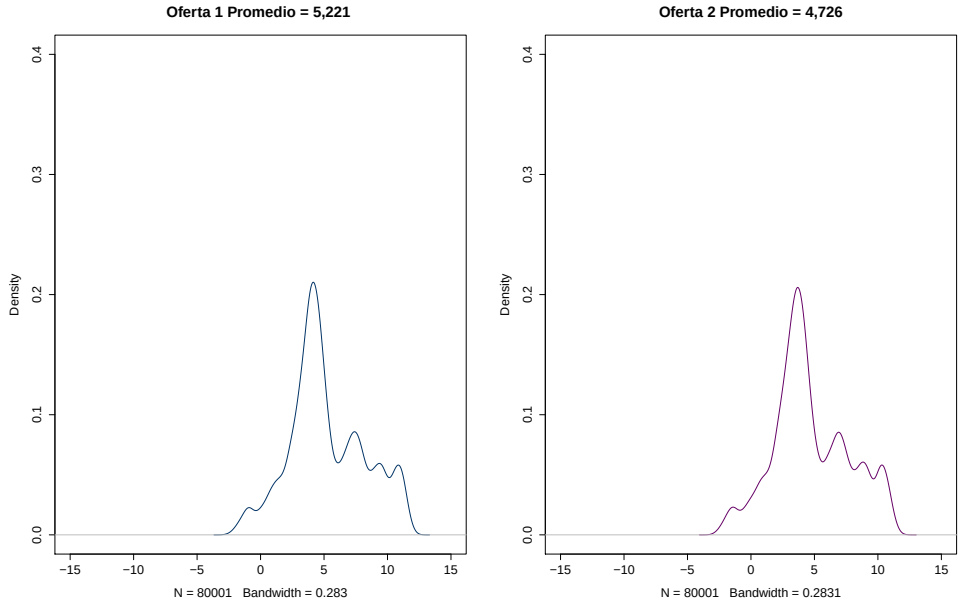
La primera partición que se selecciono fue entre 20.000 y 35.000 simulaciones y la segunda partición fue entre 70.000 y 85.000 simulaciones, los resultados se pueden observar en la tabla 3:

No Particiones	Limite Inferior	Limite Superior
Particion 1	-3.123,77	-3.123,42
Particion 2	-3.123,62	-3.123,27

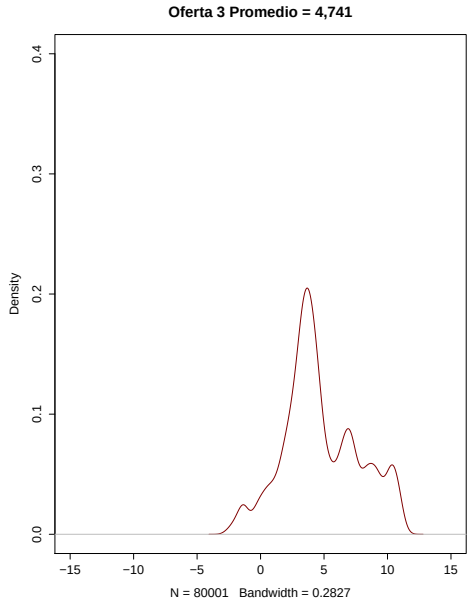
Tabla 3: Intervalos de las particiones

Como los dos intervalos comparten limites se puede determinar que si existe una convergencia en la distribución a posteriori log verosimil del parámetro β, como esta metodología se realizo después de las primeras 20.000 simulaciones es recomendable quemar estas primeras simulaciones.

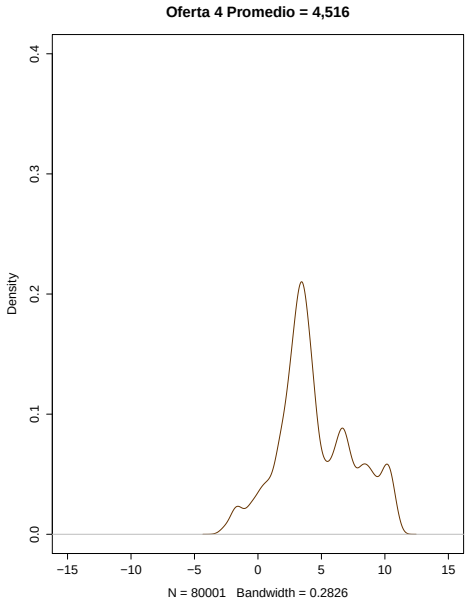
Se procede a construir la distribución a posteriori del parámetro β de los individuos quemando las primeras 20.000 simulaciones. La figura 7 muestra la distribución a posteriori del parámetro β del individuo 57 y la figura 8 la distribución a posteriori del parámetro β del individuo 324.



