Notas de clase de Probabilidad y Estadística

Volumen 9: Regresión logística multinomial

Versión 2 (Julio, 2019)

Dr. rer. nat. Humberto LLinás Solano
Doctor en Estadística (Mainz-Alemania)
Profesor Titular
Investigador Asociado (Colciencias)
hllinas@uninorte.edu.co

Departamento de Matemáticas y Estadística
Universidad del Norte
(www.uninorte.edu.co).

Índice general

Prefacio

Introducción

Estas notas de clase hacen parte de un compendio de varios volúmenes y están dirigido a todo tipo de público que requiere de algún conocimiento básico en Estadística.

El autor

Humberto Jesús LLinás Solano es Licenciado en Ciencias de la Educación, con énfasis en Matemáticas, Física y Estadística de la Universidad del Atlántico (Colombia). Magister en Matemáticas, convenio Universidad del Valle-Universidad del Norte (Colombia). Doctor en Estadística (Dr. rer. nat.) de la Universidad Johannes Gutenberg de Mainz (Alemania). Desde 1998 se desempeña como profesor de tiempo completo de la Universidad del Norte y forma parte de los grupos de investigación Matemáticas y Enfermedades tropicales de dicha institución. Autor de los productos[1]:

- *Estadística descriptiva y distribuciones de probabilidad* (2005, [6])
- *Estadística inferencial* (2006, [9])
- *Una visión histórica del concepto moderno de integral* (como editor, 2006, [4])
- *Medida e integración* (2007, [10])
- *Applets de estadística* (2007, [12])
- *Introducción a la estadística con aplicaciones en Ciencias Sociales* (2012, [13])
- *Procesos estocásticos con aplicaciones* (como coautor, 2013, [2])
- *Introducción a la estadística matemática* (2014, [14])
- *Introducción a la teoría de la probabilidad* (2014, [15])

[1] Se cita el título del texto o applet, el año de publicación y la referencia bibliográfica respectiva. Cuando sea necesario, un comentario adicional.

1 El modelo logístico multinomial

1.1 Modelos logísticos y modelos relacionados

1.1.1. Modelo básico

La variable de interés Y puede asumir tres niveles: 0, 1 o 2. Para cada $r = 0,1,2$, sea $p_r := P(Y = r)$ la probabilidad de que Y tome el valor r.

Haciendo n observaciones independientes de Y, se obtiene la muestra $Y = (Y_1, \ldots, Y_n)$ con los datos $y_i \in \{0,1,2\}$, $i = 1, \cdots, n$, donde y_i es un posible valor de Y_i, las cuales son independientes entre sí.

Para construir la funci{on de verosimilitud, debemos crear tres variables binarias con valores 0 y 1 e independientes, de la siguiente manera:

$$U_{ri} = \begin{cases} 1, & \text{if } Y_i = r; \\ 0, & \text{de otra forma.} \end{cases}$$

donde $r = 0,1,2$ y $i = 1, \ldots, n$. Observe que $U_{ri} \sim \mathscr{B}(1, p_{ri})$, donde $p_{ri} = P(Y_i = r)$.

En términos de las variables U_{ri}, las variables muestrales serán $Y_i = (U_{0i}, U_{1i}, U_{2i})$, con valores $y_i = (u_{0i}, u_{1i}, u_{2i})$, siendo $\sum_{r=0}^{2} u_{ri} = 1$, para i fijo. Se llega a un modelo estadístico donde:

$$P(Y_i = y_i) = \prod_{r=0}^{2} p_{ri}^{u_{ri}}, \quad i = 1, \ldots, n$$

Fijando $y = (y_1, \cdots, y_n)^T$, obtenemos el logaritmo de la función de verosimilitud

$$\mathscr{L}(p) = \sum_{i=1}^{n} [u_{0i} \ln p_{0i} + u_{1i} \ln p_{1i} + (1 - u_{0i} - u_{1i}) \ln(1 - p_{0i} - p_{1i})], \tag{1.1}$$

evaluada en el parámetro $2n$-dimensional $p = (p_{01}, p_{11}, \ldots, p_{0n}, p_{1n})^T$.

