



UNIVERSIDAD NACIONAL DE COLOMBIA

Comparación de dos metodologías en la construcción de diseños óptimos para modelos heterocedásticos

Jaime Andrés Gaviria Bedoya

Universidad Nacional de Colombia
Escuela de Estadística
Medellín, Colombia
2013

Comparación de dos metodologías en la construcción de diseños óptimos para modelos heterocedásticos

Jaime Andrés Gaviria Bedoya

Tesis presentada como requisito parcial para optar al título de:
Magister en Ciencias- Estadística

Director:
Víctor Ignacio López Ríos, profesor asociado
Escuela de Estadística

Universidad Nacional de Colombia
Escuela de Estadística
Medellín, Colombia
2013

(Dedicatoria o un lema)

A mi familia: Consuelo, Rosa, Estefany y
Oscar

Agradecimientos

Agradezco muy especialmente al profesor Víctor López por haberme asesorado en esta tesis, a los profesores Juan Carlos Correa y Maria Eugenia Castañeda por haber revisado este trabajo y a la escuela de estadística por todo el apoyo recibido durante estos últimos años.

Resumen

Los diseños de experimentos óptimos son una herramienta que le permite al investigador saber en que niveles del factor debe experimentar para obtener una estimación óptima de los parámetros del modelo en estudio bajo determinado criterio estadístico. Uno de los criterios de optimalidad más conocidos es el criterio de D optimalidad que consiste en encontrar el diseño que minimice la varianza generalizada del vector de parámetros. Ahora, el diseño depende de un modelo de regresión que tiene entre otros supuestos que la componente del error tiene varianza constante. Sin embargo, en la práctica se encuentran casos en los que no necesariamente se cumple dicho supuesto. Para resolver este problema existen dos metodologías: una es aplicar algún tipo de transformación sobre el modelo para volver homogénea la varianza y encontrar el diseño para el modelo transformado. La segunda consiste en incorporar la función que modela la varianza del error en el modelo y encontrar el diseño D óptimo. En este trabajo se encontrarán diseños D-óptimos mediante las dos metodologías, se compararán los diseños mediante el cálculo de las D eficiencias y se simularán observaciones en el modelo con el fin de comparar los errores relativos y los errores cuadráticos medios mediante ambas metodologías.

Palabras clave: Diseños óptimos, modelos heterocedásticos, Matriz de información, criterio D optimalidad.

Abstract

The optimal design of experiments is a tool that allows the researcher to know which factor levels should experiment to obtain a best estimate of the parameters of the model under certain statistical criteria. One of the best known criteria is D optimality which involves finding the design that minimizes the generalized variance of the parameter vector. Now, the design depends on a regression model which has among other assumptions that the error component has constant variance. However, in practice there are cases where this assumption is not satisfied necessarily. To solve this problem there are two methods: one is to apply some transformation on the model to get back the variance homogeneous and find the design for

the transformed model. The second consists of incorporating a function that models error variance in the model and find the D optimal design. In this thesis we will find D-optimal designs using the two methodologies, designs will be compared by calculating the D efficiencies and simulate the model with observations in order to compare the relative errors and mean square errors by both methodologies.

Keywords: Optimal designs, heteroscedatic models, Information matrix, D optimality criteria.

Contenido

Agradecimientos	vii
Resumen	ix
1. Introducción	2
2. Marco teórico	5
2.1. El modelo	5
2.1.1. Matriz de información	6
2.2. Criterios de optimalidad	7
2.2.1. Criterio D optimalidad	8
2.2.2. Criterio G-optimalidad	8
2.3. Teorema General de Equivalencia	9
3. Metodologías	11
3.1. Modelación de la varianza del error	11
3.1.1. Diseños óptimos modelando la varianza	12
3.2. Transformación del modelo	20
3.2.1. Transformaciones de Box-Cox	20
3.2.2. Diseños D-óptimos para el modelo transformado	21
4. Ejemplos	26
4.1. Ejemplo: Modelo heterocedástico no lineal	26
4.1.1. Ajuste del modelo no lineal	27
4.2. Modelación de la varianza del error	28
4.2.1. Diseño D óptimo local	30
4.2.2. Diseños D óptimos variando el parámetro τ	32
4.2.3. Simulación	34
4.2.4. Eficiencias	35
4.3. Segunda metodología: Transformación del modelo	37
4.3.1. Ajuste del modelo transformado	37
4.3.2. Diseño D óptimo para el modelo transformado	38
4.3.3. Simulación	40

5. Conclusiones y recomendaciones	42
5.1. Conclusiones	42
5.2. Recomendaciones	44
A. Algoritmos	47

1. Introducción

En cualquier investigación que involucre la realización de un experimento aleatorio, el investigador debe determinar previamente las variables de interés, los factores controlables y en especial en qué niveles de dichos factores debe experimentar. Posteriormente debe considerar un modelo estadístico que explique de manera adecuada el comportamiento entre los factores controlables (variables explicativas \mathbf{x}) y su relación con la variable de interés (variable respuesta \mathbf{Y}) (Atkinson, Donev y Tobias, 2007). Generalmente dicho modelo tiene la forma de los modelos de regresión: $\mathbf{Y} = \eta(\mathbf{x}, \theta) + \epsilon$, donde $\eta(\mathbf{x}, \theta)$ es una función (lineal o no) del vector de p parámetros θ y el vector de variables explicativas \mathbf{x} y ϵ es el error. Adicionalmente, el investigador debe determinar los niveles del factor en los que debe medir la variable respuesta. Si bien la escogencia de estos puntos de muestreo se puede basar en estudios previos o en la experiencia del investigador, dicha escogencia no deja de ser subjetiva. En este sentido, los diseños de experimentos óptimos surgen como una herramienta que le permite al investigador saber en qué niveles de la variable explicativa medir la variable respuesta para obtener una mejoría en la estimación (de acuerdo a un criterio de optimalidad) de los parámetros del modelo bajo estudio a un menor costo (López-Ríos y Ramos-Quiroga, 2007).

Para la construcción del diseño óptimo, se asume que en el modelo mencionado se cumplen los siguientes supuestos sobre los errores: normalidad, incorrelación y varianza constante. La violación del supuesto distribucional y la incorrelación son tratados por varios autores (para más información ver Steinberg y Hunter, 1984). En este trabajo se asumirá que se cumplen los supuestos de normalidad e incorrelación de los errores. El supuesto de varianza del error constante en la práctica puede no cumplirse. Por ejemplo, en bioquímica el modelo de Michaelis-Menten se utiliza para modelar la tasa de velocidad de reacción de formación de un producto en una reacción química dada una concentración de sustrato determinada. En este tipo de modelos el supuesto de homogeneidad de la varianza no necesariamente se cumple, ya que por lo general la varianza de la velocidad de la variable respuesta es una función de la parte determinística del modelo (López-Fidalgo y Wong, 2002). Cuando el modelo utilizado es heterocedástico, el problema de encontrar diseños óptimos requiere una metodología distinta a la utilizada en el caso homogéneo. Uno de los principales problemas de hallar diseños D óptimos para modelos heterocedásticos es el de encontrar una expresión para la matriz de información de Fisher. Esto se debe principalmente a la forma de la función de verosimilitud cuando la varianza no es constante, ya que en ese caso no solo se involucran los parámetros de la parte determinística sino otros adicionales. En este caso la matriz de información se deberá obtener a partir de las derivadas parciales de la función de

verosimilitud que pueden tener unas expresiones bastante complejas.

Para resolver el problema de la heterocedasticidad, existen dos metodologías que permiten obtener diseños D óptimos. Una primera que consiste en modelar la varianza del error mediante alguna función conocida. Para el caso de modelos lineales con estructura de varianza el trabajo más representativo puede ser el realizado por Atkinson y Cook (1995). Estos autores construyen diseños D-óptimos locales (diseños óptimos en los que se requiere conocer un valor local para el vector de parámetros) para modelos heterocedásticos lineales considerando determinadas estructuras en la varianza del error. Una de ellas es cuando el logaritmo de la varianza de la variable respuesta es una función lineal de una de las variables explicativas y donde la función que modela la varianza es exponencial. Otra estructura posible es cuando la varianza de la respuesta es una función lineal de su media. Como una aplicación de esta estructura Dette y Wong (1999), construyen diseños D-óptimos locales para el modelo de Michaelis-Mentel, en el cual la varianza de la respuesta Y varía de manera exponencial como función de su valor esperado $E(Y)$. Un trabajo más reciente es el de Dette y Müller (2012) en el que se encuentran diseños D óptimos locales y bayesianos en modelos de regresión donde el coeficiente de variación es constante.

