ENFOQUE BAYESIANO DEL MODELO DE REGRESION LOGISTICA USANDO CADENAS DE MARKOV MONTE CARLO

Lucio Díaz González*[1], Dante Covarrubias Melgar* y Vivian Sistachs Vega**
* Universidad Autónoma de Guerrero, Chilpancingo Guerrero, México.
** Universidad de La Habana, La Habana, Cuba.

ABSTRACT
The Logistic Regression is a highly used model on different areas of science where the response variable of studied problems is binary. This model can be studied under the Bayesian approach, nonetheless calculations might be complicated even using computation methods, is for that reason we use MCMC methods, which are iterative methods to obtain an approximation of the posterior distribution of model parameters. We use the R software to obtain the distribution with the MCMC method. On this work we implement the logistic regression through a simulation study under the Bayesian Approach applied to the cognitive health state on elders in the state of Guerrero.

KEYWORDS: Bayesian Logistic Regression, priori distribution, MCMC methods.

MSC: 62J12

RESUMEN

El modelo de Regresión Logística es un modelo muy utilizado en diferentes ramas de la ciencia donde se estudian problemas en los cuales la variable respuesta es binaria. Este modelo puede ser estudiado bajo el enfoque bayesiano, sin embargo, los cálculos pueden ser complicados incluso utilizando medios de cómputo, por esto es que se utilizarán los métodos MCMC, estos son un método iterativo para obtener de forma aproximada la distribución posterior de los parámetros en el modelo.
La obtención de la distribución con el método MCMC se realizará en el software R. En este trabajo se hace la implementación del modelo de regresión logística bajo el enfoque bayesiano aplicado el problema de salud del estado cognitivo de los adulto mayores del estado de Guerrero, a través de un estudio de simulación.

1. INTRODUCCIÓN

La regresión logística es uno de los métodos estadísticos más versátiles usados para modelar relaciones entre una variable dependiente dicotómica y varias variables independientes, estas pueden ser de cualquier naturaleza, convirtiéndose en un método estándar para el análisis de regresión cuando los datos son binarios (silva, 1994; Hosmer y Lemeshow, 2013).
El análisis de regresión logística es un método estándar para construir modelos de predicción para un evento binario y ha sido especialmente popular en investigaciones médicas, en las cuales la variable dependiente es que un paciente tenga o no una enfermedad (Santos, Mahola y Tachibana, 2009). En este sentido, uno de los objetivos de la regresión logística es expresar la probabilidad de que un paciente padezca cierta enfermedad; así como también describir la relación entre las variables respuesta y un conjunto de variables explicatorias (covariables) independientes.
Estudios previos basados en técnicas de regresión bayesiana para respuesta dicotómica han sido presentados en distintos campos de aplicación (Albert y Chib, 1993; Albert y Chib, 1995; Holmes y Held, 2006 y O'Hagan et al., 1991; entre otros). Chen et al. (1999) también aplican una estimación Bayesiana de los parámetros desde un punto de vista asimétrico, proponiendo una modelización que recoja la posibilidad de asimetría en la información. Otros autores han analizado el número de reclamaciones mediante la detección de las variables más importantes, como por ejemplo Lee et al. (2003), Richaudeau (1999) y Ordaz y Melgar (2010). Este último artículo incluye una base de datos española para la aplicación de un modelo probit.
La inferencia Bayesiana es un enfoque atractivo para la estimación de los parámetros del modelo de regresión logística (MRL) y en los últimos años ha crecido en popularidad, debido a que permite una fácil

[1] luciodiaz@uagro.mx

interpretación de las estimaciones de los parámetros del MRL y ofrece mejores resultados en muestras pequeñas. Sin embargo, para obtener las distribuciones de densidad a posterior de los parámetros del modelo es indispensable el uso de los métodos de aproximación, porque analíticamente no podemos especificar las densidades marginales a posterior. En este caso podemos usar Cadenas de Markov Monte Carlo (MCMC) o el método de Laplace (ver Santos, Mahola y Tachibana, 2009). En el presente trabajo utilizamos los métodos MCMC para encontrar las estimaciones para los parámetros del MRL en el estado cognitivo de los Adultos mayores en el estado de Guerrero.
El trabajo esta estructura de la siguiente manera: en el epígrafe 2 de presenta un resumen del modelo de regresión logística, en el epígrafe 3 se menciona los métodos MCMC y en el epígrafe 4 de describe el algoritmo metrópolis Hastings, así como los criterios de convergencia y la implementación en el software estadístico R y por último (epígrafe 4) se presenta la aplicación de los métodos MCMC al estado cognitivo de los adultos mayores.