1.1.2. El modelo completo

El *modelo completo* es caracterizado por el supuesto de que todos p_{ri} (con $r = 0,1,2$ y $i = 1,\ldots,n$) son considerados como parámetros.

Teorema 1.1

En el modelo completo, las ML-estimaciones de p_{ri} son $\widehat{p}_{ri} = U_{ri}$ con valores $\widehat{p}_{ri} = u_{ri}$ para $r = 0,1,2$ y $i = 1,\ldots,n$. Además, $\mathscr{L}_c := \mathscr{L}(y) = 0$.

1.1.3. El modelo nulo

El *modelo nulo* es caracterizado por el supuesto de que para cada $r = 0,1,2$, todos los p_{ri} $(i = 1,\cdots,n)$ son considerados iguales; es decir, se tienen dos parámetros p_0 y p_1. En este caso, (1.1) será:

$$\mathscr{L}(p) \quad = \quad n[\overline{u}_0 \ln p_0 + \overline{u}_1 \ln p_1 + (1 - \overline{u}_0 - \overline{u}_1)\ln(1 - p_0 - p_1)] \tag{1.2}$$

siendo $\overline{u}_r = \frac{\sum_{i=1}^{n} u_{ri}}{n}$.

Teorema 1.2

En el modelo nulo, la ML-estimación de p_r es $\hat{p}_r = \overline{U}_r$ con valor $\hat{p}_r = \overline{u}_r$. Además, $\mathscr{L}_o := \mathscr{L}(\widehat{p}) < 0$ si y sólo si $0 < \overline{u}_0 + \overline{u}_1 < 1$.

1.1.4. El modelo saturado y supuesto

El modelo saturado está caracterizado por los siguientes supuestos:

1. Se supone que:

   a) Se tienen K variables explicativas $X_1,\cdots,X_K$ (algunas pueden ser numéricas y otras categóricas) con valores $x_{1i},\cdots,x_{Ki}$ para $i = 1,\cdots,n$ (fijadas u observadas por el estadístico, según sean variables determinísticas o aleatorias).

   b) Entre las n kuplas $(x_{1i},\cdots,x_{Ki})$ de los valores de la variable explicativa X haya J kuplas diferentes, definiendo las J poblaciones. Por tanto, $J \leq n$.

Notación Para cada población $j = 1,\cdots,J$ se denota:

- el número de observaciones Y_{ij} (o de observaciones U_{rij} en la categoría r) en cada población j por n_j, siendo $n_1 + \cdots + n_J = n$;
- Para cada $r = 0,1,2$ fijo, la suma de las n_j observaciones U_{rij} en j por $Z_{rj} := \sum_{i=1}^{n_j} U_{rij}$ con valor $z_{rj} = \sum_{i=1}^{n_j} u_{rij}$, siendo $\sum_{j=1}^{J} z_{rj} = \sum_{i=1}^{n} u_{ri}$.

Para mayor simplicidad en la escritura, se abreviará la j-ésima población $(x_{1j},\cdots,x_{Kj})$ por el símbolo $\star$.

2. Para cada $r = 0, 1, 2$ fijo, cada población $j = 1, \cdots, J$ y cada observación $i = 1, \cdots, n$ en j, se supone que:

   - $(U_{rij}|\star) \sim \mathscr{B}(1, p_{rj})$
   - Las variables $(U_{rij}|\star)$ son independientes entre sí

   A continuación, se oprimirá el símbolo $\star$.

