REVISTA INVESTIGACION OPERACIONAL VOL. 37, NO. 1, 103-111, 2016

Applications/Aplicaciones

UTILIZACIÓN DEL BMA EN EL MODELO DE REGRESIÓN LOGÍSTICA Y SU COMPARACIÓN CON OTROS CRITERIOS DE SELECCIÓN DE MODELOS

Lucio Díaz González*, Vivian del Rosario Sistachs Vega**, Dante Covarrubias Melgar*,
Ma. del Carmen Cruz Velázquez***, Imelda S. Hernández Nava***, Martha L. Sánchez Castillo***
*Unidad Académica de Matemática, Universidad Autónoma de Guerrero, México.
**Facultad de Matemática y Computación. Universidad de La Habana, Cuba.
***Unidad Académica de Enfermería No. 1, Universidad Autónoma de Guerrero, México.

ABSTRACT
The identification of risk factors associated with a disease is an important part on the model construction for public health. The investigators frequently use Logistic Regression (LR) as a statistical tool to achieve such end. LR it is commonly used with stepwise selection methods to select risk factors associated with a particular disease. Nonetheless stepwise methods for model selection tend to select over estimated models. This happens because standard selection methods don't take into account the inherent uncertainty on model selection. A criteria that takes into account the model selection uncertainty is the Bayesian Model Average (BMA) recently applied on LR for case and study controls. This paper presents the application utility of the BMA criteria and illustrates its use in a study about the presence of preeclampsia during pregnancy on Guerrero Mexico, said application is performed on the R software it also compares the results obtained with other selection criteria.

KEYWORDS: BMA, model selection, logistic regression.

MSC: 62P10

RESUMEN:
La identificación de factores de riesgo asociados con una enfermedad es una parte muy importante en la construción de modelos en salud pública. Una de las herramientas estadisticas de uso frecuente que los investigadores utilizan para dicho fin es la Regresión Logística (RL). Los investigadores comunmente usan la RL con métodos de selección paso a paso (stepwise) para la selección de factores de riesgo asociados a una enfermedad particular. Sin embargo los métodos paso a paso para la selección de modelos tienden a seleccionar modelos sobreestimados. Esto es por que los métodos de selección estándar no toman en cuenta la incertidumbre inherente en la selección de los modelos. Un criterio que toma en cuenta la incertidumbre en la selección de los modelos es el Promedio Bayesiano de Modelos (BMA), recientemente aplicado a Regresión Logística para estudios de casos y controles. En este trabajo se presenta la utilidad de aplicar el criterio BMA y se ilustra el uso del mismo en un estudio sobre la presencia de preeclampsia durante el embarazo en Guerrero, México, dicha aplicación se realiza en el software R, también se comparan los resultados obtenidos con otros criterios de selección.

1. INTRODUCCIÓN

Muy relacionado con el tema de la incertidumbre está el problema de la selección de modelos, que tiene dos aspectos uno referido a si el modelo es adecuado y el otro, a ¿Cuál es el mejor modelo?, dentro de una colección bajo consideración (Gelfand y Dey, 1994; Diaz et al. 2014)
Se define un modelo como una especificación de una distribución de observables (los datos) y no observables (los parámetros del modelo, observaciones perdidas, etc.) y esta definición puede ser enfocada desde una perspectiva bayesiana.
En el enfoque bayesiano los parámetros y los modelos son considerados aleatorios ($f(y/M_i)$) y expresan su incertidumbre en términos de distribución de probabilidad. Entre los diferentes criterios bayesianos de selección de modelos están los Factores de Bayes (FB), como un criterio para seleccionar entre dos posibles modelos y para el caso general (más de 2 modelos) se utiliza el criterio BMA, el cual permite promediar los

modelos (ver Claeskens, G. and Hjort, N. L.2008, Ando,T. 2010), también existen otros criterios como el AIC, BIC, DIC, etc.(ver Kadane and Lazar 2004, Foster and Sober, 2011).
En el presente trabajo se muestra la utilidad del criterio BMA en la selección de modelos para ello, se presenta en el epígrafe 2 el uso del criterio BMA en la selección de modelos bajo el paradigma bayesiano, además se comentan otros criterios como el AIC, BIC y DIC. En el epígrafe 3 se ilustra la aplicación de este criterio de selección en un estudio sobre la presencia de preeclampsia durante el embarazo en Guerrero, México, también en este epígrafe se comparan los resultados de BMA con otros criterios de selección de modelos.