Sin embargo, la modelación de la varianza sólo remueve la heterocedasticidad de los errores, pero no resuelve el problema de la no normalidad. En contraste, la segunda metodología que consiste en aplicar algún tipo de transformación de la familia de Box-Cox en el modelo bajo estudio generalmente resuelve tanto la heterocedasticidad como la no normalidad de los errores (Ritz, 2008). El problema de la transformación de los modelos de regresión no lineales ha sido abordado principalmente por Carroll y Ruppert (1988). Otros autores también han investigado este problema: Beal y Sheiner (1988) abordan los distintos métodos de estimación para los modelos de regresión heterocedásticos no lineales, mientras que Kimura (1990) aborda el problema del test de razón de verosimilitud para modelos de regresión no lineales bajo normalidad y heterocedasticidad. El problema de hallar diseños D óptimos cuando se aplica la transformación de Box-Cox a la variable respuesta ha sido abordado por Atkinson y Cook (1997). Estos autores suponen que después de aplicar la transformación adecuada que homogeniza la varianza de los errores, se tiene un modelo lineal y encuentran diseños D óptimos para dicho modelo. Un trabajo más reciente publicado por Atkinson (2003) encuentra diseños D óptimos locales para modelos no lineales donde se transforman ambos lados de un modelo no lineal tanto cuando la respuesta es univariada como multivariada.

Ahora, en la bibliografía consultada no se ha encontrado un trabajo donde se comparen las dos metodologías mencionadas, así que resulta de interés obtener diseños D-óptimos con ambos métodos. Por tal motivo en este trabajo se abordará el problema de encontrar diseños D óptimos locales para modelos de regresión no lineales con varianza no constante, usando ambos métodos. En el capítulo 2 se presentan los conceptos básicos de la teoría de diseños óptimos clásica (para modelos con varianza constante). En el capítulo 3 se hace un desarrollo

teórico de los diseños D óptimos para cada una de las dos metodologías. En la sección 3.1 se estudia la primera metodología de modelar la varianza del error y se demuestra un resultado importante que permite obtener la expresión de la matriz de información cuando se incorpora en el modelo la función que modela la varianza del error. Dicho resultado aunque aparece reportado en algunos artículos, no se demuestra, a pesar de la importancia que tiene. Esto puede deberse a la complejidad y extensión de la deducción. En la sección 3.2 se estudia la segunda metodología que consiste en aplicar al modelo alguna transformación de la familia de Box-Cox. En este caso se hará una breve revisión del artículo de Atkinson y Cook (1997) antes mencionado. En el capítulo 4 se ilustrarán las dos metodologías con un modelo de regresión no lineal con varianza no constante. En la sección 4.1 se encuentra el diseño D óptimo local cuando la función que modela la varianza del error es una función potencia de la media del modelo. En la sección 4.2 se encuentra el diseño D óptimo local cuando se aplica la transformación logarítmica al modelo. En cada caso se encuentran las D eficiencias para varios diseños y para evaluar la bondad de las estimaciones con ambos métodos, se hacen simulaciones de las observaciones del modelo en los puntos de soporte del diseño encontrado con cada método y se calculan tanto los errores relativos como cuadrático medio. Finalmente en el capítulo 5 se analizan los resultados obtenidos con ambos métodos, se realizan algunas comparaciones y recomendaciones.

2. Marco teórico

2.1. El modelo

Los diseños óptimos dependen entre otras cosas del modelo bajo estudio. En este trabajo se considera un modelo de regresión de la forma

$$Y = \eta(x, \beta) + \epsilon, \quad (2-1)$$

donde Y denota la variable respuesta, x es la variable explicativa, β es el vector de p parámetros, $\eta(x, \beta)$ es la parte determinística del modelo y es función de x y del vector de parámetros β y ϵ es la componente del error (Atkinson, Donev y Tobias, 2007). Además, si se dispone de N corridas experimentales, como en la teoría clásica de regresión, para el modelo (2-1) se tienen los siguientes supuestos sobre los errores:

- $\epsilon_i, i = 1, 2, \dots, N$ tiene distribución normal con media cero y varianza constante σ^2 .
- Los errores ϵ_i son independientes.

El modelo en (2-1) puede ser o no lineal respecto al vector de parámetros β . En el caso de que el modelo sea lineal, éste se puede escribir matricialmente como

$$Y = F\beta + E, \quad (2-2)$$

donde ahora Y es el vector de $N \times 1$ respuestas, $F_{N \times p}$ es la matriz de diseño y $E_{N \times 1} \sim N(0, \Sigma)$ es el vector de errores ϵ_i , con vector de medias 0 y matriz de varianzas y covarianzas Σ diagonal (debido al supuesto de independencia).

Para el modelo dado en (2-2) uno de los problemas de la teoría de diseños óptimos, consiste en encontrar un diseño ξ con n puntos de diseño x_1, x_2, \dots, x_n , que mejor estime el vector de parámetros β de acuerdo a algún criterio de optimalidad.

Un diseño ξ se define como una medida de probabilidad con soporte finito, denotado por:

$$\xi = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ w_1 & w_2 & \cdots & w_n \end{bmatrix}, \quad (2-3)$$

donde x_1, x_2, \dots, x_n son los puntos de soporte del diseño, con pesos asociados $w_i \geq 0$ y tales que $\sum_{i=1}^n w_i = 1$ (O'Brien y Funk, 2003). Desde el punto de vista matemático, los pesos w_i

pueden ser cualquier número real entre 0 y 1, en cuyo caso se habla de diseños continuos. Sin embargo, en la práctica se trabaja con diseños exactos; esto es, con diseños donde los pesos indiquen la frecuencia de observación en cada punto de soporte (López - Ríos y Ramos - Quiroga, 2007). En este caso, el diseño exacto se escribe como:

$$\xi = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ \frac{r_1}{N} & \frac{r_2}{N} & \cdots & \frac{r_n}{N} \end{bmatrix}, \quad (2-4)$$

donde r_i es un entero positivo que indica el número de réplicas en el nivel x_i y $N = \sum_{i=1}^n r_i$ (Fedorov y Hackl, 1997).

2.1.1. Matriz de información

Para el modelo lineal (2-2), la matriz de información asociada al diseño continuo ξ está dada por :

$$M(\xi) = \int_{\mathcal{X}} f(x)f^T(x)\xi(dx) = \sum_{i=1}^n w_i f(x_i)f^T(x_i), \quad (2-5)$$

donde $f^T(x_i)$ es la i -ésima fila de la matriz de diseño F (López-Ríos y Ramos-Quiroga, 2007). Esta matriz juega un papel esencial en la construcción del diseño, pues dependiendo del criterio de optimalidad utilizado, el diseño ξ se halla maximizando o minimizando algún funcional real $\Psi(M(\xi))$ de la matriz de información sobre todos los diseños ξ posibles; es decir

$$\xi^* = \underset{\xi}{arg \text{mín}} \Psi[M(\xi)], \quad (2-6)$$

donde $\Psi[M(\xi)]$ es una función conocida como criterio de optimalidad y es utilizada dependiendo de los intereses del investigador (Fedorov y Hackl, 1997).

Además, si la matriz de información es invertible, entonces se define la varianza estandarizada de la respuesta predicha para el diseño continuo ξ como

$$d(x, \xi) = f^T(x)M^{-1}(\xi)f(x) = \text{var}(\hat{Y}(x))/\sigma^2 \quad (2-7)$$

y como se observa esta función $d(x, \xi)$ es proporcional a la varianza de la respuesta predicha, $\text{var}(\hat{Y}(x))$. Esta función como se verá más adelante juega un papel importante en la verificación de la optimalidad del diseño obtenido.

En el caso de que el modelo bajo estudio sea no lineal:

$$Y = \eta(x, \beta) + \epsilon, \quad (2-8)$$

se puede linealizar este modelo usando la aproximación de Taylor de primer orden para $\eta(x, \beta)$ alrededor del valor β_0 :

$$\eta(x, \beta) \approx \eta(x, \beta_0) + \frac{\partial \eta(x, \beta)}{\partial \beta^T} \Big|_{\beta=\beta_0} (\beta - \beta_0). \quad (2-9)$$

Luego el modelo linealizado es

$$y = \eta(x, \beta_0) + \frac{\partial \eta(x, \beta)}{\partial \beta^T} \Big|_{\beta=\beta_0} (\beta - \beta_0) + \epsilon \quad (2-10)$$

o equivalentemente

$$y - \eta(x, \beta_0) + \frac{\partial \eta(x, \beta)}{\partial \beta^T} \Big|_{\beta=\beta_0} \beta_0 = \frac{\partial \eta(x, \beta)}{\partial \beta^T} \Big|_{\beta=\beta_0} \beta + \epsilon \quad (2-11)$$

y la matriz de información para el modelo no lineal es

$$M(\xi) = \int_{\mathcal{X}} \frac{\partial \eta(x, \beta)}{\partial \beta} \frac{\partial \eta(x, \beta)}{\partial \beta^T} \xi(dx) = \sum_{i=1}^n \frac{\partial \eta(x_i, \beta)}{\partial \beta} \frac{\partial \eta(x_i, \beta)}{\partial \beta^T} w_i, \quad (2-12)$$

que corresponde a la matriz de información para el modelo linealizado con $f(x, \beta) = \frac{\partial \eta(x, \beta)}{\partial \beta}$.