   El supuesto 2 implica:

   a) Para cada $r = 0, 1, 2$ y cada población $j = 1, \cdots, J$, todos los p_{rij}, $i = 1, \cdots, n$, dentro de cada población j son iguales. Es decir, se tiene como parámetro el vector $2J$-dimensional $p = (p_{01}, p_{11}, \ldots, p_{0J}, p_{1J})^T$.

   b) Para cada $r = 0, 1, 2$ y cada población $j = 1, \cdots, J$,
   - $Z_{rj} \sim \mathscr{B}(n_j, p_{rj})$
   - Las variables Z_{rj} son independientes entre las poblaciones

En el modelo saturado, el logaritmo de la función de máxima verosimilitud será

$$\mathscr{L}(p) = \sum_{j=1}^{J} \left[z_{0j} \ln p_{0j} + z_{1j} \ln p_{1j} + (n_j - z_{0j} - z_{1j}) \ln(1 - p_{0j} - p_{1j}) \right] \tag{1.3}$$

Teorema 1.3

En el modelo saturado, las ML-estimaciones de p_{rj} son $\tilde{p}_{rj} = \frac{Z_{rj}}{n_j}$, con valores $\tilde{p}_{rj} = \frac{z_{rj}}{n_j}$, $\quad j = 1, \cdots, J$. Además,

$$\mathscr{L}(\tilde{p}) = \sum_{j=1}^{J} n_j [\tilde{p}_{0j} \ln \tilde{p}_{0j} + \tilde{p}_{1j} \ln \tilde{p}_{1j} + (1 - \tilde{p}_{0j} - \tilde{p}_{1j}) \ln(1 - \tilde{p}_{0j} - \tilde{p}_{1j})] \tag{1.4}$$

También se cumple: $\mathscr{L}_s := \mathscr{L}(\tilde{p}) < 0$ para $0 < \tilde{p}_{0j} + \tilde{p}_{1j} < 1$.

1.2 El modelo logístico

1.2.1. Supuestos

Se hacen los supuestos 1 y 2 de la sección 1.1.4, donde adicionalmente se supone que la matriz de diseño

$$C = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1K} \\ 1 & x_{21} & \cdots & x_{2K} \\ \vdots & \vdots & & \vdots \\ 1 & x_{J1} & \cdots & x_{JK} \end{pmatrix}$$

tiene rango completo $Rg(C) = 1 + K \leq J$. Para llegar a un modelo logístico se toma como referencia una de las categorías de la variable dependiente Y, digamos 0, y se hace el supuesto adicional

3.
$$g_1(x_j) = \ln\left(\frac{p_{1j}}{p_{0j}}\right) = \delta_1 + \beta_{11}x_{j1} + \cdots + \beta_{1K}x_{jK} \tag{1.5}$$

$$g_2(x_j) = \ln\left(\frac{p_{2j}}{p_{0j}}\right) = \delta_2 + \beta_{21}x_{j1} + \cdots + \beta_{2K}x_{jK} \tag{1.6}$$

donde $x_j := (1, x_{j1}, \ldots, x_{jK})^T$. Sea

$$\alpha = (\beta_1, \beta_2)^T = (\delta_1, \beta_{11}, \cdots, \beta_{1K}, \delta_2, \beta_{21}, \cdots, \beta_{2K})^T$$

el vector de los $2(1 + K)$ parámetros en el modelo.

Nótese que el supuesto sobre $Rg(C) = 1 + K$, hace identificable al parámetro α.

Para una observación x_j en la población j y para cada r, se cumple que:

$$p_{rj} = \frac{\exp\{g_r(x_j)\}}{\sum_{s=0}^{2} \exp\{g_s(x_j)\}}, \quad g_0 := 0 \tag{1.7}$$

El logaritmo de la función de verosimilitud se puede escribir en función de α como:

$$\mathscr{L}(\alpha) = \sum_{j=1}^{J} \left[z_{1j} g_1(x_j) + (n_j - z_{0j} - z_{1j}) g_2(x_j) - n_j \ln\left(\sum_{s=0}^{2} \exp\{g_s(x_j)\}\right)\right] \tag{1.8}$$

1.2.2. Relaciones entre el logístico y el saturado

Las ecuaciones del supuesto 3 de la sección 1.2 se pueden escribir vectorialmente así: $g_r = C\beta_r$, $r = 1, 2$, donde g es un vector J-dimensional con elementos $g(x_j)$, $j = 1, 2, \ldots, J$.

