

Influencia de los valores locales en los diseños cD-óptimos para el modelo logístico

Influence of local values on cD-optimal designs for logistic models

Víctor Ignacio López-Ríos¹ y David Felipe Sosa-Palacio²

Recibido: 14 de diciembre de 2017

Aceptado: 10 de mayo de 2018

Cómo citar / How to cite

V. I. López-Ríos, y D. F. Sosa-Palacio, Influencia de los valores locales en los diseños cD-óptimos para el modelo logístico. *TecnoLógicas*, vol. 21, no. 43, pp. 147-157, 2018.

© Instituto Tecnológico
Metropolitano

Este trabajo está licenciado bajo una
Licencia Internacional Creative
Commons CC BY-NC-SA



¹ Ph.D. en Estadística, Profesor Asociado, Escuela de Estadística, Grupo de Investigación en Estadística, Universidad Nacional de Colombia, Medellín-Colombia, vilopez@unal.edu.co

² MSc. en Ciencias-Estadística, Escuela de Estadística, Universidad Nacional de Colombia, Medellín-Colombia, dfsosa@unal.edu.co

Resumen

En el diseño de experimentos es común el no uso de criterios para determinar los tratamientos y el número de réplicas que se deben realizar para la obtención de una buena estimación de los parámetros del modelo, debido principalmente al desconocimiento de estos y en muchas otras ocasiones por la dificultad en implementarlos. Los diseños óptimos tratan de resolver esta falencia al dar condiciones experimentales óptimas y los niveles de los factores donde se debe medir la respuesta, con el fin de obtener una mejora en la calidad de la inferencia estadística a un menor costo. En la búsqueda de diseños óptimos se utilizan criterios de optimalidad, los cuales son función de la matriz de información de Fisher. Uno de los problemas de estimación más frecuente en los modelos no lineales es la especificación de los valores locales para los parámetros del modelo, necesarios para la optimización del criterio de optimalidad. En este artículo se realiza un estudio de robustez de los diseños óptimos obtenidos en el modelo logístico, al considerar perturbaciones en los valores locales de los parámetros, con el fin de proporcionar al investigador un rango de maniobrabilidad en la selección de los valores locales y garantizando que el diseño óptimo resultante no pierda una eficiencia considerable con respecto al valor de referencia. Para ello, a partir de los datos de un ejemplo, se encuentran las eficiencias de cada uno de los diseños obtenidos con relación al valor sin perturbar; se construyen los diseños cD-óptimos locales para la estimación de la varianza del logit, se determinó que la magnitud de la perturbación en los diseños cD-óptimos locales obtenidos alcanzan una eficiencia alrededor de un 70 %, con un radio de 0.04 de perturbación sobre el valor de referencia.

Palabras clave

Regresión Logística, diseños óptimos compuestos, criterios de optimalidad, robusticidad.

Abstract

When experiments are designed, it is uncommon to use criteria to determine the treatments and number of replicates that should be conducted to properly estimate the parameters in the model under study. This is mainly due to a lack of knowledge of said criteria and, in many other cases, to the difficulty of interpreting them. Optimal designs try to overcome this difficulty by providing optimal experimental conditions and factor levels whose response should be evaluated in order to improve the quality of the statistical inference at a lower cost. Said designs also use optimality criteria, which are functions of the Fisher-information matrix. One of the most common estimation problems in nonlinear models is specifying local values for the parameters of the model, which is necessary to minimize the optimality criterion. This work examines the robustness of optimal designs obtained from logistic models when perturbations in the local values of the parameters are considered. The objective is to provide researchers with a wide range of action to select local values while guaranteeing that the efficiency of the resulting optimal designs is not considerably compromised compared to reference values. For that purpose, based on the data of an example, the efficiency of each resulting design was contrasted with the efficiency of unperturbed values. Furthermore, cD-optimal designs were created to estimate the logit variance. The impact of said perturbations on local cD-optimal designs resulted in an efficiency of approximately 70%, with a range of 0.04 of perturbation above the reference value.

Keywords

Logistic Regression, compound designs, optimality criteria, robustness.

