PESQUIMAT Revista de la Fac. CC. MM. de la
Universidad Nacional Mayor de San Marcos
Vol. X N°1, pp 55 - 70, LIMA - PERÚ. Agosto 2007

Análisis de Diagnóstico en el Modelo de Regresión Logística: Una aplicación

Olga Solano Dávila[1] Agustina Ramírez[2]
osolanod@unmsm.edu.pe
Felix Bartolo[1] Orlando Giraldo[1] Alfredo Salinas[1]

Resumen

El análisis de regresión logística es una técnica multivariante de mucha importancia por sus aplicaciones en diferentes áreas del conocimiento humano. Su aplicación viene en aumento cada vez mas. En la investigación clínica y epidemiologica, en un estudio sobre enfermedades coronarias, el análisis de regresión logística múltiple fue aplicado por primera vez a principio de los años 60 [13]. En el estudio del análisis de regresión logística, frecuentemente el conjunto de los datos contiene algunas observaciones atípicas o extremas en relación a los datos. En la construcción de un modelo de regresión logística es importante examinar los datos que están siendo utilizados, con el objetivo de determinar la existencia de uno o varios puntos que están controlando propiedades importantes del modelo. En este trabajo se hace un estudio de diagnóstico en el modelo de regresión logística múltiple [5], sobre los factores de riesgo en la enfermedad de osteoporosis.

Palabras Clave: *Técnica multivariante, modelo de regresión logística múltiple, análisis de diagnóstico, análisis de residuos, análisis de influencia.*

Abstract

Logistic regression is a multivariate technique very important for its applications in different areas of knowingness and its applications has been growing more. In clinical and epidemiological research,in particular in a study coronary illness, analysis of logistic regression has been applied for first time around 60 years old [13] . In studies of logistic regression, it is frequent that a group of observations can be outliers. In the construction of logistic regression models is important to examine the observations to detect the existence of one or more observations that is controlling important properties of the model. We present a discussion on diagnostic to logistic regression model [5], on factors of risk in illness of bone.

Keywords: *Multivariate technic, logistic Regression model, Diagnostic analysis, analysis of residues, analysis of influence.*

[1]UNMSM, Facultad de Ciencias Matemáticas, Lima - Perú.
[2]UNFV, Facultad de Ciencias Naturales y Matemáticas - Lima, Perú.

1. Introducción

El modelo de regresión logística ha sido utilizado por muchos años, más fue solamente en [13], que se constato la importancia y aplicación de estos modelos, cuando los investigadores analizaron los datos de Framingham, que trata de un estudio sobre el corazón. Después de la publicación de este artículo, el modelo de regresión logística se torna en el método estándar para el análisis de regresión de los datos dicotómicos en muchas áreas del conocimiento humano, especialmente en las ciencias de la salud. El objetivo de esta técnica estadística multivariante puede ser la estimación, esto es, estimar la mejor relación de las variables independientes con la variable dependiente, utilizada en su mayoría en estudios etiológicos que consiste en investigar factores causales de una determinada característica de la población y estudiar que factores modifican la probabilidad en la aparición de un suceso determinado; o también predictivo, que consiste en predecir lo mejor posible la variable dependiente a través de las independientes (pueden ser variables cuantitativas o cualitativas). Generalmente es dicotómico (clasifica el valor de la variable respuesta como 1 cuando presenta la característica de interés y con el valor 0 cuando no presenta); también puede ser usada para estimar la probabilidad de cada una de las posibilidades de un suceso en mas de dos categorías (politomico).

La técnica multivariante resulta útil para determinar factores de riesgo y factores de prevención de enfermedades en muestras prospectivas donde la metodología de regresión lineal no es aplicable, dado que la variable respuesta solamente presenta dos valores (caso dicotómico), como por ejemplo presencia o ausencia de un suceso.

El objetivo de este trabajo es hacer un estudio de diagnóstico en el modelo de regresión logística múltiple, [1], y utilizamos una muestra aleatoria de los datos de [11], sobre los factores de riesgo en la enfermedad de osteoporosis.

En la siguiente sección presentamos el modelo de regresión logística múltiple.

2. Modelo de regresión logística múltiple

Considere que tenemos la disposición de p variables independientes expresada por el vector $X' = (x_1, x_{2,...,}x_p)$ y relaciona la probabilidad de que ocurra un determinado suceso independiente denotada por el vector X' con probabilidad condicional $P(Y = 1/X) = \pi(x)$ en función de p variables independientes que pueden ser cuantitativas o cualitativas según el tipo de diseño de estudio.

El logit del modelo de regresión logística múltiple se presenta por la siguiente ecuación

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p,$$

donde β_0 ,... , β_1 , ... , β_p son parámetros del modelo de regresión logística múltiple.