2. CRITERIO BMA PARA LA SELECCIÓN DE MODELOS BAJO EL PARADIGMA BAYESIANO

El enfoque bayesiano como ya se ha planteado permite expresar la incertidumbre en términos de probabilidad, y bastan las reglas básicas del cálculo de probabilidades para poder hacer inferencias y el BMA (Bayesian Model Average) no es mas que estadística bayesiana básica (ver Ando, T. 2010, Perra, S. 2013).
La probabilidad final BMA de θ viene dada por:

$$p(\theta|D) = \sum_{k=1}^{K} p\,(\theta|D, M_k)p(M_k|D) \tag{1}$$

Siendo $p(\theta|\mathrm{D}, \mathrm{M_k})$ distribución de probabilidad final de θ, dado el modelo M_k y los datos D, y $p(M_k|D)$la distribución de probabilidad final de M_k, tomado como el modelo verdadero, considerando que uno de los modelos propuestos es el verdadero.
La probabilidad final del modelo M_k, esta dada por:

$$p(M_k|D) = \frac{p(D|M_k)p(M_k)}{\sum_{i-1}^{K} p(D|M_i)p(M_i)} \tag{2}$$

En esta expresión, $p(M_k|D)$ es la integral de la función de verosimilitud del modelo M_k, resultado de integrar sobre los parámetros del modelo, es decir:

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)\, d\theta_k \tag{3}$$

Siendo θ_k él o los parámetros del modelo M_k y $p(D|\theta_k, M_k)$ la función de verosimilitud de θ_k para el modelo M_k , y $p(\theta_k|M_k)$ la probabilidad inicial de θ_k . Las probabilidades iniciales suelen considerarse iguales.

2.1 Criterio de Información de Akaike

El criterio básico basado en la información estadística, es el criterio de información de Akaike, que fue introducido por Akaike, 1973. Su idea clave es la de penalizar un exceso de parámetros ajustados. El AIC (Akaike Information Criterion) es un estimador muestral de $E[ln\, f(X|\theta)]$, esperanza de la log-verosimilitud, que viene dado por la expresión general

$$AIC(k) = -2ln\mathcal{L}\big[\hat{\theta}(k)\big] + 2k \tag{4}$$

en donde $\mathcal{L}[\theta(k)]$, es la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es el estimador máximo verosímil del vector de parámetros θ y k es el número de parámetros independientes en el modelo, mientras "ln" denota el logaritmo neperiano. El primer término del AIC puede ser interpretado como una medida de la bondad de ajuste, mientras el segundo término es una penalización, que crece conforme aumenta el número de parámetros, de acuerdo al principio de parsimonia.
El AIC enfatiza la bondad de ajuste del modelo. Como puntualizaba Takane (1987), el AIC no pretende identificar el modelo verdadero. El hecho de que un modelo se ajuste mejor a los datos, no quiere decir que sea el modelo real o verdadero. Más bien significa que el modelo es el mejor de entre los modelos candidatos, en el sentido que proporciona la aproximación más cercana a la realidad. Algunas de las ventajas del AIC que lo hace tan utilizado en la práctica, son su simplicidad (no es necesario el uso de tablas para observar un determinado valor) y facilidad para ser implementado, y el hecho de que no existe el problema de especificar subjetivamente un nivel de significación arbitrario para contrastar dos modelos.