2. EL MODELO DE REGRESIÓN LOGÍSTICA BAYESIANA

En regresión logística, cada resultado de la variable y_i (i= 1, …, n) sigue una función de probabilidad Bernoulli, toma el valor 1 con probabilidad π_i y 0 con probabilidad (1- π_i). De esta manera π_i varia sobre las observaciones como una función inversa del logaritmo de un vector x_i que incluye una constante y (k – 1) variables explicatorias. Para una muestra de n observaciones y_i, $i = 1, \ldots, n$ el modelo de regresión logística para datos binarios es especificada por

$$y_i|\pi_i \sim Ber(\pi_i)$$

$$\pi_i = P\{y_i = 1\} = \frac{exp(x_i^t\beta)}{1+exp(x_i^t\beta)}, \quad i = 1, \ldots, n$$

donde y_i=1 si si se observa la variable de interés para el i-ésimo individuo y y_i = 0 en cualquier otro caso, $\beta = (\beta_1, \beta_2, \ldots, \beta_k)^t$ es el vector de parámetros desconocidos y $x_i^t = (1, x_{i1}, x_{i2}, \ldots, x_{ik})^t$ es el vector de valores de las covariables para el i-ésimo individuo. La función de verosimilitud está dada por

$$L(\beta|\boldsymbol{y},\boldsymbol{X}) = \prod_{i=1}^{n} \left(\frac{exp(x_i^t\beta)}{1+exp(x_i^t\beta)}\right)^{y_i} \left(\frac{1}{1+exp(x_i^t\beta)}\right)^{1-y_i} \quad (1)$$

El estimador máximo verosímil β se obtiene maximizando el logaritmo de la función de verosimilitud (1). Esta estimación se obtiene utilizando el algoritmo de Newton-Raphson. Cuando usamos la estimación máximo verosímil, las inferencias sobre el modelo se basan usualmente en teoría asintótica. La distribución de $\hat{\beta}$ esta dada por

$$\hat{\beta} \sim N_{k+1}\left(\beta, I(\hat{\beta})^{-1}\right)$$

donde I$(\hat{\beta})$ es la matriz información de Fisher evaluada por $\hat{\beta}$. Jennings (1986) señala que se pueden encontrar regiones de confianza a partir de la función de verosimilitud en vez de la distribución asintótica de $\hat{\beta}$.
Los métodos basados en estimación máximo verosímil son a menudo usados en investigaciones prácticas. Sin embargo, es bien conocido que las estimaciones máximo verosímiles son sesgadas para modelos de regresión logística para muestras pequeñas (ver Griffiths, 1973), tal es el caso de muchos estudios clínicos y epidemiológicos, por lo tanto el sesgo podría ser relevante para la estimación de parámetros.
En el contexto bayesiano, la inferencia está basada por la información proporcionada por la distribución a posteriori de los parámetros β dada por

$$p(\beta|\boldsymbol{y},\boldsymbol{X}) = cf(\beta) \prod_{i=1}^{n} \prod_{i=1}^{n} \left(\frac{exp(x_i^t\beta)}{1+exp(x_i^t\beta)}\right)^{y_i} \left(\frac{1}{1+exp(x_i^t\beta)}\right)^{1-y_i} \quad (2)$$

donde $f(\beta)$ es una distribución a priori y c es la constant de normalización.
Si no existe información a priori sobre los parámetros o es dificil de obtener o formalizar, la incertidumbre inicial sobre los parámetros puede ser cuantificada usando una distribución a priori no informativa. Esto es lo mismo que incluir en el análisis la información que proviene de los datos.