(a) (b)


Oferta 3 Promedio = 4,741
Density
0.0
0.1
0.2
0.3
0.4
−15
−10
−5
0
5
10
15
N = 80001 Bandwidth = 0.2827

(c)


Oferta 4 Promedio = 4,516
Density
0.0
0.1
0.2
0.3
0.4
−15
−10
−5
0
5
10
15
N = 80001 Bandwidth = 0.2826

(d)

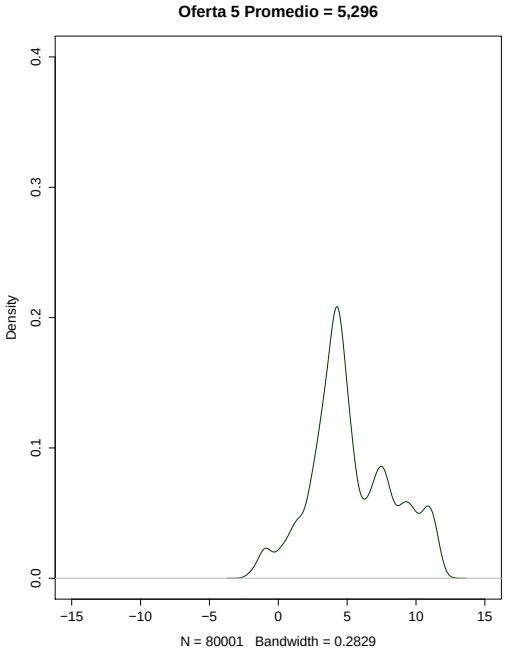
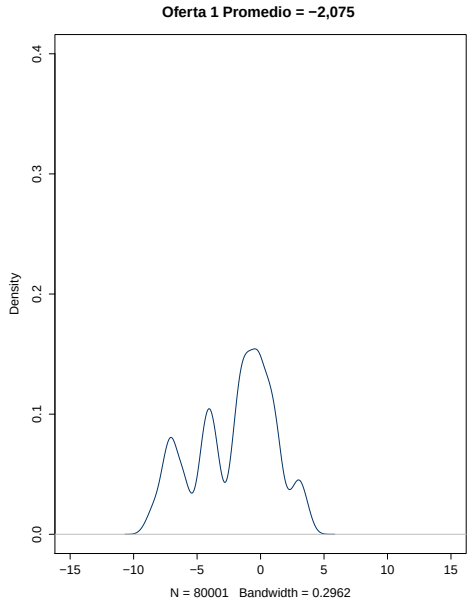
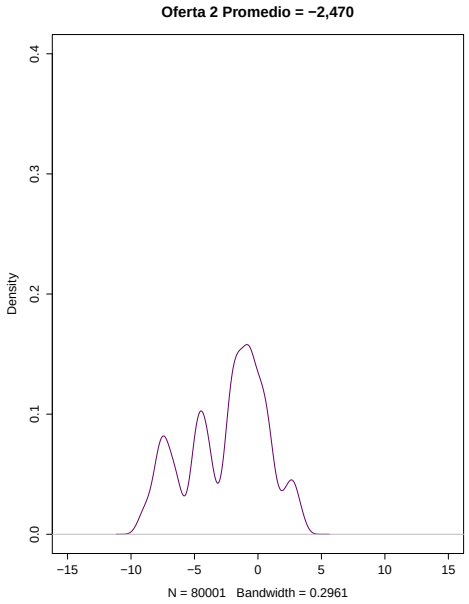

Oferta 5 Promedio = 5,296
Density
0.0
0.1
0.2
0.3
0.4
−15
−10
−5
0
5
10
15
N = 80001 Bandwidth = 0.2829