En este caso el modelo de regresión logística es

$$\pi(x) = p_j = \frac{e^{g(x)}}{1 + e^{g(x)}},$$

En la siguiente sección presentamos la estimación de los parámetros del modelo de regresión logística.

2.1. Estimación de los parámetros

Para estimar los parámetros del modelo se utiliza el método de máxima verosimilitud. Suponiendo que tenemos una muestra de n observaciones independientes (x_i, y_i) , $i = 1, 2, ..., n$. Ajustar el modelo, requiere que obtengamos estimadores del vector $\beta' = (\beta_0, \beta_1, ..., \beta_p)$. Las ecuaciones de verosimilitud que resultan pueden ser expresadas de la siguiente forma

$$\sum_{i=1}^{n} [y_i - \pi(x_i)] = 0$$

y

$$\sum_{i=1}^{n} x_{ij} [y_i - \pi(x_i)] = 0$$

para $i = 1, 2, ..., n$.
Para encontrar la solución de este conjunto de ecuaciones se utiliza métodos iterativos. Hoy en día existen paquetes estadísticos para estimar estos parámetros.

$\hat{\beta}$ denota la solución de estas ecuaciones.

En la siguiente sección mostramos las pruebas de hipótesis para evaluar el modelo de regresión logística.

2.2. Prueba de hipótesis para el modelo de regresión logística

Después de ajustado el modelo de regresión logística múltiple, empezaremos el proceso de evaluación. El primer paso en este proceso generalmente es fijar la significación de las variables en el modelo. La prueba de razón de verosimilitud para la significación total de los p coeficientes para las variables independientes en el modelo es basado en la estadística G

$$G = 2\left\{\sum_{i=1}^{n} [y_i \ln(\hat{\pi}_i) + (1 - y_i)\ln(1 - \hat{\pi}_i)] - [n_i \ln(n_i) + n_0 \ln(n_0) - n \ln(n)]\right\},$$

Los valores ajustados, $\hat{\pi}_i$, sobre el modelo, son basados sobre el vector que contienen $p + 1$ parámetros, $\hat{\beta}$, sobre la hipótesis nula de que los p coeficientes para las covariables en el modelo son iguales a cero. La estadística G tiene una distribución Chi-cuadrado con p grados de libertad.
La prueba de Wald se obtiene del cálculo de la siguiente matriz

$$W = \hat{\beta}' \left[\hat{Var}(\hat{\beta})\right]^{-1} \hat{\beta}$$

$$= \hat{\beta}'(XVX)\hat{\beta},$$

donde V es una matriz diagonal de dimensión $n \times n$ con elementos $\hat{\pi}_i\,(1-\hat{\pi}_i)$ y W tiene una distribución Chi-cuadrado con $p+1$ grados de libertad sobre la hipótesis de que cada uno de los $p+1$ coeficientes son iguales a cero.
El análisis de residuos es importante para examinar si uno o varios de los datos están controlando propiedades importantes del modelo, en la siguiente sección presentamos el análisis de residuos.

2.3. Análisis de los residuos para el modelo de regresión logística

Existen varios tipos de residuos que sugieren si una observación es atípica o no.

2.3.1. Residuos de Pearson

Los residuos de Pearson son definidos de la forma

$$r_j = \frac{y_j - m_j\hat{p}_j}{\sqrt{m_j\hat{p}_j\,(1-p_j)}},$$

donde, y_j representa el número de respuestas, $y = 1$, entre los m_j individuos con $X_j = x$ (algunos individuos que tienen el mismo valor x) , $j = 1, ..., p$.

$$\hat{p} = \hat{\pi}(x) = \frac{e^{\hat{g}(x)}}{1 + e^{\hat{g}(x)}},$$

y $g\,(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \beta_p x_p$.
El residuo de Pearson es similar al residuo estudentizado usado en la regresión lineal. Así, un residuo de Pearson en valor absoluto mayor que 2 indica un dato atípico.
La estadística χ^2 de Pearson es la suma de cuadrados de los residuos de Pearson.

$$\chi^2_P = \sum_{i=1}^{J} r_j^2$$

2.4. Residuos de Pearson Estandarizado

Los residuos de Pearson estandarizado están definidos por:

$$r_{sj} = \frac{r_j}{\sqrt{1-h_j}}$$

donde, r_j son los residuos de Pearson y h_j es el leverage, esto es, el elemento de la diagonal principal de la matriz H.