2.2. Criterios de optimalidad

Como se vió anteriormente, la construcción del diseño óptimo se basa en la optimización de algún funcional de la matriz de información. La estructura de dicho funcional es conocida como criterio de optimalidad (López - Ríos y Ramos - Quiroga, 2007). Por lo tanto existen tantos criterios de optimalidad como funciones con algún criterio estadístico puedan definirse. Uno de los criterios de optimalidad más conocido es el criterio D-optimalidad. Éste consiste en minimizar la varianza generalizada del vector de parámetros estimados, $|Var(\hat{\beta})|$ o equivalentemente minimizar $|M^{-1}(\xi)|$ (Atkinson, Donev y Tobias, 2007). Además, dado que los errores del modelo son normales, una interpretación geométrica del criterio es la de minimizar el volumen del elipsoide asociado al vector de parámetros β (O'Brien y Funk, 2003).

Una condición importante que debe cumplir un criterio de optimalidad es la propiedad de convexidad; es decir, si $M(\xi) = (1 - \alpha)M_1(\xi) + \alpha M_2(\xi)$, entonces para cualquier $0 \leq \alpha \leq 1$ se cumple que

$$\Psi[M(\xi)] \leq (1 - \alpha)\Psi[M_1(\xi)] + \alpha\Psi[M_2(\xi)]. \quad (2-13)$$

La mayoría de los criterios de optimalidad son convexos, sin embargo, el criterio D-optimalidad no lo es, pero puede transformarse en convexo tomando como criterio $\Psi[M(\xi)] = -\log|M(\xi)|$ (Fedorov y Kackl, 1997). Lo anterior implica que en los algoritmos usados para encontrar

diseños D-óptimos, se minimizará la función $\Psi[M(\xi)] = -\log|M(\xi)|$.

2.2.1. Criterio D optimalidad

Como se mencionó antes, uno de los criterios de optimalidad más conocido es el de D optimalidad. A lo largo de este trabajo se usarán algunas de las siguientes propiedades para este criterio (Atkinson, Donev y Tobias, 2007):

1. El diseño ξ^* maximiza el determinante de la matriz de información o equivalentemente minimiza $|M^{-1}(\xi)|$.
2. La D eficiencia de un diseño arbitrario ξ respecto al diseño óptimo ξ^* se obtiene como

$$D_{\text{eff}} = \left(\frac{|M(\xi)|}{|M(\xi^*)|} \right)^{1/p}, \quad (2-14)$$

donde p es el número de parámetros del modelo. Ésta es una medida de qué tan eficiente es el diseño ξ respecto al diseño D óptimo ξ^* y entre más cercano a uno sea, más eficiente será el diseño aunque no sea óptimo.

3. Para el caso de modelos lineales con varianza constante, el diseño D óptimo ξ tiene tantos puntos de soporte como parámetros tiene el diseño.
4. El diseño D óptimo es invariante frente a transformaciones lineales no degenerativas del modelo. Así, por ejemplo un diseño D óptimo para el modelo lineal $\eta = \beta^T f(x)$ también es D óptimo para el modelo transformado $\eta = \gamma^T g(x)$, donde $g(x) = Af(x)$, $|A| \neq 0$, y β, γ son los vectores de parámetros de cada modelo.

2.2.2. Criterio G-optimalidad

Siguiendo la metodología minmax para diseños óptimos, el criterio G-optimalidad consiste en minimizar el máximo de la varianza estandarizada de la respuesta predicha $d(x, \xi)$ (definida en la ecuación 2-7); es decir un diseño ξ_G^* es G-óptimo si es el diseño que minimiza sobre todos los diseños posibles el máximo sobre la región de diseño χ de la varianza estandarizada de la respuesta predicha (Wong, 1992):

$$\xi_G^* = \arg \min_{\xi} \max_{\chi} d(x, \xi). \quad (2-15)$$

Este criterio tiene una relación importante con el criterio D-optimalidad, ya que para diseños continuos se sabe que un diseño ξ_G^* G-óptimo es también un diseño D óptimo y en tal caso $\max_{\chi} d(x, \xi) = p$, donde p es el número de parámetros del modelo (Atkinson, Donev y Tobias, 2007). Esta equivalencia entre los criterios D y G optimalidad se cumple para

modelos con varianza constante y se establece más formalmente a continuación (para modelos heterocedásticos ver Wong, 1995).

2.3. Teorema General de Equivalencia

En la práctica, la manera de encontrar y verificar que un diseño es óptimo es mediante los teoremas de equivalencia. Es de aclarar que dichos teoremas solo son válidos para diseños continuos (Atkinson, Donev y Tobias, 2007). Antes de enunciar el teorema general de equivalencia, se necesita definir la derivada direccional de Ψ en la dirección $\bar{\xi}_x$.

Sea $\bar{\xi}_x$ la medida de probabilidad que asigna la unidad al punto x y ξ' la medida dada por

$$\xi' = (1 - \alpha)\xi + \alpha\bar{\xi}_x. \quad (2-16)$$

Entonces la derivada direccional de Ψ en la dirección $\bar{\xi}_x$ se define como:

$$\phi(x, \xi) = \lim_{\alpha \rightarrow 0^+} \frac{1}{\alpha} [\Psi[M(\xi')] - \Psi[M(\xi)]], \quad (2-17)$$

donde $M(\xi') = (1 - \alpha)M(\xi) + \alpha M(\bar{\xi}_x)$.

Se enuncia el siguiente teorema para el caso de que el modelo (2-1) sea lineal (Atkinson, Donev y Tobias, 2007).

Teorema 2.3.1 (Teorema General de Equivalencia). *Si Ψ es un criterio de optimalidad convexo y diferenciable con región de diseño χ un conjunto compacto y $f(x_i)$ funciones continuas, las siguientes afirmaciones caracterizan equivalentemente a un diseño ξ^* :*

1. *El diseño ξ^* minimiza $\Psi[M(\xi)]$ sobre todos los diseños.*
2. *El diseño ξ^* maximiza el mínimo sobre χ de $\phi(x, \xi)$.*
3. *El mínimo sobre la región χ de $\phi(x, \xi^*) = 0$, este mínimo se alcanza en los puntos de soporte x_i del diseño ξ^* .*
4. *Para cualquier otro diseño ξ distinto a ξ^* , se cumple $\phi(x, \xi) < 0$.*

Del teorema anterior se derivan una serie de casos especiales como el de Kiefer-Wolfowitz, que establece un criterio muy útil para determinar cuando un diseño es D-óptimo (Fedorov y Kackl, 1997).

Teorema 2.3.2. [Kiefer-Wolfowitz] *Si Ψ es un criterio de optimalidad convexo y diferenciable con región de diseño χ un conjunto compacto y $f(x_i)$ funciones continuas, entonces las siguientes afirmaciones caracterizan equivalentemente a un diseño ξ_D :*

1. ξ_D minimiza $|M(\xi)^{-1}|$,
2. ξ_D minimiza $\max_x d(x, \xi)$.
3. $\max_x d(x, \xi_D) = p$ donde p es el número de parámetros del modelo y $d(x, \xi) = f^T(x)M^{-1}(\xi)f(x)$ es la varianza estandarizada de la respuesta predicha.

Nota: en el caso de que el modelo (2-1) sea no lineal, el teorema anterior también es válido, ya que al linealizar el modelo, la función matriz de información está dada por

$$M(\xi, \beta) = \int_{\mathcal{X}} \frac{\partial \eta(x, \beta)}{\partial \beta} \frac{\partial \eta(x, \beta)}{\partial \beta^T} \xi(dx) = \sum_{i=1}^n \frac{\partial \eta(x_i, \beta)}{\partial \beta} \frac{\partial \eta(x_i, \beta)}{\partial \beta^T} w_i, \quad (2-18)$$

y en este caso la varianza de la respuesta predicha está dada igual que para el modelo lineal, excepto que en este caso $f(x, \beta) = \frac{\partial \eta(x, \beta)}{\partial \beta}$. Como se observa, en el caso de que el modelo sea no lineal, la matriz de información dependerá del vector de parámetros β , por lo que en el proceso de optimización se pueden usar dos enfoques. El primero de ellos es usar diseños óptimos locales si el investigador tiene una buena idea del valor verdadero de β . El segundo enfoque es el de hallar diseños óptimos promediados por una distribución distribución a priori para β (enfoque bayesiano) (López - Ríos y Ramos - Quiroga, 2007). En este trabajo se usará el enfoque de hallar diseños D-óptimos locales para modelos heterocedásticos utilizando las dos metodologías mencionadas al inicio.

3. Metodologías

En la teoría clásica de regresión se asume que en el modelo general

$$Y = \eta(x, \beta) + \epsilon, \quad (3-1)$$

los errores se distribuyen normal con media cero y varianza constante. Sin embargo, en la práctica los residuales usualmente exhiben algún tipo de heterocedasticidad y/o asimetría. Con la violación de alguno de estos supuestos, no se puede garantizar que el modelo (3-1) sea adecuado y por tanto las conclusiones obtenidas a partir del mismo no serán confiables (Ritz, 2008).