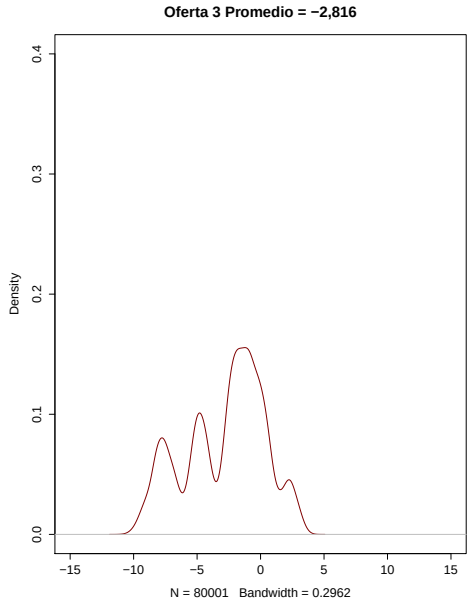
Figura 7


Oferta 1 Promedio = −2,075
Density
N = 80001 Bandwidth = 0.2962

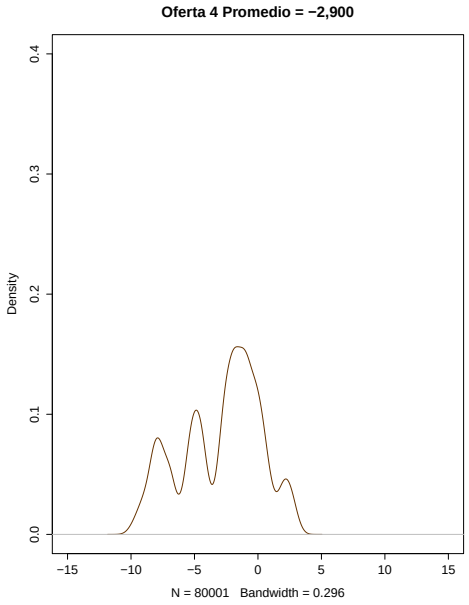
(a)


Oferta 2 Promedio = −2,470
Density
N = 80001 Bandwidth = 0.2961

(b)


Oferta 3 Promedio = −2,816
Density
N = 80001 Bandwidth = 0.2962

(c)


Oferta 4 Promedio = −2,900
Density
N = 80001 Bandwidth = 0.296

(d)

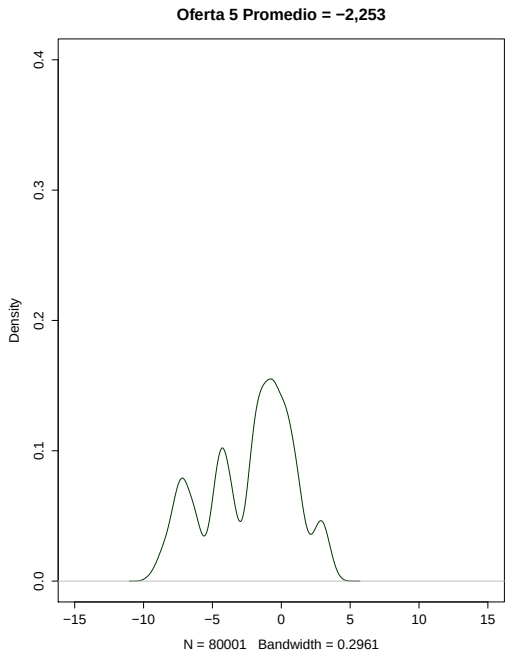


Figura 8

Como se puede observar en los gráficos de los individuos 57 y 324, se observa que se asemejan a una distribución normal pero con la diferencia en su respectiva ubicación sobre la recta. Esta diferencia se puede evidenciar cuando se calcula el promedio en cada una de las distribuciones de los individuos 57 y 324, por ejemplo si se observa el promedio de la distribución de la primera oferta, el promedio del individuo 57 es 5,221 y el promedio del individuo 324 es -2,075.

Dentro de la gráfica de cada distribución se puede encontrar el resultado del valor h de Bandwidth, el cual consiste en que por medio de este valor se puede calcular el estimador de la función de densidad de Kernel, este estimador consiste en poder encontrar la convergencia de la verdadera función de densidad del parámetro de interés, donde valores pequeños de valor h de Bandwidth dan estimaciones muy aproximadas. Por lo general este valor depende del tamaño de la muestra, para mas detalles consultar Wasserman (2006).

Ya con los β encontrados para todos los individuos, el siguiente paso es calcular el promedio de cada uno de los β y este valor servirá como un estimador del parámetro dentro del modelo para poder calcular dichas probabilidades para cada individuo. A continuación se ilustra el calculo de la probabilidad de que el individuo 57 seleccione la oferta numero 2.

La tabla 4 contiene los valores del parámetro β en cada una de las 5 ofertas.