2.4.1. Residuos de Deviance

Los residuos de Deviance están definidos de la forma:

$$d_j = \pm \left\{ 2 \left[y_j \ln \left(\frac{\mathrm{y_j}}{\mathrm{m_j \hat{p}_j}} \right) + (\mathrm{m_j - y_j}) \ln \left(\frac{\mathrm{m_j - y_j}}{\mathrm{m_j}\,(1 - \hat{\mathrm{p}}_\mathrm{j})} \right) \right] \right\}^{1/2}$$

La Deviance es la suma de cuadrados de los residuos de la deviance

$$\chi^2_D = \sum_{i=1}^{p} d_j^2$$

Si la deviance es mayor que 4 en valor absoluto entonces la observación correspondiente es atípica.

Las medidas de influencia en el modelo de regresión logística son presentadas en el siguiente parágrafo.

2.5. Medidas de influencia en regresión logística

Son de gran utilidad para evaluar la presencia de datos influyentes sobre el modelo de regresión logística ajustado.

2.5.1. Leverage

Son los elementos de la diagonal de la matriz de predicción H . El leverage para la observación *i-ésima* es el elemento *i-ésimo* de la diagonal principal de la matriz H , h_i , toma valores entre 0 y 1 con un valor medio de p/n .

$$H = X_* \left(X'_* X \right)^{-1} X'_*$$

donde $X_* = V^{1/2} X$ y $V = diag(\hat{\pi}_i(1 - \hat{\pi}_i))$.
las estadísticas de influencia derivadas de h_i son:

1. La distancia de Cook ($\Delta \mathrm{B}_i$): La distancia de Cook, mide la influencia de la estimación de los parámetros del modelo, β , y tiene la forma

$$\Delta B_j = \frac{r_{sj}^2 h_j}{(1 - h_j)},$$

donde r_{sj} : residuos de Pearson estandarizado Si $\Delta \mathrm{B}_j > 1$ la observación es influyente en los valores de los parámetros estimados.

2. Estadística Delta Chi-cuadrado de Pearson ($\Delta\chi^2_{P(i)}$): La estadística Delta Chi-cuadrado de Pearson χ^2 es calculada eliminando las observaciones de la variable x_j del modelo, esta estadística tiene la forma

$$\Delta\chi^2_{P(j)} = \frac{r_j^2}{(1-h_j)},$$

donde r_j son los residuos de Pearson y h_j es el leverage para una observación y es el elemento de la diagonal principal de la matriz H.
Valores grandes de la estadística son considerados influyentes sobre los valores estimados de los parámetros.

3. Estadística Delta Deviance Mide la deviance que resulta al excluir las observaciones de la variable x_j del modelo, esta estadística es de la forma

$$\Delta\chi^2_{D(j)} = \frac{d_j^2}{(1-h_j)}$$

$\Delta\chi^2_{P(j)}$ y $\Delta\chi^2_{D(j)}$ miden como el modelo estimado es afectado por los parámetros del modelo.
Valores grandes indican que el modelo estimado se torna mejor al excluir esa observación de los datos. Como regla tenemos que valores mayores que 4 indican influencia sobre las estimativas de los parámetros del modelo.

2.5.2. Gráficos de diagnóstico

Los gráficos de diagnóstico son de gran utilidad para evaluar la presencia de datos influyentes. Para tener una descripción rápida de la información proporcionada por las estadísticas estudiadas anteriormente, utilizamos, [5], los siguientes gráficos:

1. Delta Chi-cuadrado vs la probabilidad estimada.
2. Delta Deviance vs la probabilidad estimada.
3. Distancia de Cook vs la probabilidad estimada.
4. Delta Chi-cuadrado vs leverage.
5. Delta Deviance vs leverage.
6. Distancia de Cook vs leverage.

3. Análisis del Conjunto de datos de los factores de riesgo en la enfermedad de la osteoporosis

En esta sección analizamos un conjunto de datos reales. En el presente trabajo seleccionamos una muestra aleatoria simple de los datos de pacientes atendidos en la Clínica Good Hope, situada en Malecón Balta 956, en el distrito de Miraflores, de diciembre de 2000 a marzo de 2001, [11].
Para calcular el tamaño de la muestra se utiliza la variable prevalencia de osteoporosis ($p = 0,402$), un margen de error de 0,10 y un nivel de confianza del 95 %, lo que resulto una muestra de 48 pacientes.
Las zonas de riesgo consideradas fueron detectadas por la densitometría ósea, mediante la absorciometria de energía dual de rayos X (DEXA); es una técnica estándar o prueba de oro que indica si un individuo tiene o no osteoporosis.
La descripción de las variables en estudio es:
Variables independientes:
Edad (años)
L1L4: Columna lumbar antero posterior (gr/cm^2)
CF: Cuello femoral (gr/cm^2)
CC: Cadera completa (gr/cm^2)
T: Trocanter (gr/cm^2)

Variable dependiente
Tiene osteoporosis: (1: Si, No: 0)
Para obtener esta variable se utilizo el ultrasonido cuantitativo de calcáneo (US de calcáneo) para detectar si una persona tiene o no osteoporosis.