1. INTRODUCCIÓN

Los diseños óptimos aproximados fueron propuestos originalmente por Kiefer (1959) [1], principalmente en modelos de regresión. El origen del trabajo formal sobre este tipo de diseños se remonta al realizado por Smith (1918) [2]. La teoría de los diseños óptimos tiene como base fundamental la teoría del diseño experimental clásico, donde el investigador determina en qué niveles del factor o factores debe hacer el experimento, junto con el número de réplicas, para obtener una mejor estimación de los parámetros del modelo bajo estudio o una función lineal o no lineal de estos, con un determinado criterio estadístico. Por ejemplo, el criterio D-optimalidad proporciona el diseño donde se minimiza el volumen del elipsoide de confianza, mientras que el criterio c-optimalidad proporciona los puntos experimentales donde se minimiza la varianza de una combinación lineal o de una función no lineal del vector de parámetros estimados. Si se considera una combinación convexa de ambos criterios, se obtiene el criterio cD-optimalidad y el diseño obtenido permite estimar simultáneamente el vector de parámetros y una función de éste.

En cualquier investigación que involucre la realización de un experimento, el investigador debe determinar previamente las variables de interés. Posteriormente, debe considerar un modelo estadístico que explique de manera adecuada el comportamiento entre las variables explicativas x 's y su relación con la variable de interés o respuesta Y . Generalmente, dicho modelo tiene la forma de los modelos de regresión: $Y = \eta(x, \theta) + \epsilon$ donde $\eta(x, \theta)$ es una función (lineal o no) del vector de parámetros θ y del vector de variables explicativas x , donde ϵ es el término de error, con los supuestos usuales de independencia y homogeneidad de varianza. Más aun, el investigador debe determinar en qué niveles o valores de las

variables explicativas se debe medir la variable respuesta. Si bien, la selección de estos niveles se puede apoyar en estudios previos o en la experiencia del investigador, dicha decisión no deja de ser subjetiva. En este sentido, los diseños óptimos surgen como un medio para conocer en qué niveles de las variables explicativas es posible medir la variable respuesta para obtener una mejor estimación de los parámetros del modelo bajo estudio a un menor costo [3].

En los modelos no lineales, en particular en el modelo de regresión logística, donde la variable respuesta es binaria o dicótoma, los diseños que se obtienen dependen de la elección de los valores para los parámetros del modelo, conocidos como valores locales, independiente del criterio utilizado; para trabajos en esta dirección se pueden consultar: [4], [5], [6] y [7], entre otros.

En la práctica, el investigador no conoce de antemano el valor verdadero de los parámetros, por lo que surgen las siguientes preguntas: ¿cuál es el efecto en la eficiencia de un diseño cD-óptimo si el investigador elige un valor que no sea el correcto? ¿Es posible determinar una región donde el diseño cD-óptimo pierda poca eficiencia con respecto al diseño cD-óptimo de referencia? Con el fin de responder las preguntas anteriores se procede a realizar un estudio de robustez de los diseños cD-óptimos locales, a partir de la realización de perturbaciones, en coordenadas polares, sobre los valores locales de los parámetros.

En la sección 2 se presentan algunos conceptos básicos de los diseños óptimos. En la sección 3 se realiza un estudio de robustez, allí se presenta el procedimiento propuesto para la realización del estudio; y por último, se presentan los resultados obtenidos y las conclusiones.

2. CONCEPTOS BÁSICOS DE LOS DISEÑOS ÓPTIMOS

2.1 Supuestos y Modelo

Los diseños óptimos dependen, entre otros aspectos, del modelo bajo estudio seleccionado previamente por parte del investigador. En este artículo se considera el modelo de regresión presentado en la ecuación (1):

$$Y = \eta(x, \theta) + \varepsilon \quad (1)$$

Donde Y denota la variable respuesta del modelo, x es la variable explicativa, θ es un vector de p -parámetros, $\eta(x, \theta)$ es una función (lineal o no) del vector de parámetros, y ε es el error aleatorio con distribución normal con media cero y varianza constante.

Un diseño ξ se define como una medida de probabilidad con soporte finito, representado en la ecuación (2).

$$\xi = \begin{bmatrix} x_1 & \dots & x_n \\ w_1 & \dots & w_n \end{bmatrix} \quad (2)$$

Con x_1, \dots, x_n y w_1, w_2, \dots, w_n los puntos de soporte y los pesos del diseño, tales que $\sum_{i=1}^n w_i = 1$ [8]. Si los pesos w_i son valores reales entre 0 y 1, los diseños obtenidos se denominan diseños continuos o aproximados.

2.2 Matriz de información

La matriz de información permite cuantificar la información aportada por un diseño dado. La expresión de esta se encuentra en la ecuación (3):

$$M(\xi, \theta) = \sum_{i=1}^n w_i f(x_i, \theta) f^T(x_i, \theta) \quad (3)$$

Donde $f^T(x_i, \theta) = \frac{\partial \eta(x_i, \theta)}{\partial \theta}$ es el vector gradiente para el i -ésimo punto de soporte, x_i [3]. En este caso la matriz de

información depende del vector de parámetros θ .