2.1.1 Selección de modelos utilizando los pesos de Akaike

Wagenmaker y Farrell(2004) demuestran que los valores del AIC para cada modelo pueden ser transformados a las llamadas ponderaciones o pesos de Akaike (Akaike, 1978, 1979; Bozdogan, 1987, Burnham y Anderson, 2002), los cuales pueden ser interpretados como probabilidades condicionales para cada modelo.
En primer lugar se calcula, para cada modelo, la diferencia en el valor AIC con respecto al AIC del candidato a mejor modelo (el que minimiza el valor AIC), es decir

$$\triangle_i\ (AIC) = AIC_i - mín(AIC) \tag{5}$$

donde i representa al correspondiente modelo.
Dado que el AIC es definido para ser un estimador asintóticamente insesgado de menos dos veces la log-verosimilitud de un modelo determinado mediante las estimaciones de máxima verosimilitud de los parámetros, Akaike (1978, 1979) sugiere que $exp\left\{\frac{1}{2}AIC\right\}$ será asintóticamente una definición razonable para la verosimilitud de un modelo especificado por los parámetros determinados mediante el método de máxima verosimilitud. De esta manera, a partir de las diferencias en el AIC dadas por la ecuación 5, se puede obtener una estimación de la verosimilitud relativa al modelo M_i, puesto que

$$\mathcal{L}(M_i|datos) \propto exp\left\{-\frac{1}{2}\Delta_i(AIC)\right\},$$

donde $\propto$ indica proporcionalidad entre los dos extremos de la expresión.
Finalmente, las verosimilitudes relativas a cada modelo son normalizadas (es decir, divididas por la suma de las verosimilitudes de todos los modelos) para obtener los pesos de Akaike, $w_i(AIC)$ (ver Burnham y Anderson, 2002), cuyo valor esta dado por

$$w_i(AIC) = \frac{exp\left\{-\frac{1}{2}\Delta_i(AIC)\right\}}{\sum_{k=1}^{K} exp\left\{-\frac{1}{2}\Delta_k(AIC)\right\}}, \tag{6}$$

donde $M_i,\ i = 1, \dots, K$, son los modelos candidatos, de forma tal que $\sum_{k=1}^{K} w_i(AIC) = 1$.
Los pesos $w_i(AIC)$ dados por la expresión 6 pueden ser interpretados como la probabilidad de que M_i sea el mejor modelo (es decir, el que minimiza la discrepancia de Kullback-Leibler), dados los datos y el conjunto de modelos candidatos, siendo por tanto estos pesos medidas de evidencia a favor de un modelo concreto (Burnham y Anderson, 2002).
El nivel de evidencia en favor de un modelo respecto a otro, es obtenido mediante la división de sus respectivos pesos de Akaike. La razón de evidencia de los pesos de Akaike de un modelo A sobre un modelo B puede ser calculada mediante la expresión (Wagenmakers y Farrell, 2004)

$$\frac{w_A(AIC)}{w_B(AIC)} = exp(ln\mathcal{L}_A - ln\mathcal{L}_B + k_A - k_B) = \frac{\mathcal{L}_A}{\mathcal{L}_B}exp(k_A - k_B),$$

donde $\mathcal{L}_A$ y $\mathcal{L}_B$ son, respectivamente las verosimilitudes de los modelos A y B, mientras que k_A y k_B son, respectivamente, el número de parámetros libres estimados dentro de los modelos A y B. Así, dados dos modelos , A y B con $w_A(AIC) = r_1$ y $w_B(AIC) = r_2$, y $r_1 > r_2$, se dice que la evidencia de que el modelo A sea mejor que B es $\frac{r_1}{r_2}$ veces favorable al modelo A, en términos de la discrepancia de Kullback-Leibler.

2.2 Criterio de Información de Schwarz (BIC)

Observe que el término de penalización del AIC en la ecuación (4), $2k$, no depende del tamaño de muestra n de la población considerada. Esto conduce al hecho de que un mismo número de parámetros comunes es seleccionado mediante el AIC, tanto para muestras pequeñas como para muestras grandes. Es por ello que el AIC no es un estimador consistente del número de parámetros, aunque habría que decir también que la consistencia es una propiedad asintótica y cualquier problema real tendrá un tamaño de muestra finito n.
En el contexto de procedimientos basados en la verosimilitud, Schwarz (1978) sugirió que el AIC podría no ser asintóticamente justificable, y presentó un criterio de información alternativo a partir de un enfoque bayesiano, el BIC (*Bayesian Information Criterion*). Con este criterio, se penaliza el número de parámetros con $ln\,n$, en lugar de 2. Así,

$$BIC(k) = -2ln\mathcal{L}\left[\hat{\theta}(k)\right] + (ln\, n)k, \quad (7)$$

donde $\mathcal{L}[\theta(k)]$ es la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es el estimador máximo verosímil del vector de parámetros y k es el número de parámetros independientes en el modelo, mientras que n es el tamaño muestral.