Con base en lo anterior, se pueden distinguir los dos siguientes casos:

1. $J = 1 + K$
   En este caso, C es una matriz invertible. Por lo tanto,

   $$\beta_r = C^{-1} g_r, \quad r = 1, 2$$

   Es decir, hay una relación uno a uno entre los parámetros del modelo saturado y los del logístico. O sea, los dos modelos expresan lo mismo.

   Particularmente, las ML-estimaciones de las probabilidades p_{rj} son iguales en ambos modelos: $\widehat{p}_{rj} = \widetilde{p}_{rj}$ para cada $j = 1, 2, \ldots, 1 + K$.

2. $J > 1 + K$
   En este caso, primero hay que calcular $\hat{\alpha}$ y a partir de éstas, se pueden calcular las $\widehat{p}_{rj}$ En general, resultan que $\widehat{p}_{rj} \neq \widetilde{p}_{rj}$.

1.2.3. Estimación de los parámetros logísticos

La estimación de los parámetros de este modelo se hace por el método de máxima verosimilitud. Esto ususlamente requiere procedimientos numéricos. La mayoría de los paquetes incluyen tal procedimiento, el cual también es *el método iterativo de Newton-Raphson.*

1.2.4. ODDS

Se define como

$$O_r(j) = \frac{p_{rj}}{1-p_{0j}}$$

1.2.5. Razones ODDS

Se define como

$$OR_r(i;j) = \frac{O_r(i)}{O_r(j)}$$

Para el caso en que la variable independiente sea binaria (codificada con 0 ó 1), utilizaremos la notación $OR_r = OR_r(1;0)$.

1.3 Pruebas de comparación de modelos y selección de modelos

En esta sección se presentan estadísticas para distintas pruebas de comparación de modelos:

- H_0: Logístico vs H_1: Saturado,
- H_0: Nulo vs H_1: Logístico,
- H_0: Logístico vs H_1: Completo,
- H_0: Submodelo vs H_1: Logístico,
- H_0: Submodelo con una variable explicativa menos vs H_1: Logístico.

Estas estadísticas tienen distribución asintótica chi-cuadrada.

1.3.1. Comparación de un modelo logístico con el modelo saturado correspondiente

Teorema 1.4

La LR-estadística de prueba (según el método de cocientes de funciones de verosimilitud) para la hipótesis

H_0: el modelo logístico (con $X_1,\ldots,X_K$),
vs la alternativa
H_1: el modelo saturado correspondiente (con sus J poblaciones)

es equivalente a la llamada deviación que tiene el modelo logístico del modelo saturado

$$D^*(M) := 2\ln\left(\frac{L(\tilde{p})}{L(\hat{\alpha})}\right) = 2[\mathscr{L}(\tilde{p}) - \mathscr{L}(\hat{\alpha})]$$

la cual tiene distribución asintótica chi-cuadrada con $\nu = 2[J-(1+K)]$ grados de libertad cuando $n\to\infty$ y J es fijo.

1.3.2. Comparación de un modelo logístico con el modelo nulo

Teorema 1.5

Para la hipótesis

H_0: el modelo nulo (sólo con el intercepto),
vs la alternativa
H_1: el modelo logístico (con $X_1, \ldots, X_K$)

la estadística de prueba es

$$D^*(0) = 2\ln\left(\frac{L(\hat{\alpha})}{L(\hat{\delta}_o)}\right) = 2[\mathscr{L}(\hat{\alpha}) - \mathscr{L}(\hat{\delta}_o)]$$

y tiene distribución asintótica chi-cuadrada con $2K$ grados de libertad cuando $J \to \infty$.

Aquí $\hat{\delta}_o = logit(\overline{Y})$ es la estimación de δ en el modelo nulo.