Parámetro	**Valor**
β_1	5,221
β_2	4,726
β_3	4,741
β_4	4,516
β_5	5,296

Tabla 4

El modelo queda construido de la siguiente manera:

$$P(y_{57} = 2) = \frac{e^{\beta_{57,2}}}{1 + \sum_{j=2}^{5} e^{\beta_{57,j}}},$$

$$P(y_{57} = 2) = \frac{e^{4,726}}{1 + e^{4,726} + e^{4,741} + e^{4,516} + e^{5,296}},$$

$$P(y_{57} = 2) = 0,2775.$$

La probabilidad de que el individuo seleccione la oferta 2 es de 27,75 % La tabla 5 muestra las probabilidades de los primeros 10 individuos de seleccionar cada una de las ofertas.

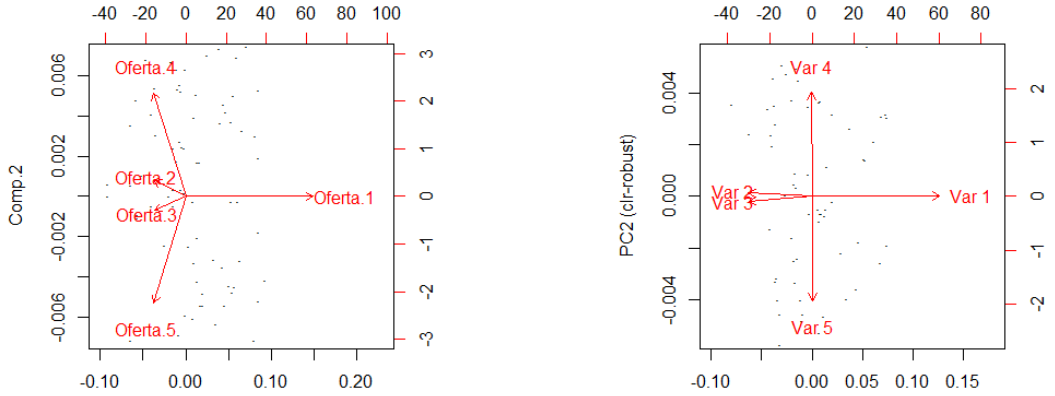
Ind.	**Oferta 1**	**Oferta 2**	**Oferta 3**	**Oferta 4**	**Oferta 5**
1	0,6629625	0,0876176	0,0697981	0,0846035	0,0950182
2	0,6421631	0,0955423	0,0681362	0,0783451	0,1158133
3	0,0129709	0,2340355	0,2710731	0,1728302	0,3090903
4	0,1322760	0,2038025	0,1585783	0,1862458	0,3190975
5	0,0115092	0,2210188	0,2187852	0,1671480	0,3815389
6	0,6677197	0,0851054	0,0711364	0,0678407	0,1081979
7	0,1730564	0,2033914	0,1800100	0,1376562	0,3058860
8	0,9840200	0,0040965	0,0034693	0,0033351	0,0050790
9	0,3984647	0,1575636	0,1231595	0,1471581	0,1736542
10	0,3400891	0,1335599	0,1087060	0,1689494	0,2486956

Tabla 5: Probabilidades de selección de las 5 ofertas de los primeros 10 individuos

Con las probabilidades calculadas para cada individuo y con sus respectivas variables sociodemográficas, se realiza un análisis de componentes principales para poder detectar las principales dimensiones de variabilidad.

Según Van den Boogaart & Tolosana (2013), los datos composicionales corresponden a cantidades o particiones relativas que siempre suman 1 o 100 % para cada individuo, por lo general estas particiones están escondidas de diferentes formas en las observaciones, de esta forma los análisis estadísticos multivariados contienen implicaciones en sus supuestos por la suma constante de cada individuo, en el caso de estudio la suma de probabilidades de que cada individuo seleccione cada una de las 5 ofertas es igual a 1.

La figura 9 contiene los gráficos de los planos factoriales de las componentes principales normales y de las componentes principales para datos composicionales, estos gráficos están construidos por los vectores propios de las ofertas donde se esta tomando los 2 primeros factores, de forma general se puede concluir que la oferta 1 se asocia de forma negativa con la oferta 2 y 3, con respecto a las demás ofertas no existe algún tipo de asociación, por el contrario las demás ofertas sin mencionar la oferta 1 tienen un determinado grado de asociación entre ellas.



(a) Plano factorial de las componentes principales normales (b) Plano factorial de las componentes principales para datos composicionales

Figura 9: Gráfico de las componentes principales

La tabla 6 corresponde al porcentaje de varianza retenido por los dos primeros factores, donde se puede observar que se retiene el 98,41 % de la variabilidad total cuando se realiza las componentes principales normales y se retiene un 99,80 % de la variabilidad total cuando se realiza las componentes principales para datos composicionales.