Es una técnica nueva que ha sido sugerida como un método alternativo para evaluar la masa y la calidad ósea. Consiste en aplicar ondas de ultrasonido y observar como se modifican a través del hueso a una velocidad del sonido (SOS) en m/seg. y la frecuencia de atenuación del sonido (BUA) en dB/mhz que viajan a través del hueso. Al ser mezcladas se obtienen la variable llamada *stiffness* en gr/ cm^2 , expresa en porcentaje del valor T del adulto joven normal. Estos parámetros son de gran utilidad porque se correlacionan con la densidad masa ósea (DMO) a través de su densidad, elasticidad y estructura ósea. Es un método que no tiene radiación, y de bajo costo, fácil transporte y menor tiempo de ejecución.
En la siguiente sección ajustamos el modelo de regresión logística múltiple.

3.1. Ajuste inicial en el modelo de Regresión Logística en las variables en estudio

Se desea comprobar la capacidad predictiva del modelo de regresión logística en 4 posibles zonas de riesgo de osteoporosis en una muestra aleatoria de pacientes de una Clínica en Miraflores.

Para seleccionar las variables significativas en el modelo utilizamos o método de Backward Stepwise (Condicional). A continuación mostramos algunos resultados.

Cuadro 1: Tabla de Clasificación por el Modelo 1

	Predecidos		Porcentaje Correcto
Observados	No tiene osteoporosis	Tiene osteoporosis	
No tiene osteoporosis	33	2	94,3
Tiene osteoporosis	4	9	69,2

Porcentaje Correcto Global 87,5 %

El 94,3 % de los pacientes que no tienen osteoporosis fueron correctamente clasificados por el modelo, mientras que el 5,7 % no fueron correctamente clasificados y el 69,2 % de los pacientes que tienen osteoporosis fueron clasificados correctamente por el modelo, mientras que el 30,80 % no fueron clasificados correctamente. La tabla de clasificación muestra que, con este modelo, fueron clasificados correctamente el 87,50 % de pacientes que tienen o no osteoporosis.

Cuadro 2: Variables en el Modelo 1

Variables	$\hat{\beta}$	Error estándar	Wald	Grados de libertad	Sig.	Exp ($\hat{\beta}$)
L1L4	-1,493	0,714	4,371	1	0,037	0,225
CC	-1,928	0,709	7,400	1	0,007	0,145
Constante	-7,189	2,252	10,22	1	0,001	0,001

Las siguientes zonas son de mayor influencia para predecir el riesgo en la presencia de osteoporosis, columna lumbar anteroposterior, L1L4 ($p = 0,037$) y cadera completa, CC, ($p = 0,007$). Estas zonas son significativas a un nivel de significancia del 0,05.

Cuadro 3: Variables no consideradas en el Modelo

Variables Score Grados de Libertad Sig.
EDAD 0,816 1 0,929
CF 0,008 1 0,268
T 2,196 1 0,533

Las variables de menor influencia para predecir el riesgo en la presencia de osteoporosis son cuello femoral, CF ($p = 0,268$), Trocánter, T ($p = 0,533$) y EDAD ($p = 0,929$) .
Para evaluar la bondad del ajuste del modelo utilizamos a prueba de Hosmer-Lemeshow, el Cuadro 4 presenta los resultados en cada paso.

Cuadro 4: Prueba de Bondad de Ajuste de Hosmer-Lemeshow

Paso	Chi-Cuadrado	Grados de libertad	Sig.
1	1,904	8	0,984
2	5,592	8	0,693
3	5,639	8	0,688
4	3,734	8	0,880

En el cuarto paso el ajuste es bueno ($p = 0,880$), no hay razones suficientes para rechazar la hipótesis nula con un nivel de significancia de 1 %, luego el modelo de regresión logística considerado tiene un buen ajuste.