2.3 Criterios de optimalidad

La construcción de diseños óptimos se basa en la optimización de alguna función de valor real de la matriz de información $\Psi(\xi) := \Psi(M(\xi))$. La estructura de dicha función es conocida como criterio de optimalidad [4]. Los criterios de optimalidad de interés son aquellos que tienen alguna interpretación estadística. Una de las características importantes que debe cumplir un criterio de optimalidad es la propiedad de convexidad; es decir, si $M(\xi) = (1 - \alpha)M(\xi_1) + \alpha M(\xi_2)$, con ξ_1 y ξ_2 diseños, entonces, para cualquier $0 \leq \alpha \leq 1$, se cumple la desigualdad dada en la ecuación (4):

$$\Psi[M(\xi)] \leq (1 - \alpha)\Psi[M(\xi_1)] + \alpha\Psi[M(\xi_2)] \quad (4)$$

Con el fin de facilitar los cálculos numéricos para encontrar el diseño D-óptimo se minimiza el funcional equivalente: $\Psi[M(\xi, \theta)] = -\log |M(\xi, \theta)|$ [9].

2.3.1 Criterio D-optimalidad

Uno de los criterios de optimalidad más usado es el criterio D-optimalidad [10]. Este consiste en hallar el diseño que minimice la varianza generalizada del vector de parámetros estimados, $|Var(\hat{\theta})|$, o equivalentemente a aquel diseño que minimice $|M^{-1}(\xi, \theta)|$. Cabe anotar que, si los errores del modelo son normales, una interpretación geométrica de este criterio es que el diseño hallado minimiza el volumen del elipsoide de confianza asociado al vector de parámetros θ [8].

2.3.2 Criterio c-Optimalidad

En ocasiones, el interés del investigador no es precisamente la estimación de todos los parámetros del modelo, sino de una parte o combinaciones de algunos, es decir,

si se quiere estimar una combinación lineal de $\theta_1, \dots, \theta_p$, equivalente a $c^T \theta$ donde $c^T = [c_1, \dots, c_p]$ es un vector de constantes conocidos. El criterio que tiene como objetivo la minimización de la varianza del estimador de esta combinación lineal se conoce como c-optimalidad. Este criterio se generaliza al hallar el diseño que minimiza la varianza de una función, no necesariamente lineal, de los parámetros estimados del modelo, $g(\hat{\theta})$, a partir de una linealización de primer orden de esta. La función objetivo en cualquiera de los dos casos, se presenta en la ecuación (5) [11].

$$\Psi_c(\xi, \theta) = c^T M^{-1}(\xi, \theta) c \quad (5)$$

2.4 Diseños compuestos

En algunos criterios de optimalidad, por ejemplo, c-optimalidad, los diseños óptimos encontrados, en general, tienen un número de puntos de soporte inferior al número de parámetros del modelo, es decir los diseños son singulares, lo cual no permite hacer la estimación de los parámetros. Una forma de resolver este problema, investigada en este artículo, es hallar el diseño que optimice una ponderación de dos criterios de optimalidad, buscando que aquel diseño resultante satisfaga ambos criterios simultáneamente. Estos diseños se conocen con el nombre de diseños óptimos compuestos.

Sea $\Psi_i(M(\xi, \theta)), i = 1, 2$, dos criterios de optimalidad definidos en una región χ y sea a , un real positivo menor a uno, entonces la combinación lineal dada en la ecuación (6):

$$\Psi(\xi, \theta) = a\Psi_1(M(\xi, \theta)) + (1 - a) \Psi_2(M(\xi, \theta)) \quad (6)$$

Es un criterio de optimalidad, conocido como el criterio compuesto de Ψ_1 y Ψ_2 . Si $a=0.5$, y $\Psi_1(M(\xi, \theta))$ y $\Psi_2(M(\xi, \theta))$ son los criterios D- y c-optimalidad respectivamente, se obtiene el criterio

compuesto cD-optimalidad igualmente ponderados, ver ecuaciones (7), (8) y (9):

$$\Psi_{cd}(\xi, \theta) = 0.5 \Psi_D(M(\xi, \theta)) + 0.5 \Psi_c(M(\xi, \theta)) \quad (7)$$

con:

$$\Psi_D(M(\xi, \theta)) = \log|M(\xi, \theta)|^{-1} \quad (8)$$

$$\Psi_c(M(\xi, \theta)) = \log[c^T M^{-1}(\xi, \theta) c] \quad (9)$$

Ambos en escala logarítmica, como resultado de considerar el logaritmo de la media geométrica de ambos criterios de optimalidad.