Tipo de análisis	% de varianza
Componentes principales	98,41 %
Componentes principales con datos categoricos	99,80 %

Tabla 6: Porcentaje de varianza retenido con dos factores

Como no existe alguna diferencia significativa entre las dos metodologías, se toma la decisión de seguir los métodos de clasificación con el análisis de componentes principales normales.

Para determinar los nichos del mercado se toma la decisión de realizar una clasificación jerárquica para poder segmentar a los sujetos, en la figura 10 se construyó un dendrograma para poder observar la segmentación realizada.

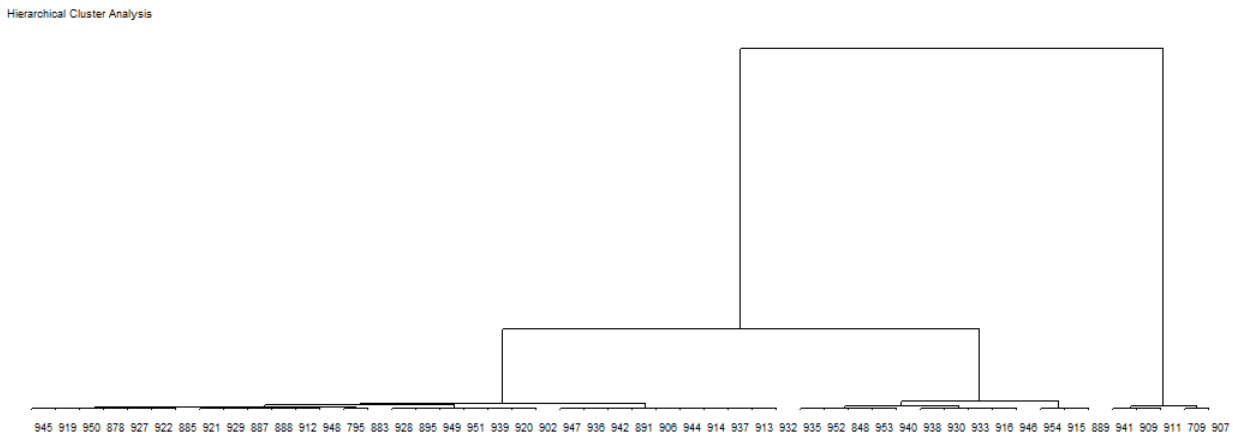


Figura 10: Dendrograma

De una forma visual se puede concluir que se puede segmentar a los individuos en 2 o en 3 grupos, para tener un criterio valido de saber cuantos grupos se deben realizar, la figura 11 nos ayuda a determinar cuantos grupos se deben realizar utilizando las curvas de nivel, se toma la decisión de que se deben realizar 2 grupos por que la varianza entre dos grupos es mas alta que cuando se divide en tres grupos.

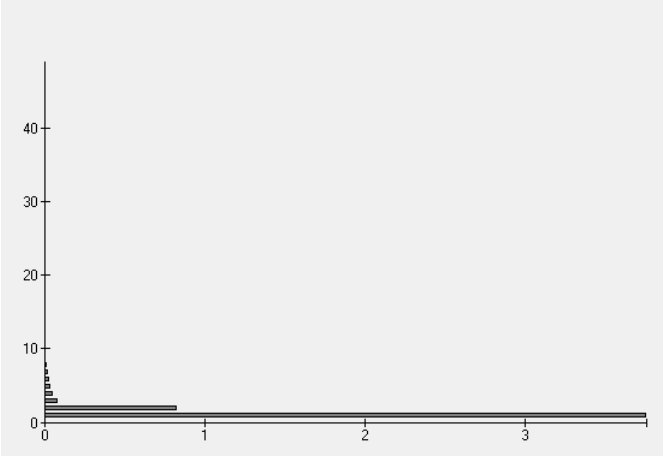


Figura 11: Curvas de nivel

La figura 12 representa de forma visual el grupo al que pertenece cada sujeto. El 61,7 % de los sujetos quedaron clasificados en el grupo 1 y el restante 38,3 % en el grupo 2.