3.2. Análisis de Diagnóstico

Los gráficos de diagnóstico son de gran utilidad para detectar la presencia de datos influyentes en el modelo de regresión logística. Para obtener una descripción rápida de la información proporcionada por las estadísticas estudiadas anteriormente, utilizamos los gráficos presentados por [5]. La Figura 1 muestra un diagrama de dispersión sobre la estadística Delta Chi-cuadrado y las estimativas de las probabilidades; las observaciones 13 y 26 son influyentes sobre las estimativas de los parámetros del modelo de regresión logística. En la Figura 2 tenemos un gráfico sobre la estadística Delta Deviance y las estimativas de las probabilidades; el individuo 13 y 26 son influyentes sobre las estimativas de los parámetros del modelo. La Figura 3 muestra que el individuo 26 se destaca del resto. En la Figura 4 se observa que, para el conjunto de datos considerados, los individuos 13 y 26 se destacan del resto. En la Figura 5 ambas aproximaciones producen las mismas observaciones influyentes, los individuos 13 y 26 se destacan del resto. La Figura 6 muestra un diagrama de dispersión

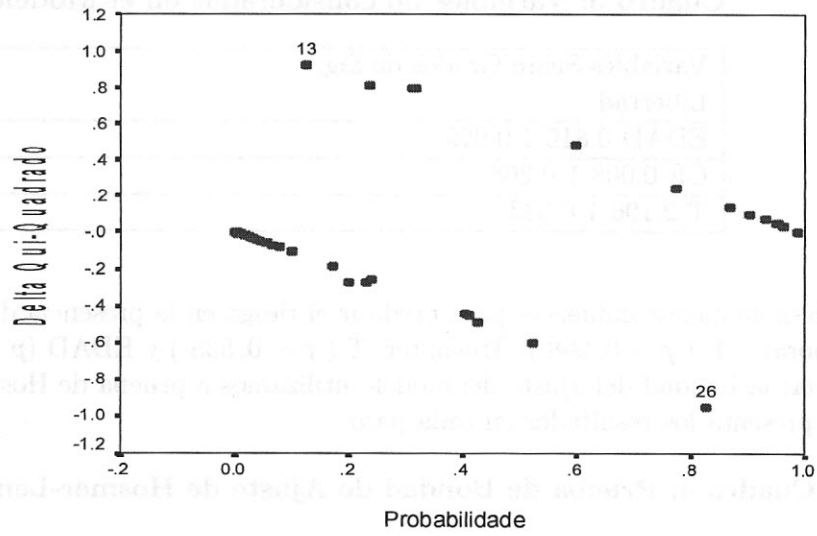

Figura 1: Delta Chi-Cuadrado vs Probabilidad

sobre las estadísticas de la Distancia de Cook y el Leverage. La observación 26 se destaca del resto.

Por lo que analizamos anteriormente en los gráficos concluimos que las observaciones 13 y 26 son influyentes en las estimativas de los parámetros del modelo de regresión logística. Eliminamos estas dos observaciones y nuevamente evaluamos el modelo.

Tenemos algunos resultados del Modelo 2, después de evaluar el modelo eliminando las observaciones 13 y 26.

Cuadro 5 : Tabla de Clasificación del Modelo 2

	Predecidos		Porcentaje correcto
Observados	No tiene osteoporosis	Tiene osteoporosis	
No tiene osteoporosis	32	2	94,1
Tiene osteoporosis	3	9	75,0

Porcentaje Correcto Global 89,1 %

El 94,1 % de los pacientes que no tienen osteoporosis fueron correctamente clasificados por el modelo, mientras que el 5,9 % no han sido clasificados correctamente, el 75,0 % de los pacientes que tienen osteoporosis fueron clasificados correctamente por el modelo, mientras