2.5 Teorema de Equivalencia para el criterio cD-optimalidad

Kiefer y Wolfowitz en 1959 [12] presentan el primer teorema de equivalencia entre los diseños D y G-óptimo. El teorema establece la equivalencia entre D-optimalidad, usado para optimizar la estimación de los parámetros del modelo, y G-optimalidad, donde el interés se centra en la predicción de la respuesta de manera óptima. Esta equivalencia es válida para diseños continuos o aproximados, para diseños exactos es posible que no se cumpla. El teorema de equivalencia facilita la construcción de algoritmos para encontrar los diseños óptimos de los criterios presentes ya que se puede caracterizar un diseño con otro.

Una versión general de este teorema, para cualquier par de criterios, se desarrolla en [13], incluso para modelos no lineales. A continuación, se enuncia el teorema de equivalencia para cD-optimalidad, cuya deducción es una consecuencia de la versión generalizada presentada en [13].

Resultado

Un diseño ξ es cD-óptimo si y solo si se cumple la ecuación (10):

$$D(x, \xi, \theta) := aD_1(x, \xi, \theta) + (1 - a)D_2(x, \xi, \theta) \leq 0 \quad (10)$$

Con $D_1(x, \xi, \theta)$ y $D_2(x, \xi, \theta)$ dadas en las ecuaciones (12) y (13). Además para todo $x \in \chi$ y en los puntos de soporte del diseño se satisface la ecuación (11):

$$D(x, \xi, \theta) = 0 \quad (11)$$

con:

$$D_1(x, \xi, \theta) = f^T(x, \theta)M^{-1}(\xi, \theta)f(x, \theta) - p \quad (12)$$

$$D_2(x, \xi, \theta) = \frac{[f^T(x, \theta)M^{-1}(\xi, \theta)c]^2}{c^T M^{-1}(\xi, \theta)c} - 1 \quad (13)$$

Las funciones $D_i(x, \xi, \theta)$, $i=1,2$, se conocen como las funciones de sensibilidad o derivadas direccionales de los criterios D- y c- optimalidad [14].

2.6 Diseños óptimos locales

Introducidos por [15], son los primeros diseños que aparecieron para el caso no lineal. Consiste en dar inicialmente un valor para θ , que esté cercano al valor verdadero del parámetro. Esta asignación a priori dependerá del conocimiento previo que el investigador tenga acerca del problema bajo estudio, puede optar por estimaciones a partir de datos previos, o estudios similares. Es de vital importancia una buena selección de los valores iniciales con el fin de garantizar un mejor experimento para la estimación de los parámetros. Una vez se tenga el valor o los valores supuestos para θ se construyen diseños óptimos a partir de funcionales de la matriz de información. Los diseños resultantes son diseños óptimos locales.

Varios autores han usado esta alternativa para la construcción de los diseños, entre ellos: [16], [17], [18], entre otros.

3. ESTUDIO DE ROBUSTEZ DE LOS DISEÑOS cD-ÓPTIMOS PARA EL MODELO LOGÍSTICO

Con el fin de dar respuesta a los interrogantes presentados en la sección 1, se hizo un estudio de robustez de los diseños cD-óptimos a la elección de los valores locales, para esto se consideran diferentes perturbaciones para los valores locales a partir de un valor de referencia para θ tomado como el valor correcto. Se hacen las perturbaciones usando posiciones estratégicas en un círculo de un radio dado y con centro en el valor de referencia. Lo anterior se hace usando transformaciones en coordenadas polares, ver ecuaciones (15) y (16). Se cuantifica las cualidades de los nuevos diseños obtenidos calculando la eficiencia con respecto al diseño obtenido usando el valor de referencia. El estudio se realiza con el modelo logístico, ver ecuación (14), usando los datos reales de un ejemplo obtenido en la literatura [19].

3.1 Metodología

A continuación, se esquematiza la metodología utilizada.

- a. Para un conjunto de n-datos (x_i, Y_i) se ajusta el modelo de regresión logística dado en la ecuación (14):

$$P[Y_i = 1|X = x_i] := E(Y|x_i) := \pi(x_i, \theta) = \frac{1}{1 + \exp(-\gamma(x_i - \mu))} \quad (14)$$

y $\theta^T = (\gamma, \mu)$ es el vector de parámetros. Denote por $\theta_0^T = (\gamma_0, \mu_0)$ las estimaciones resultantes para γ, μ considerados como los valores locales o de referencia para el vector de parámetros θ .