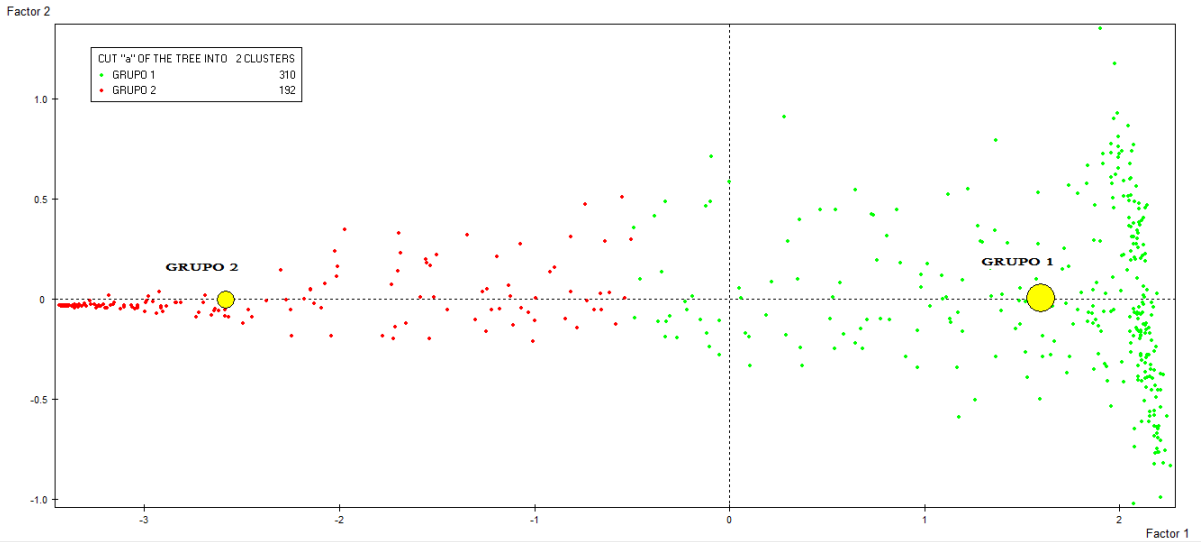


Figura 12: Clasificación de los sujetos

Después de realizar la respectiva clasificación, se puede interpretar que características diferencian a cada uno de los grupos y lo mas importante que oferta es la mas adecuada para cada grupo. El grupo 1 esta conformado principalmente por mujeres solteras, que pertenezcan al estrato 2, cuya ocupación sean empleadas o amas de casa y que residan en la ciudad de Bogotá, a este grupo es ideal ofrecerles las ofertas 2, 3, 4 y 5; el grupo 2 esta compuesto por hombres independientes, que su estado civil es que sea casado o este en unión libre, que pertenezcan al estrato 5 que y residan en la ciudad de Cali, a este grupo es ideal ofrecerle la oferta 1.

Para complementar el análisis anterior, la tabla 7 muestra el promedio de cada una de las probabilidades dentro de cada uno de los grupos, donde se puede concluir que el promedio de las probabilidades de seleccionar la oferta 1 del grupo 1 es demasiado baja comparada con las otras ofertas, mientras que el mismo promedio de la oferta 1 en el grupo 2 es demasiado alta con respecto a las probabilidades de seleccionar las demás ofertas.

Segmento	Oferta 1	Oferta 2	Oferta 3	Oferta 4	Oferta 5
Grupo 1	0,096408	0,222493	0,188541	0,195772	0,296786
Grupo 2	0,845852	0,037937	0,032994	0,031673	0,051544

Tabla 7: Probabilidades de selección de las 5 ofertas dentro de los grupos

5. Conclusiones

Para dar solución a la problemática de que como estimar las preferencias del consumidor, se realizó un Modelo Jerarquico Bayesiano Logistico Multinomial. Como la variable respuesta tiene distribución multinomial, se construyo el modelo para calcular la probabilidad de que cada sujeto seleccione cada una de las 5 ofertas propuestas.

Para la estimación del parámetro β del modelo (este modelo se divide en dos niveles por que el parámetro β esta asociado a una distribución que depende de las variables sociodemográficas), se utilizó la función

rhierMnlDP del paquete bayesm del programa estadístico R.

Por ultimo utilizando el conjunto de probabilidades de cada sujeto y con sus respectivas variables sociodemográficas, se realizó un análisis de componentes principales y un método de clasificación para poder segmentar a los sujetos en dos grupos y se pudo identificar los nichos de mercado para cada una de las diferentes ofertas.