Para resolver el problema de la heterocedasticidad y/o no normalidad de los errores, existen dos metodologías. La primera consiste en modelar la varianza del error con alguna función conocida e incorporarla en el modelo (3-1). Sin embargo, la modelación de la varianza sólo remueve la heterocedasticidad de los errores, pero no resuelve el problema de la no normalidad (Ritz, 2008). En contraste, la segunda metodología que consiste en aplicar algún tipo de transformación en el modelo bajo estudio generalmente resuelve tanto la heterocedasticidad como la no normalidad de los errores.

3.1. Modelación de la varianza del error

Una forma donde los errores del modelo (3-1) tienen varianza no constante es cuando la varianza de la variable respuesta, $var[Y]$, sea proporcional a su valor esperado, $E[Y]$. Por ejemplo, en la figura (3-1) se observa que cuando el valor esperado de la respuesta es pequeño, también lo es la varianza, pero cuando la media aumenta, la variabilidad también. Más adelante se va a analizar más a fondo este conjunto de datos.

Existen distintas formas de modelar la varianza cuando ésta es una función creciente. Una de ellas es si la función que modela la varianza es lineal en la respuesta (Huet et al., 2004):

$$var(\epsilon_i) = \sigma^2(1 + \tau\eta(x_i, \beta)). \quad (3-2)$$

En este caso como $1 + \tau\eta(x_i, \beta) > 0$ entonces el parámetro τ tiene la restricción $\tau > \frac{-1}{\eta(x_i, \beta)}$.

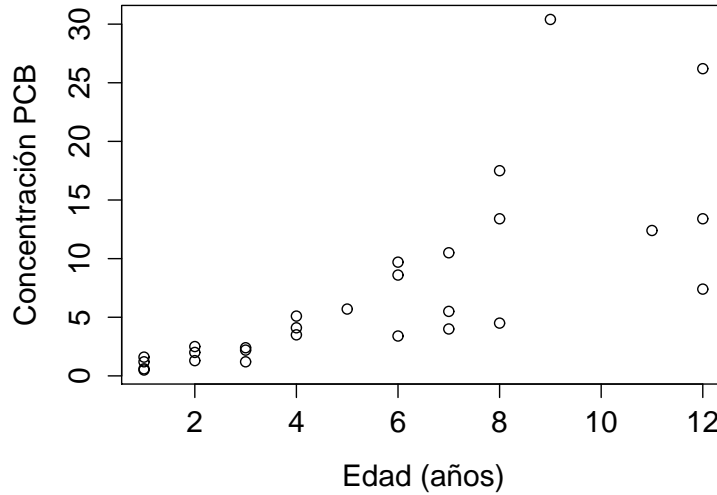


Figura 3-1.: Gráfico de los datos

Otra forma de modelar la varianza del error es mediante una función potencia de la media:

$$\text{var}(\epsilon_i) = \sigma^2(\eta(x_i, \beta))^{2\tau}. \quad (3-3)$$

El modelo heterocedástico (3-1) con estructura de varianza dada en (3-3) es conocido en la literatura como el modelo con varianza potencia de la media (*the-power-of-the-mean variance model*):

$$Y_i = \eta(x_i, \beta) + \epsilon_i, \text{ con } \text{var}(\epsilon_i) = \sigma^2(\eta(x_i, \beta))^{2\tau}, \quad (3-4)$$

y la homogeneidad de varianza corresponde a $\tau = 0$, en cuyo caso el modelo homocedástico (3-1) es un submodelo del modelo (3-4) y por lo tanto puede implementarse una prueba estadística para determinar si el modelo heterocedástico (3-4) proporciona una mejoría en el ajuste comparado con el modelo homocedástico (Ritz, 2008).

3.1.1. Diseños óptimos modelando la varianza

Una primera aproximación al problema de hallar diseños D óptimos modelando la varianza del error es la dada por Wong (1995), quien considera el modelo lineal

$$Y = f(x)^T \beta + \epsilon(x), \quad (3-5)$$

donde se asume que los errores son normales e independientes y con varianza inversamente proporcional a una función conocida, $\lambda(x)$. La matriz de información para este modelo está dada por:

$$M(\xi) = \int_x \lambda(x) f(x) f(x)^T \xi(dx). \quad (3-6)$$

En este trabajo se propone una generalización al trabajo realizado por Wong (1995), en el sentido de que la función que modela la varianza del error no solo depende de x , sino también puede depender de algunos parámetros. Por ejemplo, en farmacocinética un modelo muy utilizado es el modelo de Michaelis-Menten:

$$Y = \frac{mx}{k+x} + \epsilon, \quad (3-7)$$

que modela la tasa de velocidad de formación de un producto en una reacción química dada una concentración de sustrato x (López-Fidalgo y Wong, 2002). Este modelo se caracteriza por ser un modelo heterocedástico donde la varianza del error es una función potencia de la media, que se puede expresar como una función de la forma (Dette y Wong, 1999):

$$\text{var}(Y) = v(\eta(x, \theta)), \quad (3-8)$$

donde $\eta(x, \theta) = \frac{mx}{k+x}$ y $\theta = (m, k)^T$. En este caso la función que modela la varianza del error depende del vector de parámetros θ .

El modelo general es como lo proponen Downing, Fedorov y Leonov (2001):

$$Y = \eta(x, \theta) + \epsilon, \text{ con } \text{var}(\epsilon) = S(x, \theta), \quad (3-9)$$

donde la función $S(x, \theta)$ es una función positiva. Por ejemplo, en los experimentos de dosis respuesta, se usan modelos heterocedásticos donde se modela la varianza del error, como por ejemplo en el modelo logístico

$$Y = \eta(x, \theta) + \epsilon, \quad (3-10)$$

donde $\eta(x, \theta) = \theta_1 + \frac{\theta_2 - \theta_1}{1 + (x/\theta_3)^{\theta_4}}$ y la función que modela la varianza es (Downing, Fedorov y Leonov, 2001):

$$s(x, \theta) = \theta_5 \eta(x, \theta)^6. \quad (3-11)$$

Nótese que en este caso la función $s(x, \theta)$ no solo depende de los parámetros de la parte determinística, sino también de otros parámetros adicionales.

En particular para el modelo (3-4) la función $S(x, \theta)$ que modela la varianza del error es una función potencia de la media:

$$S(x, \theta) = \sigma^2(\eta(x_i, \beta))^{2\tau}, \quad (3-12)$$

con $\theta = (\beta, \tau, \sigma^2)$.

El problema que se plantea es un caso más general que el considerado por Wong, ya que primero el modelo puede ser un modelo no lineal en el vector de parámetros β y segundo se asume que la varianza del error es una función que no sólo depende de la variable regresora sino que también depende de un vector de parámetros $\theta = (\beta, \tau, \sigma^2)$.

Ahora, para poder encontrar el diseño D óptimo es necesario deducir una expresión para la matriz de información asociada al modelo (3-11). Para ello se demostrará el siguiente resultado (Downing, Fedorov y Leonov, 2001):

Teorema 3.1.1 (Matriz de información). *Sea Y con distribución normal con $E[Y|x] = \eta(x, \theta)$ y $Var[Y|x] = S(x, \theta)$, donde $S(x, \theta) > 0$. Si además las N observaciones $\{y_i, x_i\}_{i=1}^N$ son independientes, la matriz de información de Fisher para el diseño aproximado ξ es*

$$M(\xi, \theta) = \int_x I(x, \theta) d\xi(x), \quad (3-13)$$

donde

$$I(x, \theta) = \frac{1}{S(x, \theta)} \frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T} + \frac{1}{2} \frac{1}{S(x, \theta)^2} \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T}. \quad (3-14)$$

Demostración. Como $Y|x \sim N(\eta(x, \theta), S(x, \theta))$, el estadístico Score es $S(\theta) = \frac{\partial \ln L(\theta|x)}{\partial \theta}$ donde la verosimilitud está dada por

$$L(\theta|x, Y) = \frac{1}{\sqrt{2\pi S(x, \theta)}} \exp\left(-\frac{(Y - \eta(x, \theta))^2}{2S(x, \theta)}\right) \quad (3-15)$$

y

$$\ln L(\theta|x, Y) = -\ln \sqrt{2\pi} - \frac{1}{2} \left(\ln S(x, \theta) + \frac{(Y - \eta(x, \theta))^2}{S(x, \theta)} \right). \quad (3-16)$$

Por lo tanto se tiene que

$$\begin{aligned} \frac{\partial}{\partial \theta} \left(\ln S(x, \theta) + \frac{(Y - \eta(x, \theta))^2}{S(x, \theta)} \right) &= \frac{1}{S(x, \theta)} \frac{\partial S(x, \theta)}{\partial \theta} \\ &\quad - \frac{\left(2(Y - \eta(x, \theta))S(x, \theta) \frac{\partial \eta(x, \theta)}{\partial \theta} + (Y - \eta(x, \theta))^2 \frac{\partial S(x, \theta)}{\partial \theta} \right)}{S(x, \theta)^2} \\ &= \frac{1}{S(x, \theta)} \frac{\partial S(x, \theta)}{\partial \theta} - \frac{2(Y - \eta(x, \theta)) \frac{\partial \eta(x, \theta)}{\partial \theta}}{S(x, \theta)} - \frac{(Y - \eta(x, \theta))^2 \frac{\partial S(x, \theta)}{\partial \theta}}{S(x, \theta)^2} \end{aligned} \quad (3-17)$$

