

TESIS DE MAESTRÍA EN ESTADÍSTICA APLICADA

Gabriela Susana Boggio*

Directora

Elsa Clotilde Servy*

EVALUACIÓN DE CRITERIOS DE BONDAD DE AJUSTE PARA MODELOS LOGIT MARGINALES**

PREMIO PROVINCIAL 2008 A TESIS DE MAESTRÍA***

Resumen. En este trabajo se estudia el comportamiento de medidas de bondad de ajuste de modelos logit marginales para datos binarios correlacionados. Algunas de ellas son de carácter descriptivo, basadas generalmente en la noción de pérdida incurrida al usar un modelo aproximado para describir la realidad, y otras constituyen estadísticas globales de bondad de ajuste. En general son extensiones naturales de las disponibles para el caso de modelos logit clásicos, es decir bajo el supuesto de independencia de las observaciones. Para la evaluación de estas medidas se realiza un estudio por simulación mediante la utilización de un algoritmo simple de generación de datos con estructura de asociación intra-unidad constante. Entre las estadísticas de tipo descriptivo se destaca el buen comportamiento de la extensión del criterio de Akaike, la cual constituye una medida útil para seleccionar el mejor subconjunto de covariables a incluir en el modelo. En lo que se refiere a los tests globales de bondad de ajuste, la extensión de la prueba chi cuadrado de Pearson y la basada en la suma no ponderada de residuos son las que presentan en general mejor comportamiento en cuanto al control del error tipo I, si bien la potencia resulta baja.

* Docentes-Investigadoras de la Facultad de Ciencias Económicas y Estadística. Elsa Servy fue Directora de la Escuela de Estadística.

Contacto: gboggio@fcecon.unr.edu.ar

** Defendida en Rosario el 7 de diciembre de 2007.

*** Otorgado por la Secretaría de Ciencia, Tecnología e Innovación del Gobierno de la Provincia de Santa Fe.

Palabras clave: Datos binarios correlacionados; Modelos marginales; Bondad de ajuste

Abstract. This thesis presents a study about the performance of goodness of fit statistics for marginal logit models with correlated binary data. Some of them are descriptive measures, usually based on the concept of information lost when an approximating model is used to describe the reality, while others are global goodness of fit statistics. In general they are natural extensions of those available for conventional logit models, i.e. under the assumption of independent binary data. In order to evaluate these measures, a simulation study is carried out using a simple data generation algorithm with exchangeable correlation structure. The Akaike criterion extension stands out among the descriptive statistics and it is also a measure useful to select the best covariable subset to include in the model. Among the global goodness of fit statistics, the extensions of the Pearson statistic and the unweighted sums of squares statistic have the best behavior in terms of type I error rates but they have low power.

Key words: Correlated binary data; Marginal models; Goodness of fit

1. Introducción

Una preocupación recurrente por parte de los investigadores es hallar el modelo estadístico que mejor represente la información proporcionada por los datos, ya que todas las inferencias que se realicen a posteriori se basan en la veracidad de ese modelo. La búsqueda de un buen modelo se complica aún más cuando no se cumple el supuesto de independencia entre las observaciones. En estos casos han sido diseñadas nuevas estrategias, entre ellas el ajuste de los denominados modelos marginales, el cual puede realizarse a través del método de las ecuaciones de estimación generalizadas (GEE). Esta metodología no requiere supuestos distribucionales para las observaciones, pero al no estar disponible su verosimilitud completa también deben modificarse los criterios de bondad de ajuste con que los modelos son evaluados.

El objetivo de esta tesis es estudiar el comportamiento de criterios de bondad de ajuste de modelos logit marginales para datos binarios correlacionados, abarcando medidas de carácter descriptivo y pruebas estadísticas, mediante un estudio de simulación.

2. Medidas de bondad de ajuste para el modelo logit marginal

En los últimos años, algunos autores han formulado estadísticas de bondad de ajuste para modelos logit marginales. Algunas de ellas constituyen extensiones naturales de los tests corrientemente utilizados en la regresión logística clásica. Otras no se basan en la verosimilitud del modelo sino que parten de la noción más general de “pérdida” producida por elegir un modelo determinado para aproximarse a la realidad. Entre estos últimos criterios se pueden mencionar la extensión del coeficiente de determinación y la extensión de la entropía para su aplicación en modelos logit marginales (Zheng, 2000), las cuales hacen uso de la suma de las comparaciones individuales entre cada observación y su estimación a partir del modelo no sólo a través de las unidades independientes sino también a través de las sub-unidades correlacionadas correspondientes a cada una de ellas. Pan (2001) presenta una extensión del criterio de información de Akaike para estos modelos, definiendo una nueva medida de discrepancia entre el modelo propuesto y el “verdadero”, en la que reemplaza la verosimilitud por la cuasi-verosimilitud bajo el modelo de independencia de trabajo disponible en el contexto de GEE.