6. Trabajos futuros

Para futuras investigaciones, se puede pensar en aplicar un modelo autorregresivo espacial bayesiano para determinar si las preferencias de los sujetos en el momento de seleccionar la oferta se ve influenciado por el comportamiento o preferencias de otros sujetos que pertenecen a su circulo social. También se podría profundizar mas acerca de la metodología de la estimación de la función de densidad del parámetro de interés, ya que en este trabajo solo se menciono el resultado obtenido puesto que no era uno de los objetivos de investigación.

Ademas se podría considerar en comparar los resultados del modelo propuesto contra los resultados de otro modelo construido bajo una metodología diferente (modelo clásico, modelo no parámetrico, modelo mixto, etc.) y así determinar que modelo es el mas adecuado cuando se realiza una metodología trade off donde se mantiene la oferta ganadora.

Finalmente también se puede realizar la comparación de los resultados finales de la metodológica de trade off propuesta contra la metodología clásica del trade off (selección de ofertas al azar) y así determinar que metodología es la mas adecuada para identificar los nichos de mercado.

7. Apéndice

```
library(bayesm)


#### Variables Sociodemograficas
Z1


#### Variables respuestas y matriz diseño
seleccion

Z1[,1]=Z1[,1]-mean(Z1[,1])
Z1[,2]=Z1[,2]-mean(Z1[,2])
Z1[,3]=Z1[,3]-mean(Z1[,3])
Z1[,4]=Z1[,4]-mean(Z1[,4])
Z1[,5]=Z1[,5]-mean(Z1[,5])
Z1[,6]=Z1[,6]-mean(Z1[,6])
Z1[,7]=Z1[,7]-mean(Z1[,7])
Z1[,8]=Z1[,8]-mean(Z1[,8])
Z1[,9]=Z1[,9]-mean(Z1[,9])
Z1[,10]=Z1[,10]-mean(Z1[,10])
Z1[,11]=Z1[,11]-mean(Z1[,11])
Z1[,12]=Z1[,12]-mean(Z1[,12])
Z1[,13]=Z1[,13]-mean(Z1[,13])
Z1[,14]=Z1[,14]-mean(Z1[,14])
```

```
Z=as.matrix(Z1)

hh=levels(factor(seleccion$LLAVE))
nhh=length(hh)

#### da inicio a la variable Data como una lista de longitud 0

lgtdata=NULL

#### se realiza un ciclo para que cada elemento contenga los respectivos datos del encuestado

for (i in 1:nhh) {
y=seleccion[seleccion[,1]==hh[i],2]
nobs=length(y)
X=as.matrix(seleccion[seleccion[,1]==hh[i],c(3:7)])
lgtdata[[i]]=list(y=y,X=X)
}

lgtdata2¡-lgtdata
for(j in 1:503){
for(i in 1:4){
lgtdata2[[j]]$X < − rbind(lgtdata2[[j]]$X,lgtdata[[j]]$X)
}
}

lgtdata < − lgtdata2

Data=list(p=5,lgtdata=lgtdata, Z=Z)

#### verificamos que las opciones de los encuestados se almacenana como vectores
is.vector(Data$lgtdata[[1]]$y)

#### verificamos que los niveles de los atributos se almacene como matriz
is.matrix(Data$lgtdata[[1]]$X)

#### cadena de Markov
Mcmc=list(R=1000000,keep=10)

out=rhierMnlDP(Data=Data,Mcmc=Mcmc)

index=1*c( 41, 46, 51, 56, 61, 66)

matplot(out$Deltadraw[,index],type="l", xlab=Ïteraciones / 10", ylab=, main="Distribucion A
posteriori Deltas",ylim=c(-20,20))
```

```
delta=t(matrix(apply(out$Deltadraw[20000:100000,],2,mean),ncol=5))

plot(out$loglike, type="l", xlab= Ïteraciones /10", ylab=, main= "Distribución A posteriori Log
Verosimil ")

mean(out$betadraw[324,1,20000:100000])
mean(out$betadraw[324,2,20000:100000])
mean(out$betadraw[324,3,20000:100000])
mean(out$betadraw[324,4,20000:100000])
mean(out$betadraw[324,5,20000:100000])

plot(density(out$betadraw[324,1,20000:100000]), main = .ºferta 1 Promedio = -2,075 ", xlim=c(-15,15),
ylim=c(0,0.4),col="#003366")
plot(density(out$betadraw[324,2,20000:100000]), main = .ºferta 2 Promedio = -2,470", xlim=c(-15,15),
ylim=c(0,0.4), col="#660066")
plot(density(out$betadraw[324,3,20000:100000]), main = .ºferta 3 Promedio = -2,816", xlim=c(-15,15),
ylim=c(0,0.4), col="#800000")
plot(density(out$betadraw[324,4,20000:100000]), main = .ºferta 4 Promedio = -2,900", xlim=c(-15,15),
ylim=c(0,0.4), col="#663300")
plot(density(out$betadraw[324,5,20000:100000]), main = .ºferta 5 Promedio = -2,253", xlim=c(-15,15),
ylim=c(0,0.4), col="#003300")

#### Betas
hh=levels(factor(seleccion$LLAVE))
nhh=length(hh)

Betas=NULL
for (i in 1:nhh) {
b1=as.matrix(mean(out$betadraw[i,1,20000:100000]))
b2=as.matrix(mean(out$betadraw[i,2,20000:100000]))
b3=as.matrix(mean(out$betadraw[i,3,20000:100000]))
b4=as.matrix(mean(out$betadraw[i,4,20000:100000]))
b5=as.matrix(mean(out$betadraw[i,5,20000:100000]))
Betas[[i]]=cbind(b1,b2,b3,b4,b5)
}

p < − matrix(NA,nrow=503, ncol=4)

for(j in 1:503){
for(i in 1:4){
p[j,i]=exp(Betas[[j]][i+1]) / (1 + exp(Betas[[j]][2])+ exp(Betas[[j]][3])
+ exp(Betas[[j]][4]) + exp(Betas[[j]][5]))
}
}
Prob1=numeric()
for(i in 1:503)
{
Prob1[i]=1-sum(p[i,])
```

```
}

Probabilidades=cbind(Prob1,p)

#### Muestreo bootstrap
a=out$loglike[20000:35000]
class(a)
nsim=3000

muestra=matrix(NA,ncol=15000,nrow=nsim)

for(i in 1:nsim){
muestra[i,]= sample(a,15000,replace=TRUE)
}
q1=quantile(apply(muestra,1,mean),probs = c(0.025,0.975))

b=out$loglike[70000:85000]
class(a)
nsim=3000

muestra=matrix(NA,ncol=15000,nrow=nsim)

for(i in 1:nsim){
muestra[i,]= sample(b,15000,replace=TRUE)
}
q2=quantile(apply(muestra,1,mean),probs = c(0.025,0.975))
```