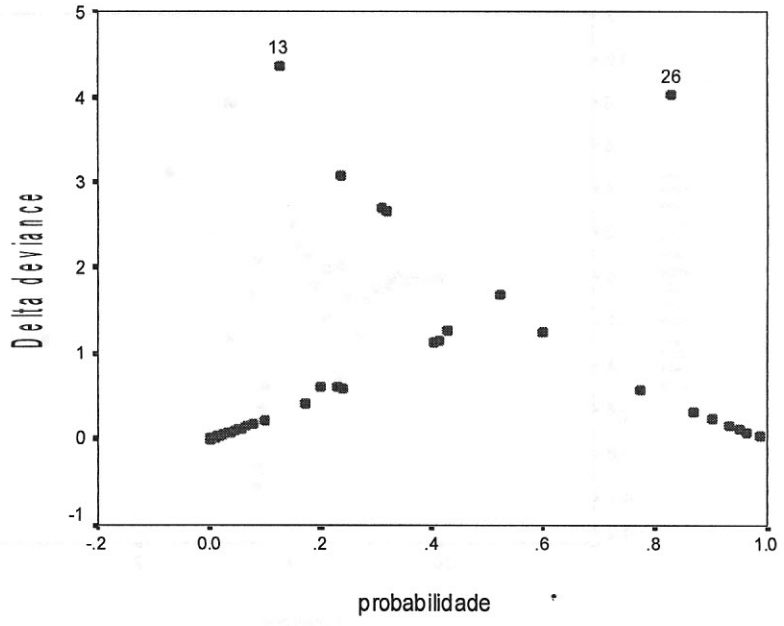

Figura 2: Delta deviance vs Probabilidad

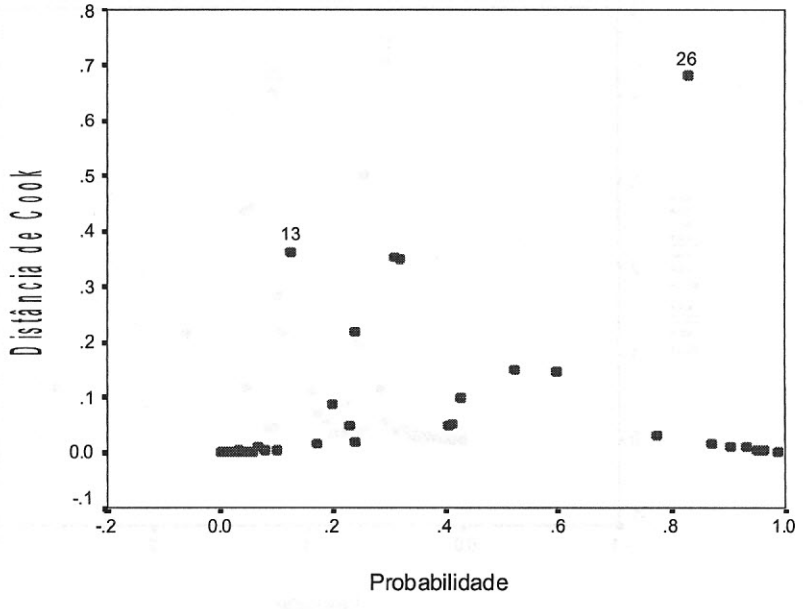

Figura 3: Distancia de Cook vs Probabilidad

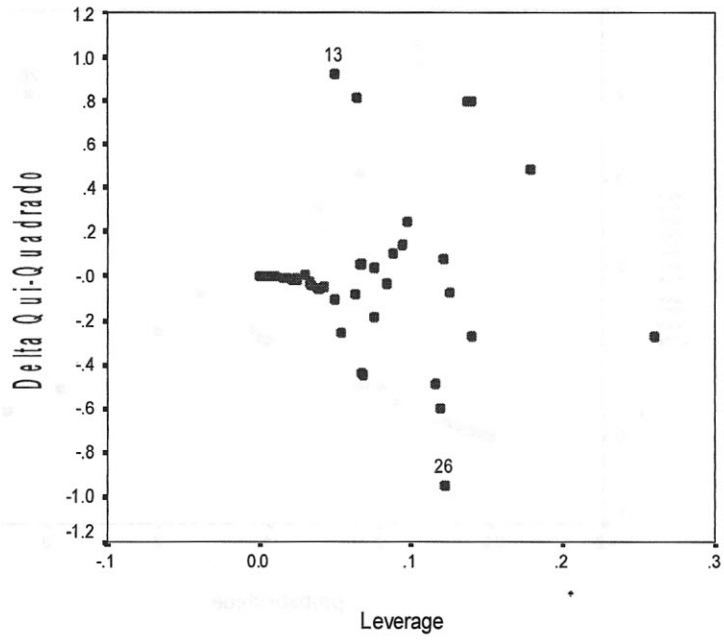

Figura 4: Delta Chi-Cuadrado vs Leverage

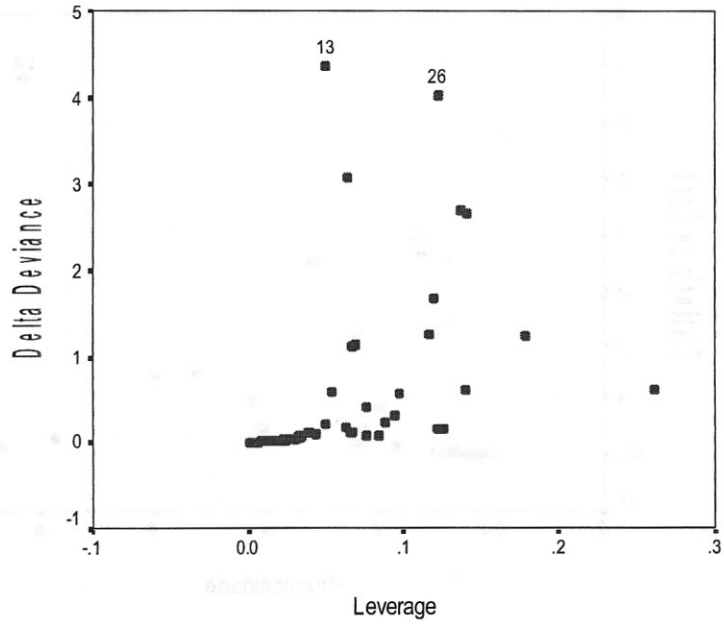

Figura 5: Delta Deviance vs Leverage

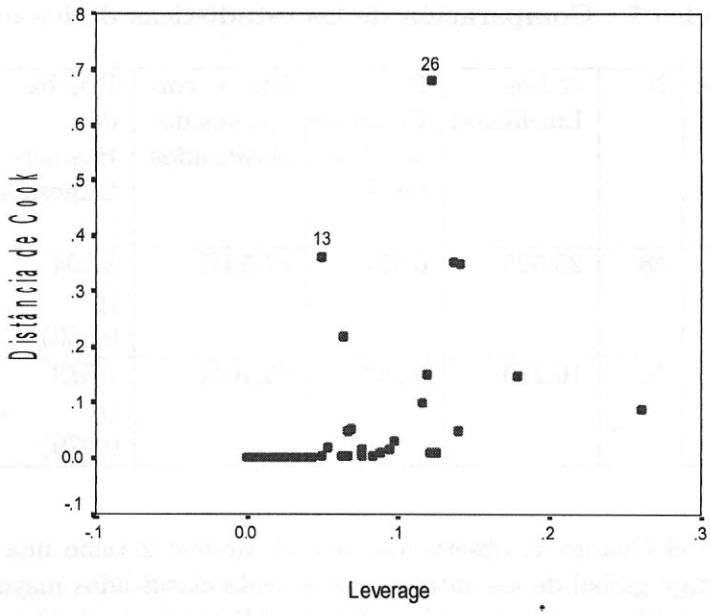

Figura 6: Distancia de Cook vs Leverage

que el 25 % no han sido clasificados correctamente. La tabla de clasificación muestra que, con este modelo, el 89,10 % de los pacientes que tienen o no osteoporosis fueron clasificados correctamente.

Cuadro 6 : Variables en el Modelo 2

Variables	$\hat{\beta}$	Error Estándar	Wald	Grados de libertad	Sig.	Exp ($\hat{\beta}$)
L1L4	-1,991	1,069	3,468	1	0,063	0,137
CC	-3,325	1,346	6,101	1	0,014	0,036
Constante	-10,564	4,117	6,584	1	0,010	0,000

Según el modelo considerado las siguientes zonas son de mayor influencia para predecir el riesgo en la presencia de osteoporosis, columna lumbar anteroposterior, L1L4 ($p = 0,063$) y cadera completa, CC ($p = 0,014$). Estas zonas son significativas a un nivel de significancia de 0,07.
El Cuadro 7 presenta las estadísticas más importantes en los dos modelos considerados.

Cuadro 7 : Comparación de las estadísticas de los dos modelos

Modelo	N	-2 Log Likelihood	R-Cuadrado de Cox Snell	Casos correctamente clasificados	Prueba de Hosmer-Lemeshow	Parámetros significativos em el modelo
1	48	25,525	0,471	87,50 %	1,904 (p = 0,880)	2/6
2	46	16,210	0,549	89,10 %	1,579 (p = 0,979)	2/6