- b. Se halla el diseño cD-óptimo local para θ .
- c. Se calculan los nuevos valores de θ , $\theta_i^T = (\gamma_i, \mu_i)$, usando las ecuaciones (15) y (16), expresiones en coordenadas polares.

$$\gamma_i = r_j \cos(\theta_k) + \gamma_0 \tag{15}$$

$$\mu_i = r_j \sin(\theta_k) + \mu_0 \tag{16}$$

Con r_j , el radio, iniciando en 0.02 hasta un valor máximo de 0.10, con incrementos de 0.02, denotado por $r_j = 0.02:0.10:0.02$ y en forma análoga para el ángulo: $\theta_k = 0^\circ:330^\circ:30^\circ$.

- d. Para cada valor de θ_i se halla el diseño cD-óptimo local, y se calcula la eficiencia usando la expresión dada en la ecuación (17) y el criterio cD-optimalidad de la ecuación (7):

$$Ef_{cD}(\xi) = \frac{\Psi_{cD}(\xi^*)}{\Psi_{cD}(\xi)} \tag{17}$$

Este valor se encuentra entre cero y uno, a medida que este valor esté cercano a uno indicará que el diseño ξ es tan competitivo como el diseño ξ^* -cD-óptimo.

3.2 Caso de estudio

En esta sección se implementa la metodología presentada en la sección 3.1 a partir de datos encontrados en la literatura [19], listan edades en años, y la presencia o ausencia de enfermedad coronaria (EC) para 100 individuos seleccionados para participar en un estudio. En la Tabla 1 se listan las primeras 10 observaciones del estudio en mención. La variable respuesta, en este caso es EC, toma valores cero o

uno, donde el valor cero indica ausencia, y el valor uno que existe presencia de enfermedad coronaria en el individuo. El interés particular es establecer la relación entre la edad del individuo y el desarrollo de la enfermedad coronaria.

Tabla 1. Conjunto de datos (10 primeras observaciones) para el caso de estudio, EC: Presencia o no de enfermedad coronaria. Fuente: [19]

Edad (años)	20	23	24	25	25	26	26	28	28	29
EC	0	0	0	0	1	0	0	0	0	0

3.2.1 Ajuste del Modelo Logístico

En primer lugar, se ajusta un modelo logístico reparametrizado dado en la ecuación (14), con γ y μ los parámetros a estimar. Usando el software estadístico R, R Development Core Team (2013) [20] con la función nls, se hallan como estimaciones para los parámetros γ y μ , los valores 0.1060 y 47.972, respectivamente.

En la Fig. 1 se muestra el ajuste del modelo reparametrizado de la forma mostrado en la ecuación (14), después de considerar los datos agrupados, se consideraron los pares de datos (x_i, p_i) , con p_i proporción de individuos con EC en la marca de clase de la edad x_i . En lo que sigue se utilizan como valores locales para θ , $\theta_0^T = (0.1060, 47.972)$. Con este valor se halla el diseño D-óptimo local, diseño que permite estimar el vector de parámetros del modelo.

3.2.2. Función de interés a estimar

Además de estimar los parámetros del modelo interesa estimar una función no lineal de ellos, por ejemplo, la varianza de la respuesta predicha en un punto x_0 , elegido como la edad a la cual un individuo tiene una probabilidad del 50 % de desarrollar enfermedad coronaria, en el caso de estudio corresponde a $x_0 = 44$ años. La ecuación (18) muestra esta varianza.

$$g(\theta) = V(\hat{y}(x_0, \theta)) = \pi(x_0, \theta)[1 - \pi(x_0, \theta)] \tag{18}$$

Con gradiente asociado $c^T = \left(\frac{\partial g(\theta)}{\partial \gamma} \quad \frac{\partial g(\theta)}{\partial \mu} \right)$ donde las derivadas con respecto a γ y μ , se muestran en las ecuaciones (19), (20) y (21) respectivamente.

$$\frac{\partial g(\theta)}{\partial \gamma} = x_0 \{1 + \exp[-\gamma(x_0 - \mu)]\}^{-2} \quad (19)$$

$$\frac{\partial g(\theta)}{\partial \mu} = \frac{\partial \pi(x_0, \theta)}{\partial \mu} - 2\pi(x_0, \theta) \frac{\partial \pi(x_0, \theta)}{\partial \mu} \quad (20)$$

$$\frac{\partial \pi(x_0, \theta)}{\partial \mu} = -\pi^2(x_0, \theta) \gamma \exp[-\gamma(x_0 - \mu)] \quad (21)$$