En relación a las pruebas globales de bondad de ajuste, Evans (1998) manifiesta que el test de Hosmer-Lemeshow (1980), desarrollado para datos independientes, puede usarse para probar el ajuste de modelos marginales salvo cuando la correlación intra-unidad es grande o el conjunto de datos es reducido. Horton et al. (1999) presentan una extensión de la prueba de Hosmer-Lemeshow para los modelos marginales. Ellos proponen formar grupos basados en combinaciones de valores de las covariables incluidas en el modelo y probar si los coeficientes de regresión de las variables indicadoras asociadas a estos grupos difieren significativamente de cero. Por último, Evans (1998), Pan (2002) y Evans y Hosmer (2004) desarrollaron dos estadísticas basadas en la comparación de valores observados versus valores predichos. Se trata de las extensiones naturales de las

estadísticas chi cuadrado de Pearson y suma no ponderada de residuos al cuadrado asociadas a los tests corrientemente utilizados en el modelo logit clásico.

3. Estudio del comportamiento de las estadísticas

Para estudiar el comportamiento de las estadísticas recién presentadas se realiza un estudio de simulación bajo diferentes escenarios, teniendo en cuenta la dimensión de la muestra (cantidad de unidades y sub-unidades dentro de ellas), la correlación entre las sub-unidades dentro de la misma unidad y el tipo y nivel de las covariables a incluir en el modelo. Los datos binarios correlacionados se generan con estructura de correlación constante entre todos los pares de observaciones dentro de cada unidad mediante el uso del algoritmo presentado por Evans y Hosmer (2004).

El comportamiento de los criterios basados en la noción de pérdida, se estudia a través de su cálculo para diferentes modelos: el que se corresponde con el de generación de los datos y otros más y menos parametrizados. Interesa evaluar si efectivamente se elige el modelo adecuado en función de cada una de las medidas.

En lo referente a los tests globales de bondad de ajuste, la evaluación se realiza en términos del control del error de tipo I y II. Para lograrlo se necesita generar datos que verifiquen la hipótesis nula en un caso y no la verifiquen en otro, calcular las estadísticas para esos datos y en base a repeticiones de estos pasos, establecer el porcentaje de rechazo real de cada uno de los tests.

4. Resultados

Los resultados hallados acerca del comportamiento de los criterios de carácter descriptivo sugieren que la extensión del criterio de Akaike es el más útil para seleccionar el mejor subconjunto de covariables a incluir en el modelo. Las extensiones de la entropía y del coeficiente de determinación, si bien son simples en lo que hace a su cálculo, no facilitan demasiado la tarea de elegir el subconjunto de covariables dado que los valores obtenidos para modelos que sobre-ajusten, sub-ajusten o representen con fidelidad la realidad, pueden ser muy parecidos.

En cuanto a los tests globales de bondad de ajuste, los desarrollados por Pan y Evans y Hosmer tienen en general buen comportamiento en lo que hace al control del error tipo I. Sin embargo la potencia es generalmente muy baja, siendo sólo aceptable para el caso en que se evalúa la omisión del término cuadrático asociado a una de las covariables del modelo. El test de Hosmer-Lemeshow debe usarse con cautela, teniendo en cuenta que para tomar en base a él decisiones confiables se debe disponer de muchas unidades, la correlación intra-unidad debe ser moderada o baja y preferentemente no se debe contar con covariables que varíen a nivel de las unidades. Del mismo modo, el test de Horton et al. no es aplicable en todas las situaciones. Funciona razonablemente bien cuando las covariables son categóricas y se dispone de al menos 100 unidades.

5.Consideraciones finales

Los resultados hallados acerca del comportamiento de las estadísticas en diferentes situaciones permiten dar recomendaciones acerca del apropiado uso de una u otra estadística de bondad de ajuste según sean las características particulares de cada conjunto de datos. Pero debe tenerse presente que aunque la elección de la estadística sea adecuada, ninguna de ellas es potente en relación a la mayoría de las hipótesis alternativas analizadas. Por ello estas estadísticas proporcionan resultados de importancia cuando indican falta de ajuste del modelo ya que impulsan al analista a buscar formas de mejorarlo. En cambio no rechazar la hipótesis de bondad de ajuste no constituye una evidencia definitiva sobre que el modelo tenga un buen ajuste y en este caso es necesaria la utilización de otras técnicas para la confirmación de la adecuación del modelo tales como las medidas de diagnóstico. Todas estas consideraciones refuerzan la idea compartida por muchos autores de que la búsqueda de un buen modelo tiene tanto de arte como de ciencia.

Referencias bibliográficas

- Evans, S. R. (1998). Goodness of fit in two models for clustered binary data. Unpublished doctoral dissertation. University of Massachusetts, Amherst, USA.
- Evans, S. R.; Hosmer, DW. (2004). Goodness of fit tests for logistic GEE models: Simulation Results. *Communications in Statistics: Simulation and Computation*, 33(1), 247-258.
- Hosmer, D. W.; Lemeshow, S. (1980). A goodness of fit test for the multiple logistic regression. *Communications in Statistics*, A10, 1043-1069.
- Horton, N.; Bebchuk, J.; Jones, C.; Lipsitz, S.; Catalano, P.; Zahner, G.; Fitzmaurice, G. (1999). Goodness of fit for GEE: an example with mental service utilization. *Statistics in Medicine*, 18, 213-222.
- Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57, 120-125.
- Pan, W. (2002). Goodness of fit tests for GEE with correlated binary data. *Scandinavian Journal of Statistics*, 20, 101-110.
- Zheng, B. (2000). Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in Medicine*, 19, 1265-1275.