Analizando el Cuadro 7, observamos que el Modelo 2 tiene una tabla de clasificación con un porcentaje global de los datos correctamente clasificados mayor que el Modelo 1. La verosimilitud cambio con respecto al modelo 1, el R cuadrado de Cox Snell también rechaza la hipótesis de que los coeficientes sean iguales a cero ($p = 0,979$). Por lo tanto el Modelo 2 es mejor que el Modelo 1.
Luego el modelo para diagnosticar osteoporosis, con US cuantitativo de calcáneo y mejor asociación en otras zonas de riesgo y que predice mejor el riesgo de fractura de un individuo en particular frente a probabilidades de tener osteoporosis o no, es el siguiente:

$$p_i = \frac{1}{1 + e^{-(-10,564 - 1,991 L1L4 - 3,325 CC)}}$$

3.3. Interpretación de los coeficientes del modelo

$e^{-10,564} = 0$, la probabilidad de que un individuo tenga osteoporosis y la probabilidad de no tener osteoporosis sin considerar las zonas de riesgo es 0,000.
$e^{-1,991} = 0,137$, el incremento de riesgo de poseer osteoporosis, entre una persona que tenga 1 gr/cm^2 de DMO menos que otra es 0,137 en la columna lumbar (L1L4), sin considerar la zona de riesgo en cadera completa.
$e^{-3,3325} = 0,036$, el incrementos de riesgo de poseer osteoporosis, entre una persona que tenga 1 gr/cm^2 de DMO menos que otra es 0,036 en cadera completa (CC), sin considerar la zona de riesgo en columna lumbar.
Por ejemplo, para un paciente de 70 años de edad de la base de datos de la población, con una DMO en columna lumbar anteroposterior (L1L4) de –0,37, en cadera completa (CC) de –0,96 :
La probabilidad de poseer osteoporosis es 0,0013, una probabilidad bien pequeña. (esta observación fue recolectada de la base de datos general y el paciente no tiene osteoporosis).