1.3.3. Comparación de un modelo logístico con el modelo completo

Teorema 1.6

Para la hipótesis

H_0: el modelo logístico (con $X_1, \ldots, X_K$),
vs la alternativa
H_1: el modelo completo (que no se basa en poblaciones)

la estadística de prueba es

$$D(M) := 2\ln\left(\frac{L(\hat{p})}{L(\hat{\alpha})}\right) = 2[\mathscr{L}(\hat{p}) - \mathscr{L}(\hat{\alpha})] = -2\mathscr{L}(\hat{\alpha})$$

y tiene distribución asintótica chi-cuadrada con $\nu = 2[n - (1 + K)]$ grados de libertad cuando $n \to \infty$.

1.3.4. Comparación de un modelo logístico con algún submodelo

Teorema 1.7

Para la hipótesis

H_0: un submodelo logístico con $X_1, \cdots, X_{\tilde{K}}$,
vs la alternativa
H_1: el modelo logístico con $X_1, \cdots, X_K$ con $\tilde{K} < K$,

(a) La estadística de prueba es equivalente a la estadística

$$D^*(L) := 2\log\left(\frac{L(\hat{\alpha})}{L(\hat{\alpha}_o)}\right) = 2[\mathscr{L}(\hat{\alpha}) - \mathscr{L}(\hat{\alpha}_o)]$$

Aquí: $\hat{\alpha} = (\hat{\delta}, \hat{\beta}_1, \cdots, \hat{\beta}_K)^T$ es la ML-estimación en el modelo logístico de la alternativa H_1 y $\hat{\alpha}_o = (\hat{\delta}_o, \hat{\beta}_{o1}, \cdots, \hat{\beta}_{o\tilde{K}})^T$ es la ML-estimación en el submodelo logístico de la hipótesis H_0.

(b) $D^*(L)$ tiene distribución asintótica chi-cuadrada con $2[K - \tilde{K}]$ grados de libertad cuando $J \to \infty$.

1.4 Comparación de un modelo logístico con un submodelo que tiene una variable explicativa menos

Teorema 1.8

Para la hipótesis

H_0: el submodelo (con $X_1, \cdots, X_K$ sin un X_k),
vs la alternativa
H_1: el modelo logístico (con $X_1, \cdots, X_K$)

se puede tomar, alternativamente, una de las dos estadísticas de pruebas siguientes (ambas tienen distribución asintótica chi-cuadrada con 1 grado de libertad cuando $J \to \infty$):

(a) $2\log\left(\frac{L(\hat{\alpha})}{L(\hat{\alpha}_o)}\right) = 2[\mathscr{L}(\hat{\alpha}) - \mathscr{L}(\hat{\alpha}_o)]$,

donde $\hat{\alpha}_o$ es la estimación bajo H_o

(b) $\frac{\hat{\beta}_k^2}{\hat{V}(\hat{\beta}_k)} = \left(\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}\right)^2$,

siendo $\hat{\beta}_k$ la estimación de β_k, para cada $k = 0, 1, \cdots, K$ en el modelo (bajo H_1) con su varianza estimada $\hat{V}(\hat{\beta}_k)$ y su error estándar $SE(\hat{\beta}_k) = \sqrt{\hat{V}(\hat{\beta}_k)}$

Observación:

1. En este teorema se está considerando el caso de datos no agrupados.
2. Con base en todas las pruebas parciales, para cada $k = 0, 1, \cdots, K$ se eliminará la variable explicativa que menor aporte individual tenga en la explicación. Es decir, la variable que tenga p-valor parcial más alto. Así, se sigue eliminando variable tras variable hasta que se rechacen todas las pruebas parciales (todas las tengan p-valores bajos).

1.5 ✍ Ejercicios

Para la solución de los siguientes ejercicios, téngase en cuenta los siguientes comentarios:

- Todos los datos mencionados aparecen en [18].
- Siempre debe detallar el análisis del conjunto de datos (con las variables especificadas) basado en lo realizado en el capítulo 1.
- Utilize Statgraphics y Excel para sus cálculos.
- Los resultados que no presente Statgraphics deben ser calculados con ayuda de Excel.
- Verifique cómo Statgraphics obtiene las estimaciones correspondientes, los logaritmos de las funciones de máxima verosimilitud, los estadísticos de pruebas, los p-valores, razones odds, intervalos de confianza (para pj, ODDS, razones ODDS, intercepto, pendiente, ...), etc.