Souza y Migon (2004) y Migon y Tachibana(1997) proponen normales independientes a priori, con precisión extremadamente pequeña, para las componentes de β. Estas a prioris son equivalentes a distribuciones a priori no informativas en esos parámetros, tratando a todos los valores como igualmente plausibles y están dados por

$$\beta_j \sim N(0{,}10000), \quad j = 1, \dots, k$$

Otra forma de obtener informaciones a posterior, es utilizan la información actual como información a priori y a partir de esta información calcular las distribuciones a posterior empleando métodos MCMC. En este caso particular aplicaremos el algoritmo de Metrópolis Hastings (Santos, Mahola y Tachibana, 2009).

3. MÉTODOS MCMC EN INFERENCIA BAYESIANA

Claramente podemos ver que la ecuación (5) es una function compleja de los parámetros, por lo tanto es necesario utilizar métodos numéricos para calcular la posterior marginal, momentos posteriores y densidades predictivas para cada uno de los parámetros del modelo. Se pueden obtener aproximaciones vía Métodos de Laplace (Tierney y Kadane, 1986) o métodos basados en Cadenas de Markov Monte Carlo (MCMC), para espacios paramétricos de dimensión alta. La aplicación de integración numérica puede ser no práctica y puede representarse pobremente.

3.1 El algoritmo Metrópolis Hastings

Los Métodos MCMC son algoritmos para el muestreo de distribuciones de probabilidad basados en la construcción de Cadenas de Markov. Estos muestreos MCMC se pueden usar para aproximar la distribución (por ejemplo generar un histograma), o calcular una integral (tal como un valor esperado o constante de normalización).

La aproximación MCMC muestra la distribución requerida construyendo una cadena de Markov aceptable para un tiempo largo.

Sea $g(\theta)$ la distribución de interés. Suponga al tiempo t, θ_{t+1} es elegido por el primer muestreo un candidato puntual η de una distribución propuesta $q(\cdot\,|\theta_t)$. El candidato η es entonces aceptado con probabilidad

$$\alpha(\theta, \eta) = min\left(1, \frac{g(\eta)q(\theta_t|\eta)}{g(\theta)q(\eta|\theta_t)}\right)$$

Si el candidato puntual es aceptado, el siguiente estado inicia $\theta_{t+1} = \eta$, si es rechazado, entonces $\theta_{t+1} = \theta_t$ y la cadena no se mueve. La distribución propuesta es arbitraria, y siempre y cuando la cadena sea irreducible y aperiodica, la distribución de equilibrio de la cadena podría ser $g(\theta)$ (Santos, et al., 2009) Para más detalles de MCMC en una variedad de modelos ver, Smith y Roberts (1993), Gelfand y Smith (1990) y Gilks et al. (1993), entre otros.

3.2. Algoritmo MCMC

En el presente apartado se describe la implementación MCMC usado en modelo de regresión logística. Se genera una muestra posterior por Monte Carlo simulando una cadena de Markov descrita como sigue:

1. Elegir los valores iniciales $\beta^0 = (\beta_0^0, \beta_1^0, \dots, \beta_k^0)$ una distancia razonable alejada de la moda a posterior.
2. Muestrear un Nuevo valor β^{i+1} de la distribución normal propuesta $N_{k+1}(\beta^i; I_F(\beta^i))$ donde $I_F(\beta^i)$ es la inversa de la matrix de Fisher de logaritmo de la posterior evaluada para β^i;
3. Así, la cadena es ejecutada para 25000 iteraciones y las primeras 5000 corridas son descartadas.
4. Después se obtiene una muestra del algoritmo MCMC para cada componente de β, es importante investigar temas tales como la convergencia para determinar si una muestra puede razonablemente ser tratada como un conjunto de realizaciones aleatorias de la distribución a posterior original. Mirando los gráficos de las trazas marginales es una forma simpe para examinar la salida formal de los procedimientos.

El método Metrópolis - Hasting tiene la ventaja de ser fácil de implementar. Una característica importante de este algoritmo es que, prácticamente no hay restricciones en la distribución a posterior.