Agradecimientos

Primero que todo quiero agradecer a Dios por darme la oportunidad de acabar mis estudios de pregrado; también agradecer a mis padres por brindarme el apoyo y haber depositado la confianza en mi; de igual forma le doy las gracias a mis directores de tesis los cuales me acompañaron y me ayudaron en el proceso de la elaboración de mi trabajo de grado; también agradezco a todos los docentes con los cuales tuve la oportunidad de tener clase y tutorías y finalmente le agradezco a mis compañeros con los cuales compartí momentos únicos y especiales en los últimos 4 años.

José Giovany Babativa Marquez
Director

Dabogerto Bermúdez Rubio
Director

Recibido:
Aceptado:

Referencias

Agresti, A. (2002), *Categorical Data Analysis*, John Wiley and Sons.

Bartomowicz, T. & Bak, A. (2014), 'Maximum difference scaling method in the maxdiff r package', *KIT SCIENTIFIC PUBLISHING* .

Carioca, A. (2011), Modelos dinamicos hierarquicos espaco temporais para dados na familia exponencial, Master's thesis, Universidade Federal do Rio de Janeiro.

Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004), *Bayesian Data Analysis*, Texts in Statistical Science.

Lid, N., Holmes, C., Muller, P. & Walker, S. (2010), *Cambridge series in statistical and probabilistic mathematics*, Cambridge University Press.

Migon, H., Souza, G. & Schmidt, A. (2008), *Modelos Hierarquicos e Aplicacoes*, Brasileira de EstatÃstica.

Ntzoufras, I. (2009), *Bayesian Modeling Using WinBUGS*, Hoboken Wiley.

Nydick, S. (2012), The wishart and inverse wishart distributions.

Raghavarao, D., Wiley, J. & Chitturi, P. (2011), *Choice based conjoint analysis*, Taylor and Francis Group.

Rossi, P., Allenby, G. & McCulloch, R. (2005), *Bayesian and statistics*, John Wiley and Sons.

Van den Boogaart, K. & Tolosana, R. (2013), *Analysing Compositional Data with R*, Springer Heidelberg New York Dordrecht London.

Wasserman, L. (2006), *All of parametric Statistics*, Springer Texts in Statistics.

Wyner, G. (1995), Trade off techniques and marketing issues.