4. Discusión y Comentarios

El objetivo de este trabajo fué el de estudiar el análisis de diagnóstico es el modelo de regresión logística, suponiendo que el modelo 1 es correcto. Para detectar individuos influyentes hicimos una evaluación de diagnóstico de los gráficos propuestos por [5]. Esto ayuda a tener una idea sobre que individuos influyentes distorsionan la estimación del proceso. Los gráficos considerados para hacer este análisis muestran que las observaciones 13 y 26 deben ser eliminadas (Ver Figuras 1, 2, 4 y 5).
El Modelo 2, ajustado sin considerar las observaciones 13 y 26, es adecuado para predecir la probabilidad de que un paciente que llega a la clínica tenga o no osteoporosis (ver Cuadro 7). Según el Modelo 2, las variables mas relacionadas con el diagnóstico de osteoporosis, al utilizar el US de calcáneo son: columna lumbar anteroposterior (L1L4) y cadera completa (CC) a un nivel de significancia de 7%. (ver Cuadro 6). La Tabla de clasificación muestra que con este modelo clasifica correctamente el 89,10% de los pacientes que poseen o no osteoporosis. (ver Cuadro 5).

Referencias

[1] ANDERSEN, E.B. Introduction to the statistical analysis of categorical. Springer-Verlag Berlin Heidelberg. New York, 1997.

[2] ALVAREZ, C.R. Estadística multivariante y no paramétrica con SPSS: Aplicación a las ciencias de la salud. Díaz de Santos, S.A. Madrid, (1995).

[3] ATKINSON, A.C. Plots, transformations, and regression: An Introduction to graphical methods of diagnostic regression analysis. New York : Clarendon Press Oxford. (1985). 282p.

[4] GARRET, J.M. Quantitative methods: Logistic regression and exploratory data analysis. Chapel Hill. North Carolina, (1994).

[5] HOSMER, D.W.; LEMESHOW, S. Applied logistic regression. 2nd. ed. John Wiley & Sons. New York, (2000).

[6] LONG, J.S. Regression models for categorical and limited dependent variables. Sage Publications. California, (1997).

[7] MAGARO, M.; SOLI, A.; CARCCHIO, R.; ANGELOSANTE, S.; MIRONE, L.; PALAZZONI, G. Quantitative ultrasonography in the evaluation of postmenopausal osteoporosis. Comparison with dual energy x-ray absorptiometry. Ann. Ital. Med., v10, p.218-221, (1995).

[8] MENARD, S. Applied logistic regression analysis. Sage Publications. California, (1995).

[9] PEREZ, C. Técnicas estadísticas con spss. Prentice Hall, Madrid, (2001).

[10] PREGIBON, D. Logistic regression diagnostics. Ann. Stat., v.40, p.705-724, (1980).

[11] RAMÍREZ, A. Aplicación de las curvas roc en el diagnostico de osteoporosis. (2002). 121f. Tesis (Licenciada en Estadística) – Facultad de Ciencias Naturales y Matemáticas, Universidad Nacional Federico Villarreal, Lima, Perú.

[12] SOLANO, O.; BARTOLO, F.; GIRALDO, O.; SALINAS, S. Análisis de diagnostico en el modelo de regresión logistica para determinar los factores de riesgo en la osteoporosis. Infome final presentado al Instituto de Investigación de la Facultad de Ciencias Matemáticas: UNMSM, Lima. Perú, (2004).

[13] TRUETT, J.; CORNFIELD, J.;KANNEL, W. A multivariate analysis of the risk of coronary heart disease in Framingham. J. Chronic Diseases, v.20, p.511-524, (1967).

[14] VISAUTA, B. Análisis estadístico con spss para windows: Estadística Multivariante. McGraw-Hill, Madrid, (1998).

[15] WEISBERG, S. Applied linear regression. 2nd. ed. John Wiley & Sons. New York, (1985).

[16] WORLD HEALTH ORGANIZATION STUDY GROUP. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis.World Health Organization, v. 843,p.135-242, (1994).