Pero para el estadístico Score se sabe que (Casella y Berger, 2002):

$$E[S(\theta)] = E \left[\frac{\partial \ln L(\theta|x, Y)}{\partial \theta} \right] = 0 \quad (3-18)$$

y

$$Var[S(\theta)] = E \left[\left(\frac{\partial \ln L(\theta|x, Y)}{\partial \theta} \right)^2 \right] = I(x, \theta) \quad (3-19)$$

donde al ser θ un vector entonces

$$E \left[\left(\frac{\partial \ln L(\theta|x, Y)}{\partial \theta} \right)^2 \right] = E \left[\frac{\partial \ln L(\theta|x, Y)}{\partial \theta} \frac{\partial \ln L(\theta|x, Y)}{\partial \theta^T} \right] \quad (3-20)$$

y

$$\begin{aligned} 4E \left[\frac{\partial \ln L(\theta|x, Y)}{\partial \theta} \frac{\partial \ln L(\theta|x, Y)}{\partial \theta^T} \right] &= \frac{1}{S(x, \theta)^2} \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T} - \\ &\frac{4E[Y - \eta(x, \theta)]}{S(x, \theta)^2} \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T} - \\ &\frac{2E[Y - \eta(x, \theta)]^2}{S(x, \theta)^3} \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T} + \\ &\frac{4E[Y - \eta(x, \theta)]^2}{S(x, \theta)^2} \frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T} + \\ &\frac{4E[Y - \eta(x, \theta)]^3}{S(x, \theta)^3} \frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T} + \\ &\frac{E[Y - \eta(x, \theta)]^4}{S(x, \theta)^4} \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T}. \end{aligned}$$

Pero por propiedades de la distribución normal, se sabe que los momentos centrales impares son cero y que los pares son de la forma:

$$E[Y - \mu_Y]^n = \frac{2k! \sigma^{2k}}{2^k k!}, \quad (3-21)$$

donde $n = 2k$. Por lo tanto, como $Y \sim N(\eta(x, \theta), S(x, \theta))$, entonces se tiene que

$$E[Y - \eta(x, \theta)] = E[Y - \eta(x, \theta)]^3 = 0 \quad (3-22)$$

$$E[Y - \eta(x, \theta)]^2 = Var[Y|x, \theta] = S(x, \theta) \quad y \quad (3-23)$$

$$E[Y - \eta(x, \theta)]^4 = 3(Var[Y|x, \theta])^2 = 3S^2(x, \theta), \quad (3-24)$$

de modo que

$$\begin{aligned}
4E \left[\frac{\partial \ln L(\theta|x, Y)}{\partial \theta} \frac{\partial \ln L(\theta|x, Y)}{\partial \theta^T} \right] &= \frac{4}{S(x, \theta)} \frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T} + \\
&\quad \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T} \left(\frac{1}{S(x, \theta)^2} - \frac{2}{S(x, \theta)^2} + \frac{3}{S(x, \theta)^2} \right) \\
&= \frac{4}{S(x, \theta)} \frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T} + \frac{2}{S(x, \theta)^2} \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T}.
\end{aligned}$$

De las ecuaciones anteriores se sigue que la matriz de información en una observación es

$$I(x, \theta) = E \left[\left(\frac{\partial \ln L(\theta|x, Y)}{\partial \theta} \right)^2 \right] \quad (3-25)$$

$$= \frac{1}{S(x, \theta)} \frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T} + \frac{1}{2S(x, \theta)^2} \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T} \quad (3-26)$$

que era lo que se quería demostrar. \square

Como un caso particular del modelo (3-4), se tiene el siguiente corolario:

Corolario 3.1.2. *Para el modelo con varianza potencia de la media (3-4):*

$$Y_i = \eta(x_i, \beta) + \epsilon_i, \quad (3-27)$$

donde los errores son independientes y se distribuyen normal con media cero y varianza $\text{var}(\epsilon_i) = \sigma^2(\eta(x_i, \beta))^{2\tau}$ con β, τ y σ^2 parámetros, la matriz de información está dada por

$$M(\xi, \theta) = UWU^T + VWV^T, \quad (3-28)$$

donde

$$U_{(p+2) \times r} = (u_1, u_2, \dots, u_r) \quad V_{(p+2) \times r} = (v_1, v_2, \dots, v_r), \quad (3-29)$$

$$W = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & \\ \vdots & & \ddots & \\ 0 & & & w_r \end{pmatrix} \quad (3-30)$$

y para $i = 1, 2, \dots, r$:

$$u_i = \left(\frac{1}{\sigma \eta(x_i, \beta)^\tau} \frac{\partial \eta(x_i, \beta)}{\partial \beta^T}, 0, 0 \right)_{(p+2) \times 1}^T \quad (3-31)$$

$$v_i = \left(\frac{\sqrt{2}\tau}{\eta(x_i, \beta)} \frac{\partial \eta(x_i, \beta)}{\partial \beta^T}, \sqrt{2} \ln \eta(x_i, \beta), \frac{1}{\sqrt{2}\sigma^2} \right)_{(p+2) \times 1}^T. \quad (3-32)$$

Demostración. Se puede aplicar el Teorema 3.1.1 al modelo con varianza potencia de la media (3-4) con $\theta = (\beta^T, \tau, \sigma^2)^T$ y $S(x, \theta) = \sigma^2 \eta(x, \beta)^{2\tau}$.

Como $\eta(x, \beta)$ no depende de τ y σ^2 entonces:

$$\begin{aligned} \frac{\partial \eta(x, \theta)}{\partial \theta} &= \left(\frac{\partial \eta(x, \beta)}{\partial \beta^T}, \frac{\partial \eta(x, \beta)}{\partial \tau}, \frac{\partial \eta(x, \beta)}{\partial \sigma^2} \right)^T \\ &= \left(\frac{\partial \eta(x, \beta)}{\partial \beta^T}, 0, 0 \right)^T \end{aligned} \quad (3-33)$$

luego

$$\frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T} = \left(\begin{array}{c|cc} \frac{\partial \eta(x, \beta)}{\partial \beta} & \frac{\partial \eta(x, \beta)}{\partial \beta^T} & \begin{array}{c} 0_{p \times 1} \\ 0_{p \times 1} \end{array} \\ \hline 0^T & 0 & 0 \\ 0^T & 0 & 0 \end{array} \right)_{(p+2) \times (p+2)}. \quad (3-34)$$

Por otro lado

$$\frac{\partial S(x, \theta)}{\partial \theta} = \left(\frac{\partial S(x, \theta)}{\partial \beta^T}, \frac{\partial S(x, \theta)}{\partial \tau}, \frac{\partial S(x, \theta)}{\partial \sigma^2} \right)^T, \quad (3-35)$$

donde

$$\begin{aligned} \frac{\partial S(x, \theta)}{\partial \beta} &= \sigma^2 \frac{\partial \eta(x, \beta)^{2\tau}}{\partial \beta} \\ &= \sigma^2 \frac{\partial \eta(x, \beta)^{2\tau}}{\partial \eta} \frac{\partial \eta(x, \beta)}{\partial \beta} \quad \text{por la regla de la cadena} \\ &= \sigma^2 (2\tau \eta(x, \beta)^{2\tau-1}) \frac{\partial \eta(x, \beta)}{\partial \beta} \end{aligned} \quad (3-36)$$

$$\begin{aligned} \frac{\partial S(x, \theta)}{\partial \tau} &= \sigma^2 \frac{\partial \eta(x, \beta)^{2\tau}}{\partial \tau} \\ &= \sigma^2 \frac{\partial (e^{2\tau \ln \eta(x, \beta)})}{\partial \tau} \\ &= 2\sigma^2 \eta(x, \beta)^{2\tau} \ln \eta(x, \beta) \end{aligned} \quad (3-37)$$

y

$$\frac{\partial S(x, \theta)}{\partial \sigma^2} = \eta(x, \beta)^{2\tau}. \quad (3-38)$$

Entonces

$$\begin{aligned} \frac{\partial S(x, \theta)}{\partial \theta} &= \left(2\tau\sigma^2\eta(x, \beta)^{2\tau-1} \frac{\partial \eta(x, \beta)}{\partial \beta^T}, 2\sigma^2\eta(x, \beta)^{2\tau} \ln \eta(x, \beta), \eta(x, \beta)^{2\tau} \right)^T \\ &= \eta(x, \beta)^{2\tau} \left(\frac{2\tau\sigma^2}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta^T}, 2\sigma^2 \ln \eta(x, \beta), 1 \right)^T. \end{aligned} \quad (3-39)$$