El criterio es obtenido analizando el comportamiento de la probabilidad a *posteriori* del *j*-ésimo modelo cuando n tiende a infinito, bajo la hipótesis de algunas distribuciones a *priori* de los parámetros.

$$lnf\left(X\middle|M_j\right) = \mathcal{L}_j\left(\hat{\theta}_j\middle|X\right) + lnP\left(\hat{\theta}_j\middle|M_j\right) + \frac{k_j}{2}ln(2\pi) - \frac{k_j}{2}ln\, n + \frac{1}{2}ln\left|R_j\right|,$$

donde

- $f\left(X\middle|M_j\right)$ es la verosimilitud marginal de los datos para el modelo M_j
- $\mathcal{L}_j\left(\hat{\theta}_j\middle|X\right)$ es la verosimilitud del modelo M_j, donde $\hat{\theta}_j$ es el estimador máximo verosímil del parámetro θ_j.
- $P\left(\hat{\theta}_j\middle|M_j\right)$ es la probabilidad *a priori* para los parámetros.
- k_j es el número de parámetros independientes estimados dentro del modelo M_j.
- $\left|R_j\right|$ es el determinante de $R_j = n\, S_j$, donde S_j es la matriz de covarianzas del estimador $\hat{\theta}_j$.

Para un n suficientemente grande,

$$\ln f\left(X\middle|M_j\right) = \mathcal{L}_j\left(\hat{\theta}_j\middle|X\right) - \frac{k_j}{2}\ln n \quad (8)$$

La expresión (8) fue obtenida por vez primera por Schwarz (1978), que propuso escoger el modelo que conduzca a un valor máximo de esta cantidad. Una forma equivalente de este criterio es calcular para cada modelo la cantidad

$$BIC\left(M_j\right) = -2\ln\mathcal{L}_j\left(\hat{\theta}_j\middle|X\right) + k_j\ln n,$$

y seleccionar aquel modelo para el cual esta cantidad es mínima. De esta manera, este criterio pondera la desviación del modelo, medida por $-2\ln\mathcal{L}_j\left(\hat{\theta}_j\middle|X\right)$, con el número de parámetros. Si introducimos más parámetros en el modelo, mejorará el ajuste, con lo que aumentará el soporte o disminuirá la desviación que aparece en $k_j\ln n$.

De esta manera, los términos principales del desarrollo asintótico conforman un criterio válido para muestras grandes, más allá del contexto bayesiano, dado que ellos no dependen de la distribución *a priori*.

Este criterio, a diferencia del AIC, considera el número n de observaciones en el término de penalización. El valor óptimo de k se calcula minimizando $BIC(k)$ en la ecuación (7).

2.3 Criterio de información de Deviance (DIC)

El DIC es otro criterio de selección de modelos que es comúnmente clasificado en la familia bayesiana. Algunos criterios de información, tal como el AIC y BIC, expresan el ajuste y la complejidad de un modelo. Esto es difícil de lograr en un contexto especifico que algunas veces encontramos en la práctica, por ejemplo modelos jerárquicos o modelos complejos (ver Henderson et al. 2010). Esto es cuando nosotros representamos la marginal de los datos x en un modelo de probabilidad como

$$p(x) = \int_{\theta\in\Theta} p(x|\theta)p(\theta)d\theta \quad (9)$$

con parámetro θ y densidad *a priori* $p(\theta)$, podemos algunas veces elegir representar que la distribución *a priori* es dominada por un hiperparámetro Ψ.

$$P(\theta) = \int_{\psi\in\Psi} P(\theta|\psi)P(\psi)d\psi$$

Sin embargo existe un inconveniente sobre cual podría considerarse como la función de verosimilitud de los datos: $P(x|\theta,\psi), P(x|\theta), P(x|\psi)$ (ver Bayarri et. al, 1988).