1.1 Demuestre el teorema 1.1.2.

1.2 Demuestre el teorema 1.1.3.

1.3 Demuestre el teorema 1.1.4.

1.4 ¿En qué cambian los resultados de este capítulo si en vez de tener tres categorías 0, 1, 2, la variable de respuesta tuviese R categorías $0,1,2,\ldots,R-1$. Reformule nuevamente los resultados para este último caso.

1.5 Considere los datos **meexp**. Use un subconjunto de estos datos y ajuste un modelo de regresión logística multinomial. Por ejemplo, escoja sólo los primeros 200 individuos. Tome como variable dependiente a **ME** y use el valor 0 como referencia.

1.6 Considere los datos **lowbwt**. Defina una nueva variable BWT4 como

$$BWT4 = \begin{cases} 0 & \text{si } BWT > 3500, \\ 1 & \text{si } 3000 < BWT \leq 3500 \\ 2 & \text{si } BWT \leq 3000 \end{cases}$$

Use esta variable como de respuesta y ajuste un modelo de regresión logística multinomial.

A Apéndice

A.1 Pseudo-R-cuadrado

A.1.1. Introducción

Como punto de partida, recordemos que un (no-pseudo) R^2 es un estadístico generado en la regresión ordinaria. A menudo es usado como una medida de bondad de ajuste y es definida como

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2}$$

donde n es el número de observaciones en el modelo, $\overline{y}$ es la media de las observaciones y_i y $\widehat{y}_i$ es el valor predicho por el modelo.

Cuando analizamos los datos con el modelo de regresión logística, no existe un R^2 equivalente. Las estimaciones de los parámetros logísticos son estimaciones de máxima verosimilitud obtenidos por procesos iterativos. Ellos no son calculados con el fin de minimizar la varianza, así que no se puede aplicar la aproximación de la regresión ordinaria para la bondad de ajuste. Sin embargo para evaluar la bondad de ajuste de los modelos logísticos se han desarrollado los llamados PSEUDO-R^2, los cuales son medidas basadas en estimaciones de máxima verosimilitud, como ya se explicó anteriormente. Desafortunadamente, algunas de estas medidas no alcanzan 0 ó 1, muchos de ellos pueden tener valores diferentes y es importante reclacar que no se pueden interpretar como se haría con el R^2 de la regresión ordinaria. En la práctica, se habla de un ajuste aceptable para valores desde 0,2 y como bueno de 0,4.

Algunos de los pseudos-R^2 más importantes y utilizados son los de Mc-Fadden, Mc-Fadden ajustado, Cox-Snell, Nagelkerke, enttre otros. Explicaremos a continuación cada uno de ellos.

A.1.2. Pseudo-R^2 de Mc-Fadden

Se calcula con la fórmula:

$$R^2_{MF} = 1 - \frac{\mathscr{L}(\widehat{\alpha})}{\mathscr{L}(\widehat{\alpha}_0)}$$

Aquí $\mathscr{L}(\widehat{\alpha}_0)$ es el Log-verosimilitud en el modelo nulo y $\mathscr{L}(\widehat{\alpha})$ es el Log-verosimilitud en el modelo logístico. Toma sólo valores menores que 1.

A.1.3. Pseudo-R^2 de Mc-Fadden ajustado

En analogía al R^2 ajustado de la regresión lineal (que también tiene en cuenta el número de parámetros de la regresión y el número de casos), se propone la correspondiente versión para el Pseudo-R^2 de Mc-Fadden. Desafortunadamente, al respecto, la discusión no termina allí porque en la literatura podemos encontrar muchas otras propuestas. Una de estas propuestas es la siguiente:

$$R^2_{MFADJ} = 1 - \frac{\mathscr{L}(\widehat{\alpha}) - m}{\mathscr{L}(\widehat{\alpha}_0)}$$

Aquí m es el número de parámetros en el modelo (sin la constante). Esta medida puede tomar valores negativos.