Por lo tanto se obtiene la matriz simétrica

$$\frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T} = (\eta(x, \beta)^{2\tau})^2 \begin{pmatrix} \frac{4\tau^2\sigma^4}{\eta(x, \beta)^2} \frac{\partial \eta(x, \beta)}{\partial \beta} \frac{\partial \eta(x, \beta)}{\partial \beta^T} & \frac{4\tau\sigma^4 \ln \eta(x, \beta)}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} & \frac{2\tau\sigma^2}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} \\ \frac{4\tau\sigma^4 \ln \eta(x, \beta)}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} & 4\sigma^4 (\ln \eta(x, \beta))^2 & 2\sigma^2 \ln \eta(x, \beta) \\ \frac{2\tau\sigma^2}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} & 2\sigma^2 \ln \eta(x, \beta) & 1 \end{pmatrix} \quad (3-40)$$

y por el Teorema 3.1.1 la matriz de información de Fisher para el diseño aproximado ξ es

$$M(\xi, \theta) = \int_{\mathcal{X}} I(x, \theta) d\xi(x), \quad (3-41)$$

donde

$$I(x, \theta) = \frac{1}{S(x, \theta)} \frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T} + \frac{1}{2} \frac{1}{S(x, \theta)^2} \frac{\partial S(x, \theta)}{\partial \theta} \frac{\partial S(x, \theta)}{\partial \theta^T}, \quad (3-42)$$

que por las ecuaciones 3-35 y 3-40 se expresa como

$$\begin{aligned} I(x, \theta) &= \frac{1}{\sigma^2\eta(x, \beta)^{2\tau}} \begin{pmatrix} \frac{\partial \eta(x, \beta)}{\partial \beta} \frac{\partial \eta(x, \beta)}{\partial \beta^T} & 0_{p \times 1} & 0_{p \times 1} \\ 0^T & 0 & 0 \\ 0^T & 0 & 0 \end{pmatrix} \\ &+ \frac{1}{2(\sigma^2(\eta(x, \beta)^{2\tau}))^2} (\eta(x, \beta)^{2\tau})^2 \begin{pmatrix} \frac{4\tau^2\sigma^4}{\eta(x, \beta)^2} \frac{\partial \eta(x, \beta)}{\partial \beta} \frac{\partial \eta(x, \beta)}{\partial \beta^T} & \frac{4\tau\sigma^4 \ln \eta(x, \beta)}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} & \frac{2\tau\sigma^2}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} \\ \frac{4\tau\sigma^4 \ln \eta(x, \beta)}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} & 4\sigma^4 (\ln \eta(x, \beta))^2 & 2\sigma^2 \ln \eta(x, \beta) \\ \frac{2\tau\sigma^2}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} & 2\sigma^2 \ln \eta(x, \beta) & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sigma(\eta(x, \beta)^\tau)^2} \frac{\partial \eta(x, \beta)}{\partial \beta} \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma(\eta(x, \beta)^\tau)^2} \frac{\partial \eta(x, \beta^T)}{\partial \beta} & 0 & 0 \end{pmatrix} \\ &+ \begin{pmatrix} \frac{\sqrt{2}\tau}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta} \\ \sqrt{2} \ln \eta(x, \beta) \\ \frac{1}{\sqrt{2}\sigma^2} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}\tau}{\eta(x, \beta)} \frac{\partial \eta(x, \beta)}{\partial \beta^T} & \sqrt{2} \ln \eta(x, \beta) & \frac{1}{\sqrt{2}\sigma^2} \end{pmatrix}. \end{aligned} \quad (3-43)$$

O sea que

$$I(x_i, \theta) = u_i u_i^T + v_i v_i^T, \quad (3-44)$$

donde

$$u_i = \left(\frac{1}{\sigma \eta(x_i, \beta)^\tau} \frac{\partial \eta(x_i, \beta)}{\partial \beta}, 0, 0 \right)_{(p+2) \times 1}^T \quad (3-45)$$

$$v_i = \left(\frac{\sqrt{2}\tau}{\eta(x_i, \beta)} \frac{\partial \eta(x_i, \beta)}{\partial \beta}, \sqrt{2} \ln \eta(x_i, \beta), \frac{1}{\sqrt{2}\sigma^2} \right)_{(p+2) \times 1}^T, \quad (3-46)$$

y entonces la matriz de información de Fisher es

$$M(\xi, \theta) = \sum_{i=1}^r I(x_i, \theta) w_i = \sum_{i=1}^r u_i u_i^T w_i + \sum_{i=1}^r v_i v_i^T w_i, \quad (3-47)$$

o equivalentemente

$$M(\xi, \theta) = U W U^T + V W V^T, \quad (3-48)$$

donde

$$U_{(p+2) \times r} = (u_1, u_2, \dots, u_r) \quad V_{(p+2) \times r} = (v_1, v_2, \dots, v_r) \quad (3-49)$$

$$W = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & \\ \vdots & & \ddots & \\ 0 & & & w_r \end{pmatrix} \quad (3-50)$$

que era lo que se quería mostrar. \square

Para verificar que el diseño ξ^* es el D óptimo, se usa el siguiente corolario del Teorema General de Equivalencia (Dette y Wong, 1999), (López-Ríos, 2008):

Teorema 3.1.3. [Teorema de Equivalencia D optimalidad]

Sea $M(\xi, \theta)_{m \times m}$ la matriz de información del diseño ξ una matriz definida positiva, $\Psi(\xi, \theta) = \log |M(\xi, \theta)|$ el criterio D-optimalidad y el rango de regresión χ un conjunto compacto. Entonces el diseño ξ^* es D óptimo si la derivada direccional de ϕ en ξ^* en la dirección de ξ_x cumple

$$\phi(M(\xi^*, \theta), M(\xi_x, \theta)) \leq 0 \quad \forall x \in \chi, \quad (3-51)$$

donde $\phi(M(\xi^*, \theta), M(\xi_x, \theta)) = \text{Tr}(M(\xi_x)M^{-1}(\xi^*)) - m$ (donde Tr denota la traza de la matriz) y ξ_x es el diseño que asigna toda su masa en x . Además, la derivada direccional $\phi(M(\xi^*), M(\xi_x)) = 0$ en los puntos de soporte del diseño ξ^* .

Del corolario anterior se tiene un criterio útil para verificar que un diseño ξ^* sea D óptimo, ya que sólo basta con graficar la derivada direccional $\phi(M(\xi^*, \theta), M(\xi_x, \theta))$ para $x \in \chi$ y verificar que dicha gráfica sea menor o igual a cero en todo el rango de regresión y que además en los puntos de soporte del diseño ξ^* se cumple la igualdad.

3.2. Transformación del modelo

Como se mencionó al principio de este capítulo, con frecuencia los errores del modelo considerado exhiben tanto heterocedasticidad como no normalidad. Sin embargo, el modelar la varianza del error no resuelve el problema de la no normalidad. Ahora, si bien el supuesto de normalidad no es necesario en la estimación por mínimos cuadrados generalizados, en lo que respecta a los diseños óptimos si es un requerimiento importante. Por lo tanto, la metodología de transformar el modelo surge como una herramienta útil a la hora de encontrar los diseños óptimos para un modelo heterocedástico. En esta sección se estudiarán algunas de las transformaciones más importantes.

3.2.1. Transformaciones de Box-Cox

Una de las transformaciones más conocidas son las propuestas por Box-Cox (1964):

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log y & \text{si } \lambda = 0 \end{cases} \quad (3-52)$$

para $y > 0$ (notar que $y^{(\lambda)}$ y y^λ son diferentes).

Esta transformación es continua para cualquier valor de λ . Si $\lambda = 0$, la transformación adecuada es la logarítmica. Para $\lambda \neq 0$, las transformaciones de Box y Cox son una familia de transformaciones potencia. Así, para $\lambda = 2$ se obtiene la transformación cuadrática, para $\lambda = 1/2$ la transformación raíz cuadrada, para $\lambda = -1$ la transformación recíproca. El valor de $\lambda = 1$ indica que ninguna transformación es necesaria (Ritz y Streibig, 2008).

En algunos casos la transformación deberá hacerse antes de la etapa de diseño, por lo que el investigador deberá tener algún conocimiento previo del problema que le dé alguna idea del valor posible del parámetro λ . Por ejemplo, si la variable respuesta es un volumen, la transformación adecuada podría ser una raíz cúbica. Mientras que una raíz cuadrada podría ser adecuada si la respuesta es un conteo. En el caso de que la respuesta tenga unos órdenes de magnitud grandes, la transformación logarítmica podría ser adecuada (Atkinson y Cook, 1997).

Usando la aproximación de Taylor de primer orden para una variable aleatoria Z :

$$E(g(Z)) \approx g(E(Z)) \quad (3-53)$$

con $g(z) = z^{(\lambda)}$, se tiene que

$$E[Y^{(\lambda)}] \approx (E[Y])^{(\lambda)} \quad (3-54)$$

y se llega a que la transformación de Box y Cox (3-52) logra homogeneizar la varianza:

$$\begin{aligned} \text{var}[Y^{(\lambda)}] &\approx \text{var}[Y] \left[\frac{dY^{(\lambda)}}{dy} \right]_{y=E[Y]}^2 \\ &= \text{var}[Y] (E[Y])^{2\lambda-2} \end{aligned} \quad (3-55)$$

para algún λ donde $\text{var}[Y^{(\lambda)}]$ sea constante (Seber y Wild, 2003).