3.2 Criterios de convergencia

Geweke propone un criterio basado en métodos de series de tiempo. Consiste en dividir la cadena en dos conjuntos, el primero contiene el primer x% de iteraciones y el segundo, el y% (en CODA es el 10% y 50%, respectivamente). Si la cadena en conjunto es estacionaria, las dos secuencias deben ser similares. Este método calcula las medias y varianzas asintóticas de estos dos segmentos por medio de la estimación espectral de densidades y construye un estadístico Z que es la diferencia entre las dos medias dividido por el error estándar de su diferencia. Conforme la longitud de la cadena crece, la distribución muestral de Z se aproxima a una N(0,1). Esto significa que valores de $|z| \leq 2$ indican que lasvariables has alcanzado la convergencia.

3.3. Aplicación del paquete MCMCpack en R

Para utilizar el paquete MCMCpack, es necesario activarlo mediante la instrucción **library(MCMCpack)**. Una vez activado podemos estimar la distribución a posterior de los parámetros mediante la instrucción **posterior=MCMClogit(modelo, data=datos)**, para mostrar los resultados utilizamos la instrucción **summary(posterior)**. Esta instrucción nos muestra los valores estimados de los parámetros con sus respectivas distribuciones a posterior , así como sus momentos.
Otro paquete muy útil en la simulación de estos modelos es CODA (Convergence diagnosis and Output Analysis) que nos permite obtener una serie de resultados como son representaciones gráficas de las distribuciones a posterior de los parámetros. El paquete CODA está diseñado para complementar y validar los resultados generados por los procedimientos MCMC. Permite el análisis estadístico, representación gráfica, aplicación de los criterios de convergencia existentes, estableciendo las regiones de mayor probabilidad para los parámetros (Geweke, J., 2005).
Para más detalles sobre la implementación de este algoritmo ver (Alonso, A. 2008, Santos, M. et al., 2009, Plummer M. 2012)

4. APLICACIÓN

En ésta sección presentamos la aplicación del modelo de regresión logística al estado cognitivo de los adultos mayores del estado de Guerrero. Se considera el algoritmo de Metrópolis Hastings para especificar las distribuciones a posterior de los parámetros del modelo de regresión logística. El conjunto de datos consiste en observaciones recolectadas en el 2005 en un estudio sobre el problema de salud de los adultos mayores en Guerrero. La muestra está compuesta por 592 registros y seis variables que se utilizaron como variables independientes. Estos datos han sido utilizados anteriormente por Díaz y Sistachs (2013).
Las seis variables son Sexo, tipo de población , estado de salud, Índice de masa corporal, Actividades Instrumentales de la vida diaria y estado anímico, donde estas fueron seleccionadas utilizando el criterio BMA en un estudio previo en la selección del modelo, representando potenciales factores de riego (Ver Díaz y Sistachs, 2013), ver tabla 1 para detalles.

Tabla 1. Caracterización de las variables en el estudio

Variable	Identificación	Codificación
Sexo	Sexo	0=Femenino, 1=Masculino
Población	Poblaci	0=Rural, 1=Urbana
Estado de salud	Es_sano	0=Enfermo, 1=Sano
Indice de masa corporal	Esca_imc	0=Normal, 1=Sobrepeso
Actividades instrumentales de la vida diaria	AIVD	0=Independiente, 1=Dependiente
Estado anímico	EEA	0=No afectado, 1=Afectado

El modelo de regresión logística fue ajustado usando el software R con el algoritmo MCMC generado por la función "MCMClogit". La densidad candidato generada es una densidad normal multivariada centrada en los valores actuales de beta con matriz de varianzas y covarianzas que es una aproximación de la densidad a posterior basada en las estimaciones máximo verosímiles y la precisión a priori multiplicada por la precisión del parámetro al cuadrado. Las cadenas son corridas para 25000 iteraciones y las primeras 5000 son descartadas.