El DIC mide la complejidad del modelo en términos de sus propiedades de estimación, la complejidad del modelo se refiere al número efectivo de parámetros en el modelo. Desde el punto de vista bayesiano el DIC,

consiste en incorporar la información *a priori* de los parámetros: "parece razonable que una medida de complejidad pueda depender de la información *a priori* relacionada con los datos específicos que son observados" (Spiegelhalter et. al, 2002).
Para hacer explicito esto, Spiegelhalter et. al (2002) proponen medir la complejidad comparando la devianza esperada en los datos (basado en la distribución *a posteriori*) con la devianza en el estimador $\tilde{\theta}(x)$ que podríamos usar. Esto podría dar una idea de la dificultad en la estimación. Así necesitamos medir la devianza en los datos x relativa a la hipótesis puntual θ. La medida canónica de devianza entre los datos x y el modelo es $-\log p(x|\theta)$ (Good, 1952; Bernardo, 1979). La devianza se define como

$$-\log p(x|\theta) = \sum_i \log p(x_i|\theta)$$

Regresando a la comparación de las devianzas, $\tilde{\theta}(x)$ denota un estimador bayesiano estándar de nuestra cantidad de interés θ, concretamente la media *a posteriori* de θ. Entonces podemos comparar la devianza en los datos (distribución *a posteriori* condicional de θ) con la devianza observada bajo nuestra estimación estándar de θ. Esta cantidad P_D muestra la dificulta de ajustar eficientemente los parámetros de un modelo M_θ.

$$\begin{aligned} P_D &= \mathrm{E}_{\theta|x}[-2\log p(x|\theta)] - 2\left(-\log p\left(x\middle|\tilde{\theta}(x)\right)\right) \\ &= 2\log p\left(x\middle|\tilde{\theta}(x)\right) - 2\int_{\theta\in\Theta} \log p(x|\theta)p(\theta|x)d\theta \end{aligned} \quad (10)$$

De hecho, P_D se ha usado para evaluar valores de modelos candidatos, y sirve como base para el Criterio de Información de Devianza (DIC), el DIC es definido como

$$DIC(M_\theta, x) = \mathrm{E}[D(\theta, x)] + p_D(M_\theta, x) \quad (11)$$

Donde la función $D(\cdot,\cdot)$ es definida como

$$D(\theta, x) = -2\log p(x|\theta) + 2\log f(x) \quad (12)$$

Para alguna función estandarizada $f(x)$. El valor del DIC en la ecuación (11) es la suma de la bondad de ajuste $\mathrm{E}[D(\theta, x)]$ y la complejidad del modelo $p_D(M_\theta, x)$ (para más detalles sobre el DIC del Spiegelhalter et al., 2002).

3. RESULTADOS Y DISCUSIÓN DE ESTUDIO SOBRE LA PRESENCIA DE PREECLAMPSIA DURANTE EL EMBARAZO EN GUERRERO, MÉXICO

La preeclampsia es una complicación médica del embarazo también llamada toxemia del embarazo y se asocia a hipertensión inducida durante el embarazo; está asociada a elevados niveles de proteína en la orina (proteinuria). Dependiendo de las cifras de presión arterial y la pérdida de proteínas en la orina, la preeclampsia puede ser leve, moderada o severa, sin embargo, esta clasificación puede a veces ser peligrosa, ya que en algunas oportunidades unas pacientes con preeclampsia clasificada como leve puede pasar rápidamente a ser severa, incluso a presentar convulsiones(eclampsia).
En el Estado de Guerrero durante el año 2005 por un informe de la Secretaria de Salud se manifestó que una mujer muere cada 4 días por complicaciones del embarazo, parto y puerperio, ubicándose en el primer lugar por mortalidad materna a nivel nacional.
Es por ello que se analizó una población de 4692 expedientes clínicos, donde se utilizó un muestreo no probabilístico por conveniencia y se seleccionaron 404 expedientes clínicos de gestantes atendidas durante el 2008 con el diagnóstico de enfermedades hipertensión en el Hospital de la Madre y el Niño Guerrerense (HMNG), luego la variable respuesta que se consideró fue Y= preeclampsia (donde la presencia de preeclampsia (1) y la ausencia fue (0)).
Las variables que se analizaron en este estudio son las siguientes, ya que están consideradas como factores de riesgo para padecer preeclampsia. Esta información se obtuvo de los expedientes clínicos:

1. (P1) Edad cumplida
2. (P2) Estado civil (soltera (0), casada (1))
3. (P6) Ocupación (hogar (1) y otro (0).)
4. (P11) Ha tenido abortos (si (1), no (0))
5. (P12) Ha tenido embarazos múltiples (si (1), no (0))
6. (P13) Semanas de gestación
7. (P14) Tuvo atención prenatal durante su embarazo (si (1), no (0))
8. (P17) Que antecedentes patológicos presenta (si presentó (1) y no presentó (0))

9. (P18) Qué tipo de antecedentes hereditarios familiares tiene (tiene HTA (1) y o no (0))
10. (P19) Nivel socioeconómico (bajo (1), otro (0))

El modelo que se utilizó fue el de regresión logística binaria para analizar la probabilidad de presencia de preeclampsia, que es una enfermedad producto de la hipertensión arterial, a partir de las variables que se consideraron factores de riesgo.
Bajo el paradigma Bayesiano las distribuciones *a priori* de los parámetros vector β y σ^2 *q*ue se consideraron fueron las no informativas, es decir $p(\beta,\tau) \propto \tau^{-1}$, donde $\tau = \frac{1}{\sigma^2}$
La distribución posterior del modelo con esas *a priori* seria

$$p(\beta,\tau / z) = N_p\left(\beta / \hat{\beta}_w, \tau^{-1}\left(X^tWX\right)^{-1}\right)Ga\left(\tau / \frac{n-p}{2}, \frac{n-p}{2}\hat{\sigma}^2\right)$$

donde $\hat{\beta}_w = \left(X^tWX\right)^{-1}X^tWz$, $\hat{\sigma}^2 = \frac{1}{n}\left(z - X\hat{\beta}_w\right)^t\left(z - X\hat{\beta}_w\right)$ y W matriz diagonal donde

$w_{ii} = \hat{\pi}_i\left(1-\hat{\pi}_i\right)$

Para resolver el problema se utilizó el paquete BMA (Raftery, et al 2010; Spencer, 2011) que esta en el lenguaje R (R Development core tema, 2010), y cuyas instrucciones para correr el BMA en este modelo aparecen en el siguiente cuadro.

```
Programa en R
library (leaps)
library(BMA)

datos<-read.table("hta.txt",header=T)
y=datos$y
x=data.frame(datos[,-11])
x$p2<-as.factor(x$p2)
x$P6<-as.factor(x$P6)
x$P11<-as.factor(x$P11)
x$P12<-as.factor(x$P12)
x$P14<-as.factor(x$P14)
x$P17<-as.factor(x$P17)
x$P18<-as.factor(x$P18)
x$P19<-as.factor(x$P19)
modelo.bma=bic.glm(x,y, glm.family="binomial")
summary(modelo.bma)
plot(modelo.bma, mfrow=c(3,3))
imageplot.bma(modelo.bma)
modelo.bma$bic
modelo.bma$deviance
modelo.aic1=glm(y~x$p13, family="binomial")
summary(modelo.aic1)
modelo.aic2=glm(y~x$p13+x$p18, family="binomial")
summary(modelo.aic2)
modelo.aic3=glm(y~x$p13+x$p1, family="binomial")
summary(modelo.aic3)
modelo.aic4=glm(y~x$p12+x$p13, family="binomial")
summary(modelo.aic4)
modelo.aic5=glm(y~x$p13+x$p19, family="binomial")
summary(modelo.aic5)
```

El programa seleccionó 10 modelos, de los cuales en la tabla 1 aparecen sólo los 5 mejores que tiene una probabilidad *a posteriori* acumulada del 0.862, además en dicha tabla se muestra información por columna p!, que indica el porcentaje de las probabilidades finales de inclusión de cada variables en el mejor modelo. En la columna EV se muestra los valores esperados BMA finales de los coeficientes y bajo las siglas SD se muestran las desviaciones estándar BMA finales para cada coeficiente. En las siguientes columnas aparecen los coeficientes estimados de cada modelo.
En la tabla 2 aparecen los valores de los diferentes criterios de selección de modelos que se utilizaron.