3.2.2. Diseños D-óptimos para el modelo transformado

El problema de hallar diseños D-óptimos locales cuando se aplica alguna de las transformaciones de Box-Cox a la variable respuesta, ha sido abordado por Atkinson y Cook (1997). En particular los autores suponen que después de aplicar la transformación adecuada que homogeneiza la varianza de los errores, se tiene un modelo lineal de la forma

$$Y^{(\lambda)} = f^T(x)\beta + \epsilon^*, \quad (3-56)$$

donde los nuevos errores $\epsilon_i^* \sim N(0, \sigma^2)$ para algún valor de λ .

Matriz de información para el modelo transformado

Para el modelo transformado (3-56), la densidad condicional de Y_i dado x_i es

$$k(Y_i|x_i, \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(Y_i^{(\lambda)} - f^T(x_i)\beta)^2}{2\sigma^2} \right] J_i(\lambda), \quad (3-57)$$

donde $\theta^T = (\beta, \lambda, \sigma^2)$ es el vector de parámetros, $J_i(\lambda)$ es el jacobiano de la transformación potencia y está dado por (Carroll y Ruppert, 1988):

$$J_i(\lambda) = \frac{\partial Y^{(\lambda)}}{\partial Y}. \quad (3-58)$$

Entonces se calcula el valor esperado respecto a la distribución en (3-57) (Atkinson y Cook, 1997):

$$\begin{aligned}
I(\theta) &= -E \left[\frac{\partial^2 \log f(Y_i|x_i, \theta)}{\partial \theta^2} \right] \\
&= \begin{pmatrix} ff^T & 0 & -\frac{fE(\dot{Y}^{(\lambda)})}{\sigma^2} \\ & \frac{1}{2\sigma^4} & -\frac{E(\epsilon^* \dot{Y}^{(\lambda)})}{\sigma^4} \\ & & \frac{E(\dot{Y}^{(\lambda)})^2 + E(\epsilon^* \ddot{Y}^{(\lambda)})}{\sigma^2} \end{pmatrix} \quad (3-59)
\end{aligned}$$

donde $\epsilon^* = Y^{(\lambda)} - f^T(x)\beta$, $f = f(x)$ y $\dot{Y}^{(\lambda)}$, $\ddot{Y}^{(\lambda)}$ indican la primera y segunda derivada respecto a λ , respectivamente y están dados por:

$$\dot{Y}^{(\lambda)} = \frac{Y^\lambda \log Y^\lambda - Y^\lambda + 1}{\lambda^2} \quad \text{y} \quad (3-60)$$

$$\ddot{Y}^{(\lambda)} = \frac{Y^\lambda (\log Y^\lambda - 1)^2 + Y^\lambda - 2}{\lambda^3} \quad (3-61)$$

y de (3-59) la matriz de información para el modelo transformado (3-56) está dada por

$$M(\xi, \theta) = \int_{\chi} I(\theta) \xi(dx), \quad (3-62)$$

donde ξ denota la medida de probabilidad definida sobre la región de diseño χ .

Sin embargo, la expresión obtenida para la matriz de información en (3-59) tiene el problema de que no se pueden calcular los valores esperados para $\dot{Y}^{(\lambda)}$ y $\ddot{Y}^{(\lambda)}$. Ahora, por (3-52) y (3-56) se tiene que:

$$E(Y^{(\lambda)}) = E\left(\frac{Y^\lambda - 1}{\lambda}\right) = f^T(x)\beta, \quad (3-63)$$

de donde se sigue que

$$\mu(x|\theta) = E(Y^\lambda) = \lambda f^T(x)\beta + 1. \quad (3-64)$$

Luego, usando la aproximación de Taylor de segundo orden alrededor de $\mu(x|\theta) = E(Y^\lambda)$ (para $\lambda \neq 0$):

$$E(g(Z)) \approx g(E(Z)) + \frac{1}{2} \left(\frac{\partial^2 g(z)}{\partial z^2} \right) \Big|_{z=E(Z)} \text{var}(Z). \quad (3-65)$$

Aplicando esta aproximación con $Z = Y^\lambda$ se tiene

$$\frac{E(\dot{Y}^{(\lambda)})}{\sigma^4} \approx \frac{\log \mu(x|\theta)}{\lambda \sigma^2} \quad \text{y} \quad (3-66)$$

$$\frac{E(\ddot{Y}^{(\lambda)})}{\sigma^2} \approx \frac{\log^2 \mu(x|\theta)}{\lambda^2}. \quad (3-67)$$

Además, usando la aproximación para la varianza de una variable aleatoria se tiene que

$$E(\dot{Y}^{(\lambda)})^2 \approx E(\dot{Y}^{(\lambda)})^2 + \frac{\log^2 \mu(x|\theta)}{\lambda^2} \sigma^2. \quad (3-68)$$

Finalmente sustituyendo estas tres aproximaciones en (3-59) se tiene la siguiente aproximación

$$I(\theta) \approx I_a(\theta) = \begin{pmatrix} ff^T & 0 & -\frac{fE(\dot{Y}^{(\lambda)})}{\sigma^2} \\ & \frac{1}{2\sigma^4} & -\frac{\log \mu(x|\theta)}{\lambda\sigma^2} \\ & & \frac{\log^2 \mu(x|\theta)}{\lambda^2} + \frac{E(\dot{y}^{(\lambda)})^2}{\sigma^2} \end{pmatrix} \quad (3-69)$$

donde (si $\lambda \neq 0$):

$$E(\dot{Y}^{(\lambda)}) \approx \frac{\mu(x|\theta) \log \mu(x|\theta) - \mu(x|\theta) + 1}{\lambda^2}. \quad (3-70)$$

Para $\lambda = 0$ se tiene la siguiente aproximación:

$$I(\theta) \approx I_a(\theta, \lambda = 0) = \begin{pmatrix} \frac{ff^T}{\sigma^2} & 0 & -\frac{f(f^T\beta)^2}{2\sigma^2} \\ & \frac{1}{2\sigma^4} & -\frac{f^T\beta}{\sigma^2} \\ & & 2(f^T\beta)^2 + \frac{(f^T\beta)^4}{4\sigma^2} \end{pmatrix} \quad (3-71)$$

Matriz de información total

Reemplazando (3-69) o (3-71) en (3-62) se obtiene la matriz de información total aproximada

$$M_a(\xi, \theta) = \int_{\mathcal{X}} I_a(\theta) \xi(dx), \quad (3-72)$$

y si el diseño es aproximado

$$\xi = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ w_1 & w_2 & \cdots & w_n \end{bmatrix} \quad (3-73)$$

entonces

$$M_a(\xi, \theta) = \sum_{i=1}^n I_a(\theta, x_i) w_i. \quad (3-74)$$

Finalmente, con el fin de implementar esta matriz de información en un algoritmo, resulta conveniente descomponer la matriz de información en la forma:

Para $\lambda \neq 0$ como

$$I_a(\theta, x) = z_1 z_1^T + z_2 z_2^T \quad (3-75)$$

$$\text{donde } z_1(x, \theta) = \begin{pmatrix} f/\sigma \\ 0 \\ -E(\dot{Y}^{(\lambda)})/\sigma \end{pmatrix}_{(p+2) \times 1} \quad \text{y} \quad z_2(x, \theta) = \begin{pmatrix} 0 \\ 1/\sqrt{2}\sigma^2 \\ -\sqrt{2}\log \mu(x, \theta)/\lambda \end{pmatrix}_{(p+2) \times 1}.$$

Para $\lambda = 0$

$$z_1(x, \theta) = \begin{pmatrix} f/\sigma \\ 0 \\ -(f^T \beta)^2/2\sigma \end{pmatrix}_{(p+2) \times 1} \quad \text{y} \quad z_2(x, \theta) = \begin{pmatrix} 0 \\ 1/\sqrt{2}\sigma^2 \\ -\sqrt{2}f^T \beta \end{pmatrix}_{(p+2) \times 1}.$$

Así que en general la matriz de información total para el diseño ξ se puede expresar como

$$\begin{aligned} M_a(\xi, \theta) &= \sum_{i=1}^n z_1(x_i, \theta) z_1(x_i, \theta)^T w_i + \sum_{i=1}^n z_2(x_i, \theta) z_2(x_i, \theta)^T w_i \\ &= KWK^T + LWL^T, \end{aligned} \quad (3-76)$$

donde $K_{(p+2) \times n} = [z_1(x_1, \theta), \dots, z_1(x_n, \theta)]$, $L_{(p+2) \times n} = [z_2(x_1, \theta), \dots, z_2(x_n, \theta)]$ y $W = \text{diag}[w_1, \dots, w_n]$.