Con el fin de hallar diseños no singulares que permitan al mismo tiempo estimar los parámetros del modelo y estimar la función de interés, se opta por usar diseños óptimos compuestos, donde se considera el mismo peso o importancia para cada uno de los criterios considerados, el diseño óptimo local encontrado es observable como ξ_{cD} en la expresión mostrada en (22). Además, se usó el teorema general de equivalencia con el fin de determinar si en efecto, el diseño compuesto obtenido, dado en la expresión

(22), cumplía con los requisitos para ser óptimo.

$$\xi_{cD} = \begin{pmatrix} 31.04 & 64.89 \\ 0.59 & 0.41 \end{pmatrix} \quad (22)$$

En la Fig. 2 se observa la derivada direccional de los puntos obtenidos para el diseño compuesto; se puede notar que en los puntos del diseño la derivada se hace cero y es negativa en los demás puntos de la región experimental, por lo que el teorema de equivalencia es válido.

3.2.3 Resultados y análisis

En las Tablas 2, 3 y 4 se muestran las perturbaciones correspondientes a las radios 0.02, 0.04 y 0.10 con sus respectivas eficiencias. Los valores perturbados corresponden a los puntos de un círculo con radio r , centrado en el valor local de los parámetros. En este punto se obtienen diseños con una eficiencia igual o similar a la arrojada por el diseño óptimo.

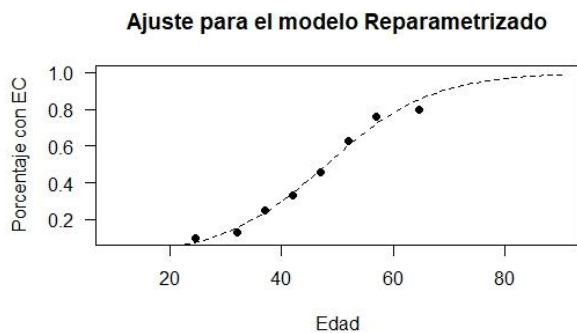


Fig. 1. Ajuste para el modelo logístico mostrado en (14). Fuente: autor.

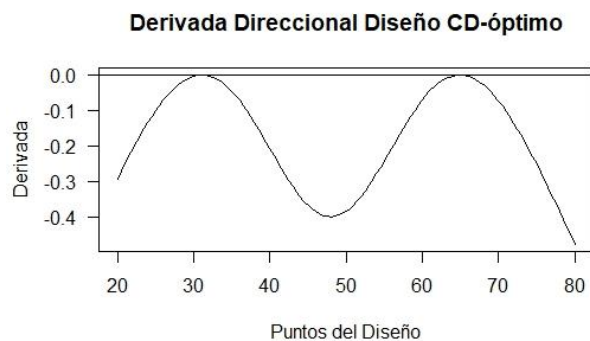


Fig. 2. Derivada direccional o función de sensibilidad del diseño cD-óptimo. Fuente: autor.

Influencia de los valores locales en los diseños cD-óptimos para el modelo logístico

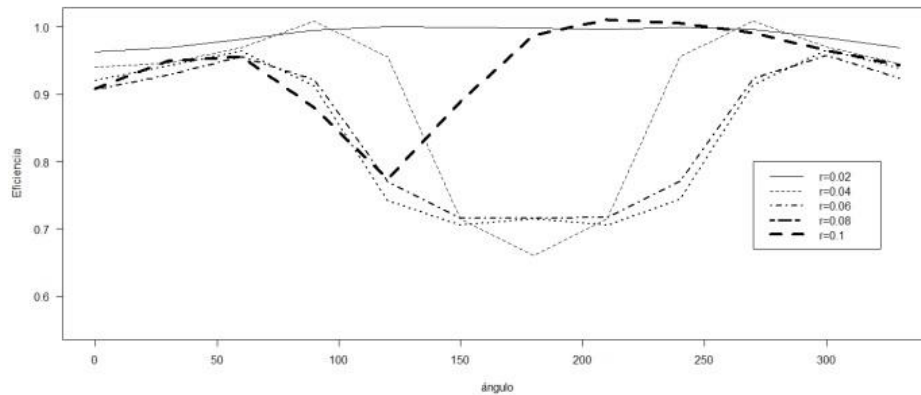


Fig. 3. Gráfico de eficiencias para los diseños óptimos para los distintos valores del radio. Fuente: autor.

Tabla 2. Diseños cD-óptimos y eficiencias para $r=0.02$ y diferentes ángulos.