4.1 Programa en R

En esta sección se presenta el programa en R utilizado en este trabajo

```
################## logistic regression##################
********ESTADO COGNITIVO DE LOS ADULTOS MAYORES*********
*******************************************************
library("MASS"); library(lattice);library("coda");library("MCMCpack")
library(splines); library("survival"); library(leaps)
datos<-read.table("base pension guerrero pocas variables2.txt", header=T)
datos
y<- datos$EEC; x<- data.frame(datos[,-6])
*******************************************************************
Codigo para la simulación, suponiendo que tenemos  prior impropia uniforme
*******************************************************************
posterior<-MCMClogit(EEC~as.factor(poblaci)+
as.factor(es_sano)+as.factor(EEA), data=datos,burnin=1000,mcmc=25000)
summary(posterior)
plot(posterior,trace=FALSE)
plot(posterior, density=FALSE)
autocorr.plot(posterior)
geweke.diag(posterior,frac1=0.1, frac2=0.5)
geweke.plot(posterior,pvalue=0.05)
*********************************************************************
Codigo para la simulación, suponiendo que tenemos  prior normal multivariada
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
posterior2<-MCMClogit(EEC~as.factor(poblaci)+as.factor(es_sano)+as.factor(EEA), b0=0, B0=.001,
data=datos,burnin=1000,mcmc=25000)
summary(posterior)
plot(posterior2,trace=FALSE)
plot(posterior2, density=FALSE)
autocorr.plot(posterior2)
geweke.diag(posterior2)
geweke.plot(posterior2)
```

5. RESULTADOS

Así, se ejecuta el muestreo de Metrópolis Hastings, las salida de la cadena a lo largo de las iteraciones para los parámetros del modelo de regresión logística se presenta en la figuras 1.
Un análisis formal para el diagnóstico de convergencia de la cadena se hace a través del gráfico en la figura 1. Note que la serie de tiempo presenta estacionalidad y la cadena se puede considerar como buena mezcla indicando que la convergencia se alcanzó. Otra fácil e intuitiva herramienta de diagnóstico implementada es el gráfico de autocorrelaciones que se muestra en la figura 2.
En el gráfico 2. Se puede apreciar que existe convergencia en todas y cada una de las variables incluidas en el estudio, por tanto esto nos indica que la nuestra que se esta usando para realizar el análisis es adecuada, también podemos observar que la convergencia es lenta.
Hay también muchos métodos de diagnóstico de convergencia formal basados en propiedades estadísticas de la cadena de Markov. Uno de los más populares fue propuesto por Geweke (1992) basado en la prueba de igualdad de medias de la primera y última parte de la cadena de Markov (por default es el primer 10% y último 50%). Si las dos muestras se obtuvieron de una distribución estacionario, las dos medias con iguales y el estadístico de Geweke tiene una distribución normal estándar asintóticamente. Por lo tanto valores de Z que caen en los extremos de la distribución normal estándar indica que la cadena no converge completamente. Esta aproximación está disponible en el software R con la paquete CODA (Plummer et al., 2005). Seleccionado la prueba de Geweke, los resultados se muestran en la Tabla 2. Los resultados para la cadena proporcionan evidencia de la convergencia de cada una de las variables en el modelo para valores de 1.96.