Tabla 1 Resultados de la corrida del paquete BMA en R, en el problema de la eclampsia.

	p!	EV	SD	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Intercp	100	-6.6110	1.2645	-6.53e+00	-6.93e+00	-7.83e+00	-4.4e+00	-6.19e+00
P1	12.5	0.0057	0.0185	-	-	4.759e-02	-	-
P2	0.0	-0.0000	0.0000	-	-	-	-	-
P6	2.6	-0.0336	0.0960	-	-	-	-	-
P11	2.9	-0.0093	0.1148	-	-	-	-	-
P12	7.4	-0.1615	0.8085	-	-	-	-	-
P13	100	0.2345	0.0244	2.332e-01	2.369e-01	2.359e-01	-2.11e-01	2.35e-01
P14	0.0	-0.0000	0.0000	-	-	-	2.34e-01	-
P17	3.3	-0.0100	0.0953	-	-	-	-	-
P18	24.0	0.2183	0.4535	-	9.149e-01	-	-	-
P19	3.6	0.0228	0.1495					-5.16E-01

Se puede observar de la tabla 1 que la variable 13 (semanas de gestación), es la más importante debido a que tiene una probabilidad de inclusión del 100%, es decir que aparece como variable importante en cualquier modelo seleccionado. También se puede observar que el mejor modelo sólo incluye a esta variable, además del intercepto, y tiene la probabilidad final más alta (ver tabla 2).
Comparando este resultado con los otros criterios (ver tabla 2), el BIC considera como mejor modelo el modelo 1, al igual que el DIC, sin embargo no ocurre así con el AIC , criterio que selecciona el modelo 2, sin embargo es necesario recordar que el AIC es un criterio que no incluye la incertidumbre en la selección de dicho modelo. Otro aspecto importante es que los valores que tienen los modelos dentro de cada criterio no permiten decidir de manera clara cual es el mejor modelo, ya que en algunos casos la decisión en la selección viene dada por que el valor del criterio sea el mayor o el menor como es el caso del DIC (cuyo valor debe ser mayor) y BIC (cuyo valor debe ser el menor). Sin embargo podemos observar que el criterio BMA permite decidir de manera clara cual es el mejor modelo.

Tabla 2 Resultados de los 5 modelos obtenidos en la selección de modelos.

Modelo	Variables	DIC	AIC	BIC	BMA
Modelo1	P13	169.65	173.65	-1923.82	0.479
Modelo2	P13, P18	165.62	171.63	-1921.96	0.190
Modelo3	P13, P1,	166.91	172.91	-1920.67	0.100
Modelo4	P12, P13	168.33	174.33	-1919.25	0.049
Modelo5	P13, P19	168.54	174.54	-1919.04	0.044

4. CONCLUSIONES

La utilización del criterio BMA permitió seleccionar de manera clara y sencilla el mejor modelo debido a que presenta la probabilidad final para modelo seleccionado y los ordena de mayor a menor. También la implementación del criterio BMA en el software R favorecer la selección de modelos bajo el enfoque bayesiano.
A diferencia de los criterios BIC y DIC, el BMA resultó el mejor para la selección cuando se incluye incertidumbre.

El modelo seleccionado apunta a que la presencia de preeclampsia dada las variables estudiadas esta asociada al factor de riesgo semanas de gestación (P13).

RECEIVED NOVEMBER, 2014
REVISED FEBRUARY, 2015

REFERENCIAS

[1] AKAIKE, H. (1973): Information theory and the extension of the maximum likelihood principle**.** In Proc. Int. Symp**. Information Theory (Edited by B. N. Petrov and F. CzÅLaki)**, 267-281. Akademia KiadoÅLo, Budapest.

[2] AKAIKE, H. (1978): On the likelihood of time series model. **The statistician**, 27, 217-235.

[3] AKAIKE, H. (1979): A Bayesian extension on the minimum AIC procedure of autoregressive model fiting. **Biometrika**, 66, 237-242.