Fuente: autor.

r	θ	$\xi = [x_1 x_2 w_1 w_2]$	Eficiencias
0.02	0	[33.63 62.31 0.58 0.42]	0.965
	30	[33.34 62.62 0.59 0.41]	0.966
	60	[32.46 63.51 0.59 0.41]	0.981
	90	[31.06 64.91 0.59 0.41]	0.999
	120	[31.06 64.91 0.59 0.41]	0.999
	150	[31.05 64.90 0.59 0.41]	0.999
	180	[31.04 64.89 0.59 0.41]	1.000
	210	[31.62 64.29 0.59 0.41]	0.992
	240	[31.03 64.87 0.59 0.41]	1.000
	270	[31.02 64.87 0.59 0.41]	1.000
	300	[32.42 63.48 0.59 0.41]	0.981
	330	[33.32 62.60 0.59 0.41]	0.969

Tabla 3. Diseños cD-óptimos y eficiencias para $r=0.04$ y diferentes ángulos.

Fuente: autor.

r	θ	$\xi = [x_1 x_2 w_1 w_2]$	Eficiencias
0.04	0	[35.54 62.31 0.56 0.44]	0.941
	30	[35.10 62.62 0.57 0.43]	0.947
	60	[33.66 63.51 0.58 0.42]	0.965
	90	[31.08 64.91 0.57 0.43]	0.999
	120	[27.40 64.91 0.61 0.39]	1.000
	150	[23.55 64.90 0.63 0.37]	0.667
	180	[21.77 64.89 0.63 0.37]	0.695
	210	[23.51 64.29 0.63 0.37]	0.667
	240	[27.33 64.87 0.61 0.39]	1.000
	270	[31.00 64.87 0.59 0.41]	1.000
	300	[33.59 63.48 0.58 0.42]	0.965
	330	[35.06 60.80 0.57 0.43]	0.947

Tabla 4. Diseños cD-óptimos y eficiencias para $r=0.10$ y diferentes ángulos.

Fuente: autor

r	θ	$\xi = [x_1 \ x_2 \ w_1 \ w_2]$	Eficiencias
0.10	0	[39.10 56.83 0.55 0.45]	0.900
	30	[38.54 57.49 0.55 0.45]	0.946
	60	[36.40 59.70 0.56 0.44]	0.959
	90	[31.14 64.99 0.59 0.41]	1.000
	120	[20.00 78.94 0.64 0.36]	0.554
	150	[20.00 80.00 0.64 0.36]	1.000
	180	[20.00 80.00 0.64 0.36]	1.000
	210	[20.00 80.00 0.64 0.36]	1.000
	240	[20.70 78.87 0.64 0.36]	1.000
	270	[30.94 64.79 0.59 0.41]	0.999
	300	[36.23 59.53 0.56 0.44]	0.958
	330	[38.44 57.39 0.55 0.45]	0.945

Al realizar las perturbaciones con un radio $r = 0.04$, ver Fig. 3, se obtienen eficiencias cercanas al diseño óptimo, la menor eficiencia en este caso es de aproximadamente 70 %. En la misma figura se observan las eficiencias obtenidas para los otros radios, r , se puede ver que para los radios 0.04, 0.06 y 0.08 las menores eficiencias se obtienen en el segundo cuadrante del plano cartesiano, ángulos entre 90° y 180° .

4. CONCLUSIONES

Se propuso un procedimiento para el análisis del efecto que tiene la elección de los valores locales en la obtención de los diseños cD-óptimos. Este procedimiento considera un gran margen de posibilidades para la elección de los valores locales.

A partir de un ejemplo con datos reales, tomados de la literatura, se implementó el procedimiento usando el modelo logístico y se pudieron obtener los siguientes resultados:

-Los diseños cD-óptimos locales para el modelo logístico, tuvieron eficiencias al menos del 70 % para un radio de 0.04.

-La magnitud de la eficiencia disminuye a medida que el radio de perturbación es mayor y también depende del ángulo. Para

un radio de perturbación de 0.10 se obtiene la menor eficiencia en el segundo cuadrante, esta es de un 55 %, lo que indica que a medida que los valores locales no sean los adecuados el diseño tiene menos eficiencia.

-Se puede identificar una región donde los diseños óptimos obtenidos, de acuerdo con la selección de los valores locales, presentan eficiencias altas. Para el radio $r = 0.04$, se identifica el primer cuadrante y en el cuarto cuadrante donde las eficiencias son cercanas a la del diseño de referencia y para el segundo y tercer cuadrante las eficiencias no disminuyen drásticamente donde la menor es de un 70 % aproximadamente con respecto al diseño de referencia.