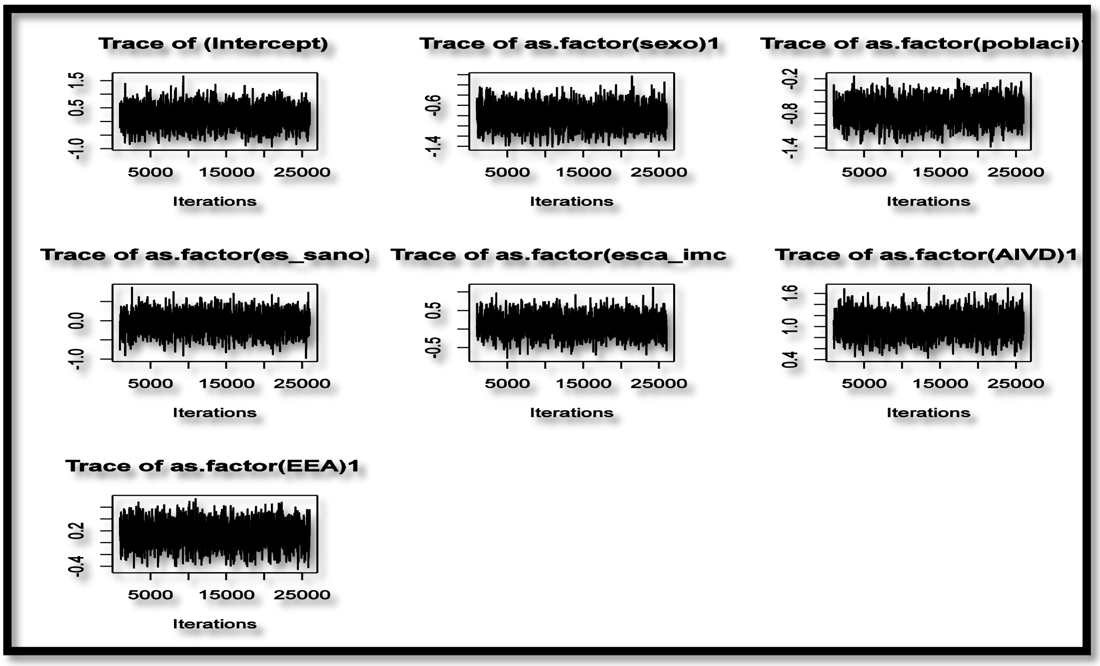

Figura 1. Gráfico de las trazas para los parámetros del modelo

Tabla 2. Diagnóstico de convergencia de Geweke

Variable	Z-score No informativa Impropia	Z-score Normal multivariada
Constante	-0.2235	0.1399
Sexo	1.9803	1.6540
Tipo de población	-2.0598	-2.0894
Estado de salud	-.05584	-0.9026
IMC (índice de masa corporal)	0.4507	-0.1904
AIVD (Actividades Instrumentales de la vida diaria)	0.4131	1.1927
EEA (Estado anímico)	0.7549	1.1275

En la tabla 3. Se muestran las estimaciones para los parámetros con las diferentes a prioris consideradas y podemos observar que los resultados son muy similares. También en la tabla 4 se muestran los intervalos de credibilidad para dichos parámetros.