[4] ANDO, T. (2010**): Bayesian Model Selection and Statistical Modeling** A Chapman& Hall Book.

[5] BAYARRI, M., DEGROOT, M., AND KADANE, J. (1988): What is the Likelihood Function?, in: S. Guota and J. Berger (ed):, **Statistical Decision Theory and Related topics** IV, 1-27. Springer: New York.

[6] BERNARDO, J. M. (1979): Reference posterior distributions for Bayesian inference. **J. R. Statist. Soc.** B41,113-147.

[7] BOZDOGAN, H. (1987): Model Selection and Akaike´s information criteria (AIC):: the theory and its analytic extension. **Psychometrika**, 52, 3, 345-370.

[8] BURNHAM, K. P. and ANDERSON, D. R. (2002): **Model selection and multimodel inference: a practical information-theoretic approach.** Springer-Verlag, . New York..

[9] CLAESKENS, G. AND HJORT, N. L. (2008): **Model Selection and Model Averaging** Cambridge University Press, Cambridge.

[10] DIAZ, L., SISTACHS, V. y COVARRUBIAS, D. (2014): Selección de modelos bajo el enfoque Bayesiano: una aplicación al estado cognitivo de los adultos mayores en el estado de Guerrero. **Revista Investigación Operacional**, 35, 180-188.

[11] FOSTER M., AND SOBER E., (2011): AIC Scores as Evidence A Bayesian interpretation, in Malcolm Forster and Prasanta S. Bandyopadhyay (eds): **The Philosophy of Statistics**. 535-549. Kluwer, Dordrech:.

[12] GELFAN, A.E. AND DEY, D. K., (1994): Bayesian Model: asymptotic and exact calculations. **Journal of the Royal Statistical Society** B56: 510-514.

[13] GEORGE, E. (2000): The variable selection problem. **Journal of the American Statistical Association**, 95,1304–1308.

[14] GOOD, I. (1952): Rational decisions. **Journal of the Royal Statistical Society**, Series B, 14, No. 1, 107-114.

[15] HENDERSON, L., GOODMAN, N. D., TENENBAUM, J. B., and WOODWARD, J. F. (2010): The structure and dynamics of scientific theories: a hierarchical Bayesian perspective. **Philosophy of Science**,77,. 172-200.

[16] KADANE, J. B. and LAZAR, N. A (2004):, Methods and Criteria for Model Selection. **Journal of the American Statistical Association**, 99, 279-290.

[17] PERRA, S. (2013): **Objective Bayesian Variable Selection for Censored data**. Universidad degli Studi di Cagliari. SESC-S/01.

[18] RAFTERY, A. (1996): Approximate Bayes factors and accounting for model uncertainty in generalised linear models. **Biometrika**, 83, 51–266.

[19] RAFTERY, A., HOETING, J., VOLINSKY, C., PAINTER, I. & YEUNG, K. Y. (2010): BMA: Bayesian Model Averaging. R package version 3.13.URL: **http://CRAN.Rproject.org/package=BMA** (consultado en noviembre 2014)

[20] R DEVELOPMENT CORE TEAM (2010): **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria.

[21] SPRENGER J., (2013): The Role of Bayesian philosophy within Bayesian Model Selection, European **Journal for the Philosophy of Science**. 3, 101–114.

[22] SPIEGELHALTER D., NICOLA G., Bradley P., VANDER A., (2002): Bayesian Measures of model complexity and fit (with discussion). **Journal of the Royal Statistical Society** B 64:583-639.

[23] SCHWATRZ, G. (1978): Estimating the dimension of a model. **The Annals of Statistics**, 6, 461-464.

[24] TAKANE, Y. (1987): Introduction to special section. **Psychometrika**, 52, 316-316.
[25] VIALLEFONT V, RAFTERY A, (2001): Richardson S. Variable selection and Bayesian model averaging in case- control studies. **Statistics in Medicine**, 20, 3215–3230.
[26] WAGENMAKERS, E. J. and FARRELL, S. (2004): AIC model selection using Akaike weights **Psychonomic Bulletin and Review**, 11, 192-196.