-A partir del estudio realizado, la elección del valor local no necesita ser tan precisa, al poder identificar regiones donde la eficiencia de los diseños obtenidos es comparable a la del diseño óptimo de referencia.

5. AGRADECIMIENTOS

Los autores agradecen a la Escuela de Estadística de la Universidad Nacional de Colombia por el apoyo recibido para la elaboración de este artículo. También

agradecen los comentarios y observaciones de los dos jurados anónimos que ayudaron a mejorar la presentación de este artículo.

6. REFERENCIAS

- [1] J. Kiefer, "Optimum Experimental Designs," *J. R. Stat. Soc.*, vol. 21, no. 2, pp. 272–319, 1959.
- [2] K. Smith, "On the Standard Deviations of Adjusted and Interpolated Values of an Observed Polynomial Function and its Constants and the Guidance they give Towards a Proper Choice of the Distribution of Observations," *Biometrika*, vol. 12, no. 1, p. 1, Nov. 1918.
- [3] V. I. López and R. Ramos, "Una introducción a los diseños óptimos," *Rev. Colomb. Estadística*, vol. 30, no. 1, pp. 37–51, 2007.
- [4] T. Mathew and B. K. Sinha, "Optimal designs for binary data under logistic regression," *J. Stat. Plan. Inference*, vol. 93, no. 1–2, pp. 295–307, Feb. 2001.
- [5] P. Parsa Maram and H. Jafari, "Bayesian D-optimal design for logistic regression model with exponential distribution for random intercept," *J. Stat. Comput. Simul.*, vol. 86, no. 10, pp. 1856–1868, Jul. 2016.
- [6] L. M. Haines, G. Kabera, P. Ndlovu, and T. E. O'Brien, "D-optimal Designs for Logistic Regression in Two Variables," in *mODa 8 - Advances in Model-Oriented Design and Analysis*, Heidelberg: Physica-Verlag HD, pp. 91–98.
- [7] M. A. Heise and R. H. Myers, "Optimal Designs for Bivariate Logistic Regression," *Biometrics*, vol. 52, no. 2, p. 613, Jun. 1996.
- [8] T. E. O'Brien and G. M. Funk, "A Gentle Introduction to Optimal Design for Regression Models," *Am. Stat.*, vol. 57, no. 4, pp. 265–267, Nov. 2003.
- [9] V. V. Fedorov and P. Hackl, *Model-Oriented Design of Experiments*, 1st ed., vol. 125. New York, NY: Springer New York, 1997.
- [10] S. D. Silvey, *Optimal Design*. Dordrecht: Springer Netherlands, 1980.
- [11] A. Atkinson, A. Donev, and R. Tobias, *Optimum Experimental Designs with SAS*. OUP UK, 2007.
- [12] J. Kiefer and J. Wolfowitz, "The equivalence of two extremum problems," *Can. J. Math.*, vol. 12, pp. 363–366, Jan. 1960.
- [13] P. Whittle, "Some general points in the theory of optimal experimental design," *R. Stat. Soc.*, vol. 35, no. 1, pp. 123–130, 1973.
- [14] C. F. Restrepo and V. I. López Ríos, "Elección de la constante de ponderación en diseños cD-óptimos compuestos," *Ing. y Cienc.*, vol. 9, no. 18, pp. 107–130, 2013.
- [15] H. Chernoff, "Locally optimal designs for estimating parameters," *Ann. Math. Stat.*, vol. 24, no. 4, pp. 586–602, 1953.
- [16] I. Ford, B. Torsney, and C. F. . Wu, "The Use of a canonical Form in the Construction of Locally Optimal Designs For Nonlinear Problems," *R. Stat. Society*, vol. 54, no. 2, pp. 569–583, 1992.
- [17] H. Dette, V. B. Melas, and A. Pepelyshev, "Optimal designs for a Class of Nonlinear Regression Models," *Ann. Stat.*, vol. 32, no. 5, pp. 2114–2167, 2004.
- [18] H. Dette, V. B. Melas, and W. K. Wong, "Optimal Design for Goodness-of-Fit of the Michaelis–Menten Enzyme Kinetic Function," *J. Am. Stat. Assoc.*, vol. 100, no. 472, pp. 1370–1381, Dec. 2005.
- [19] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. John Wiley & Sons Inc, 2000.
- [20] R. D. C. Team, "R: A Language and Environment for Statistical Computing," 2011. [Online]. Available: <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing>.