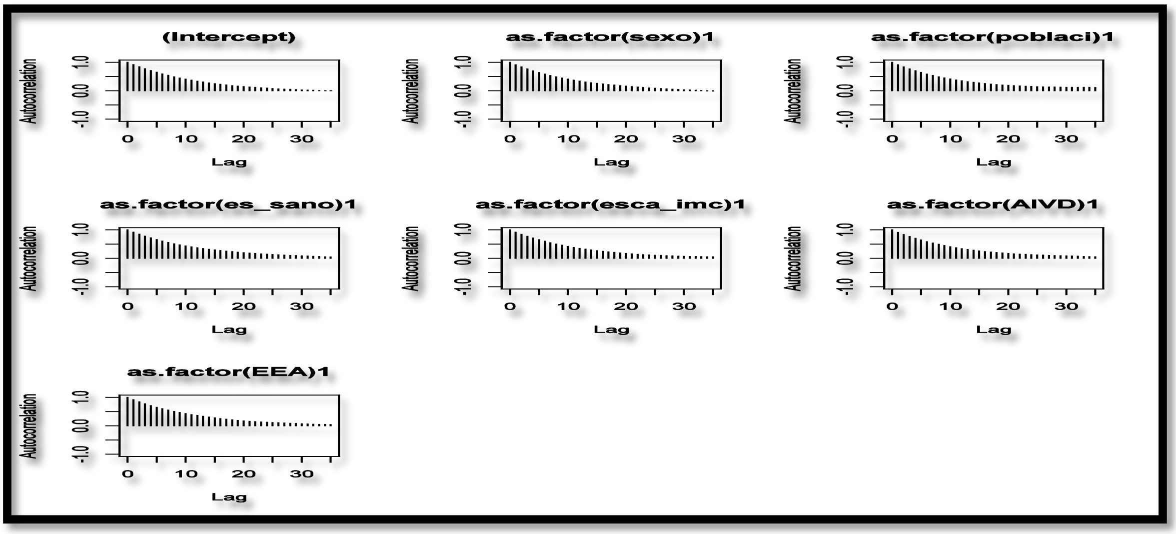

Figura 2. Gráfico de autocorrelaciones de las cadenas MCMC de los parámetros

Tabla3. Estimaciones para los parámetros

	No informativa impropia		Normal multivariada	
	Mean	SE	Mean	SE
(Intercept)	0.17925	0.010351	0.17213	0.010237
as.factor(sexo)1	-0.80675	0.005793	-0.80211	0.005771
as.factor(poblaci)1	-0.75893	0.005420	-0.75663	0.005198
as.factor(es_sano)1	-0.05592	0.007389	-0.05383	0.007675
as.factor(esca_imc)1	0.10820	0.007388	0.11370	0.007003
as.factor(AIVD)1	1.03622	0.005589	1.02996	0.005742
as.factor(EEA)1	0.14042	0.005414	0.13816	0.005233

En la Tabla 4, además de mostrar los intervalos de credibilidad, nos permite determinar cuáles variables son significativas para nuestro modelo, de acuerdo con Geweke (2005), aquellas variables que en su intervalo contengan al 0 no son importantes para el modelo que se está estudiando, en este sentido podemos observar que las variables que cumplen esta característica son: El estado de salud de los adultos mayores, Índice de masa corporal y el estado anímico de los adultos mayores.

Tabla 4. Intervalos de confianza para los modelos de regresión logística

	No informativa impropia		Normal multivariada	
	2.5%	97.5%	2.5%	97.5%
(Intercept)	-0.4857	0.8578	-0.4752	0.8335
Sexo	-1.1812	-0.4310	-1.1849	-0.4181
Tipo de población	-1.1190	-0.4143	-1.0895	-0.4051
Estado de salud	-0.5405 *	0.3918 *	-0.5369 *	0.4235 *
Indice de masa corporal	-0.3630 *	0.5828 *	-0.3470 *	0.5782 *
Actividades instrumentals de la vida diaria	0.6597	1.3972	0.6605	1.4071
Estado Animico	-0.2141 *	0.4838 *	-0.2048 *	0.4825 *

5. CONCLUSIONES

En este artículo mostramos que el modelo de regresión logística es muy útil y funciona muy bien para predecir, basado en algunas covariables relevantes en el problema del estado cognitivo de los adultos mayores en el estado de Guerrero. El uso de métodos MCMC se justifica porque la integración analítica de la distribución conjunta para los parámetros de regresión β es intratable. El desarrollo bayesiano permite usar información adicional que proporcionan estimaciones más confiables.
Nosotros mostramos en este caso particular que no importa cuál sea la distribución a priori que se tome para modelar los datos del estado cognitivo de los adultos mayores en el estado de Guerrero, los resultados son iguales al momento de aplicar los métodos MCMC.

RECEIVED DECEMBER, 2014
REVISED FEBRUARY, 2015

REFERENCIAS

[1] ALBERT, J. H. & CHIB, S. (1993): Bayesian analysis of binary and polychotomous response data. **Journal of the American Statistical Association**, 88, 669-679.
[2] ALBERT, J. H. & CHIB, S. (1995): Bayesian residual analysis for binary response regression models**. Biometrika** ,.82,747-759.
[3] ALONSO, A., (2008): El resurgir de Thomas Bayes. **Anuario jurídico y económico escurialense,**. 327-360.
[4] BROOKS, S.; GELMAN, A.; JONES, G. and MENG, X. (2011): **Handbook of Markov Chain Monte Carlo**. A Chapman & Hall/CRC.
[5] CHEN, M. H.; IBRAHIM, J.; & YIANNOUTSOS, C. (1999): Prior elicitation, variable selection, and Bayesian computation for logistic regression models. **Journal of the Royal Statistical Society**, series B, 61, 223-242.