A.1.4. Pseudo-R^2 de Cox-Snell

Se calcula con la fórmula:

$$R^2_{CS} = 1 - \left(\frac{L(\widehat{\alpha}_0)}{L(\widehat{\alpha})}\right)^{2/n}$$

Aquí $L(\widehat{\alpha}_0)$ es la verosimilitud (no el Log-verosimilitud) en el modelo nulo y $L(\widehat{\alpha})$, la del modelo logístico.

A.1.5. Pseudo-R^2 de Mc-Fadden

Se calcula con la fórmula:

$$R^2_N = \frac{R^2_{CS}}{1 - [L(\widehat{\alpha}_0)]^{2/n}}$$

Bibliografía

[1] AGRESTI, A., *Categorical data analysis*. John Wiley and Sons, Inc., New York, 1990.

[2] BARBOSA, R.; LLINÁS, H., *Procesos estocásticos con aplicaciones*, Barranquilla: Editorial Universidad del Norte, 2013.

[3] HOSMER, D. and LEMESHOW S., *Applied Logistic Regression*, Segunda edición, John Wiley and Sons, 2000.

[4] KALB, K. y KONDER, P., *Una visión histórica del concepto moderno de integral*, Barranquilla: Editorial Universidad del Norte, 2006 (editor: Dr. rer. nat. Humberto LLinás).

[5] KLEINBAUM, D. and KLEIN, M., *Logistic Regression: A self Learning Text*, Segunda edición, Ed. Springer, 2002.

[6] LLINÁS, H.; ROJAS, C., *Estadística descriptiva y distribuciones de probabilidad*. Barranquilla: Ediciones Uninorte, 2005.

[7] LLINÁS, H., *Precisiones en la teoría de los modelos logísticos*, Revista Colombiana de Estadística, Volumen 29, Número 2, pág. 239-265, 2006.

[8] LLINÁS, H., *Estadística inferencial*. Barranquilla: Ediciones Uninorte, 2006.

[9] LLINÁS, H., *Estadística inferencial*, Barranquilla: Editorial Universidad del Norte, 2006.

[10] LLINÁS, H., *Medida e integración*. Barranquilla: Editorial Universidad del Norte, 2007.

[11] LLINÁS, H., *Applet: La ley de los grandes números*. Se puede encontrar en el siguiente link: http://ylang-ylang.uninorte.edu.co/Objetos/Estadistica/LeyDeGrandesNumeros/index.html

[12] LLINÁS, H., *Applets de estadística*, 2007. Se puede encontrar en el siguiente link: *http://ylang-ylang.uninorte.edu.co:8080/drupal/?q=node/238*

[13] LLINÁS, H.; ALONSO, J. FLÓREZ, K., *Introducción a la estadística con aplicaciones en Ciencias Sociales*, Barranquilla: Editorial Universidad del Norte, 2012.

[14] LLINÁS, H., *Introducción a la estadística matemática*, Barranquilla: Editorial Universidad del Norte, 2014.

[15] LLINÁS, H., *Introducción a la teoría de probabilidad*, Barranquilla: Editorial Universidad del Norte, 2014.

[16] NELDER, J.A. and WEDDERBURN, R.W.M., *Generalized linear models*. The Journal of the Royal Statistical Society, serie A 135, pág.370-384, 1972.

[17] PÉREZ, C., *Estadística práctica con Statgraphics*. España: Prentice Hall, 2002.

[18] Página web de datos estadísticos del Institute for Digital Research and Education (IDRE) de la Universidad de California en Los Angeles (UCLA): *https://stats.idre.ucla.edu/*. En especial, consultar: *https://stats.idre.ucla.edu/other/examples/alr2/*