[6] DIAZ, L., SISTACHS, V. (2013): Selección de Modelos bajo el enfoquen bayesiano una aplicación al estado cognitivo de los adultos mayores en el estado de Guerrero. En **Modelación Matemática de Fenómenos del medio ambiente y la salud**, tomo 3, (Ed. C. Bouza, et al,): UGR, Granada118-127.
[7] HOLMES, C., HELD, K. (2006): Bayesian auxiliary variable models for binary and multinomial regression. **Bayesian analysis**, 1, 145-168.
[8] HOSMER, D. W.; LEMESHOW, S. (2013): **Applied logistic regression**. New York: John Wiley and Sons, N. York.
[9] GELFAND , A. E.; SMITH, A. F. M. (1990): Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**.85, .398-409.
[10] GEWEKE, J. (2005):. **Contemporary Bayesian Econometrics and Statistics**, Wiley, Hoboken, NJ.
[11] GEWEKE, J. (1992): Evaluating the accuracy of sampling based approaches to calculation of posterior moments. In **Bayesian Statistics**, vol. 4 (eds Bernardo J. M., Berger J. O., Dawid A. P. and Smith A. F.):. Clarendon Press, Oxford..
[12] GILK S, W. R.; CLAY TON, D . G.; SPIEGELHALTER, D . J.; BEST, N. G.; McNEIL, A. J.; SHARPLES, L. D .; K IRBY , A. J. (1993): Modelling complexity: applications of Gibbs sampling in medicine, **J . R. Stat. Soc** . Ser. B , 55, 39-52.
[13] GRIFFITHS, D. A. (1973): Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of disease. **Biometrics**, 29, 637-649.
[14] JEAN, M. M.; CHRISTIAN, R., (2013): **Bayesian essentials with R**, Second edition. Springer, Berlin.
[15] JENNINGS, D. E. (1986): Outliers and residual distributions in logistic regression. **Journal of the American Statistical Association,** 81, 987-990.
[16] LEE, K.; SHA, N.; DOUGHERTY, R.; VANNUCCI, M.; MALLIK, K., (2003): Gene selection: a Bayesian variable selection approach. **Bioinformatics**, 19, 90-97.
[17] MIGON, H. S.; TACHIBANA, V. M. (1997): Bayesian approximations in randomized response models. **Comp. Stat. Data Anal.** 24, 401-409.
[18] MELGAR, M.C.; ORDAZ, J. A. (2010): Covariate-based pricing of automobile insurance. **Insurance markets and companies: analyses and actuarial computations,** 1, 92-99.
[19] O'HAGAN, A.; WOOD WARD, E. G.; MOOD ALEY, L. C. (1991): Practical bayesian analysis of simple logistic regression: predicting corneal transplants. **Stat. Med**., 9, 1091-1101.
[20] PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. (2012): Output analysis and diagnostics for MCMC. The Coda package. Software R. Disponible en: http://cran.es.r-project.org/.
[21] RICHAUDEAU, D. (1999): Automobile insurance contracts and risk of accident: An empirical test using French individual data. **Geneva papers on risk and insurance theory**, 24, 97-114.
[22] ROBERT, C. P., CASELLA, G. (2010): **Introducing Monte Carlo Methods with R**, Springer, Berlin.
[23] SANTOS, M. A., MAOLA, F. A., TACHIBANA, V. M., (2009): Approximate Bayesian methods for logistic regression model. **Rev. Bras. Biom**., 27,288-300.
[24] SMITH, A. F. M.; ROBERTS, G. O. (1993):Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion):. *J. R. Stat. Soc*. Ser. B , 55, 3-24.
[25] SILVA, L.C. (1994): **Excursión a la Regresión Logística en la Ciencia de la Salud**. Ediciones DIAZ DE SANTOS, S.A., ESPAÑA.
[26] SOUZA, A. D. P.; MIGON, H. S. (2004): Bayesian binary regression model: An application to in-hospital death after AMI prediction. **Pesq. Ope**r, 24, 235-267.
[27] TIERNEY, L.; KADANE, J. B. (1986): Accurate approximations for the posterior moments and marginal densities. **J. Am. Stat. Assoc**., .81, 82-86.