Teorema General de Equivalencia

Usando la estructura que tiene la matriz de información total, $M_a(\xi, \theta)$, se define

$$d_j(x, \theta, \xi) = z_j^T(x, \theta) M_a^{-1}(\xi, \theta) z_j(x, \theta), \quad (3-77)$$

para $j = 1, 2$. Entonces la función de sensibilidad para el diseño ξ se puede expresar como

$$d(x, \theta, \xi) = d_1(x, \theta, \xi) + d_2(x, \theta, \xi) \quad (3-78)$$

y con el fin de verificar que el diseño encontrado sea D óptimo se tiene la siguiente versión del Teorema General de Equivalencia (Atkinson y Cook, 1997):

Teorema 3.2.1. [Teorema de Equivalencia D optimalidad]

Sea Ξ el conjunto de todas las medidas de probabilidad ξ sobre un conjunto compacto χ . Entonces las siguientes condiciones caracterizan equivalentemente al diseño ξ_D como diseño D óptimo:

1. $\xi_D = \arg \max_{\Xi} \log |M_a(\xi, \theta)|$
2. $\min_{\Xi} \max_{\chi} d(x, \theta, \xi) = \max_{\chi} d(x, \theta, \xi_D)$
3. $\max_{\chi} d(x, \theta, \xi_D) = p + 2$.

La condición 1 establece la forma como se realiza el proceso de optimización: se busca el diseño ξ_D que maximice el logaritmo del determinante de la matriz de información total. La segunda condición establece la equivalencia entre los criterios D y G optimalidad. Finalmente la condición 3 establece una forma de verificar que el diseño ξ_D es D óptimo, pues al graficar la función de sensibilidad $d(x, \theta, \xi)$ como función de x se debe cumplir que la derivada direccional es igual a $p + 2$ en los puntos de soporte del diseño ξ_D .

D eficiencia

El diseño ξ_D es D óptimo local, ya que depende del valor local para el vector de parámetros, θ_0 . Así que una mala elección de este valor local dará como resultado un diseño ineficiente. En particular una mala elección del valor de λ , puede resultar en diseños cuyo modelo está mal especificado y cuya eficiencia respecto al diseño D óptimo es muy baja. Para un valor de θ fijo, una forma de cuantificar la eficiencia de un diseño ξ respecto al diseño ξ_D es mediante la D eficiencia que está dada por

$$D_{eff}(\xi, \theta) = \left(\frac{|M(\xi, \theta)|}{|M(\xi_D, \theta)|} \right)^{1/p}, \quad (3-79)$$

donde p es el número de parámetros del modelo. Ésta es una medida de que tan eficiente es el diseño ξ comparado con el diseño D óptimo y entre más cercano a uno sea, más eficiente será el diseño aunque no sea D óptimo.

En particular, Atkinson (2005) encuentra diseños D óptimos locales para varios valores de λ cuando se aplica la misma transformación en un modelo no lineal de decaimiento exponencial de primer orden y sus resultados muestran que para valores de λ cercanos a 1, los diseños son bastante robustos en el sentido que se obtienen eficiencias bastante altas cuando se aleja del valor verdadero.

4. Ejemplos

A continuación se mostrará un ejemplo donde se aplicarán y compararán las dos metodologías mencionadas en el capítulo anterior.

4.1. Ejemplo: Modelo heterocedástico no lineal

Los siguientes son los datos de un estudio (Bates y Watts, 1988) donde se midió la concentración de *policlorobifenilos* (PCB) a un grupo de 28 truchas del lago Cayuga, New York, a distintas edades (en años). En la tabla 4-1 se muestra los datos completos y en la figura 4-1 la gráfica de los datos:

Tabla 4-1.: Datos ejemplo modelo heterocedástico

Edad	1.00	1.00	1.00	1.00	2.00	2.00	2.00
Concentración	0.60	1.60	0.50	1.20	2.00	1.30	2.50
Edad	3.00	3.00	3.00	4.00	4.00	4.00	5.00
Concentración	2.20	2.40	1.20	3.50	4.10	5.10	5.70
Edad	6.00	6.00	6.00	7.00	7.00	7.00	8.00
Concentración	3.40	9.70	8.60	4.00	5.50	10.50	17.50
Edad	8.00	8.00	9.00	11.00	12.00	12.00	12.00
Concentración	13.40	4.50	30.40	12.40	13.40	26.20	7.40

Como se observa en la figura 4-1, cuando las edades de las truchas es pequeña, también lo es la variabilidad en la concentración de *policlorobifenilos* (PCB) y a medida que las edades de las truchas son mayores, también lo es la variabilidad. Es decir, que parece ser que a medida que aumenta la edad de las truchas la variabilidad es cada vez mayor, lo que sugiere que la varianza del error no es constante, lo cual se corroborará más adelante en el análisis de los residuales.

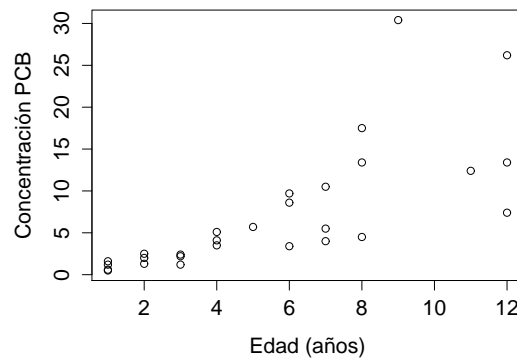


Figura 4-1.: Gráfico de los datos

4.1.1. Ajuste del modelo no lineal

En primer lugar se necesita ajustar un modelo al conjunto de datos y como se observa en la figura 4-1, la relación entre las variables parece ser de tipo no lineal. De los distintos modelos que se intentaron ajustar, el que se propone ajustar el siguiente modelo:

$$Y = \beta_1 e^{\beta_2 x} + \epsilon, \quad (4-1)$$

con β_1, β_2 parámetros a estimar.

Ahora, como se observa en el gráfico de los residuales estandarizados (figura 4-2), hay una clara tendencia de parlante que sugiere que la varianza del error no es constante y que aumenta en función del valor esperado de la variable respuesta. Para corroborar esta observación se dividieron las observaciones en dos grupos: un primer grupo de las truchas que tenían edades entre 1 y 4 años y un segundo grupo con edades entre 5 y 12 años. Luego se ajustó el modelo (4-1) y se realizó la prueba de Levene, obteniéndose un valor p de 0.001504, lo cual corrobora que el modelo bajo estudio es heterocedástico. También se realizó la prueba de Shapiro-Wilk y se obtuvo un valor p de 0.00018, por lo que la normalidad de los errores tampoco es válida. Sin embargo, como en la primera metodología se requiere la normalidad de los errores, entonces después de analizar los residuales estandarizados, se optó por eliminar las observaciones 24 y 28, que tienen los residuales más altos en valor absoluto. Luego se volvió a ajustar el modelo (4-1) y se obtuvo un valor p de 0.1215 para la prueba de normalidad.

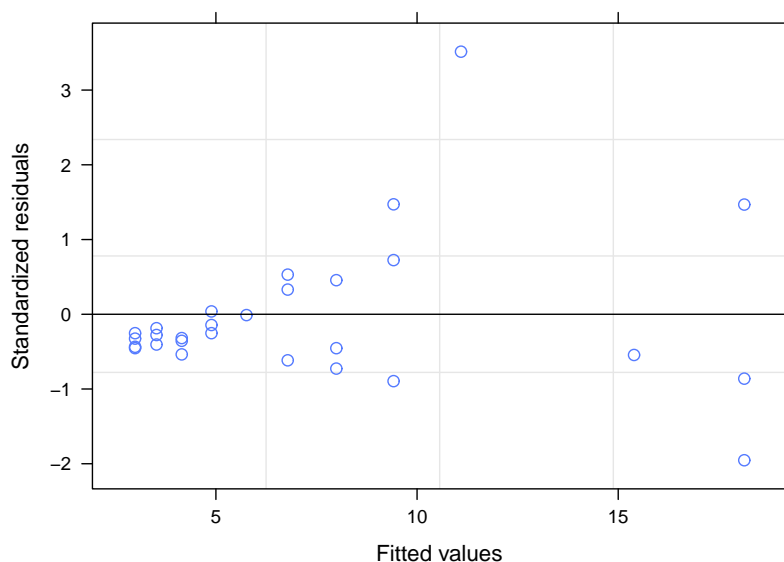


Figura 4-2.: Gráfico de los residuales estandarizados

En la tabla 4-2 se muestran los resultados al ajustar del modelo no lineal mediante la función *nls* de R. Como se observa en la tabla ambos parámetros son significativos. Como el modleo propuesto es no lineal, estos valores serán usados como valores iniciales para ajustar los modelos en cada una de las metodologías.

Tabla 4-2.: Ajuste modelo (4-1)

	Estimate	Std. Error	t value	Pr(> t)
β_1	1.87	0.53	3.52	0.00
β_2	0.20	0.03	6.93	0.00

4.2. Modelación de la varianza del error

Como se observa en la figuras 4-1 y 4-2 la variabilidad de la concentración de PCB aumenta a medida que aumenta la edad de las truchas, y además dada la estructura de parlante pareciera ser adecuado modelar la varianza del modelo como una función proporcional a una función que depende de la edad. En base a esto, se propone como una alternativa de solución incorporar una estructura para la varianza de los errores dada en el siguiente modelo:

$$Y = \beta_1 e^{\beta_2 x} + \epsilon, \quad (4-2)$$

donde suponemos que los errores $\epsilon \sim N(0, \sigma^2(\beta_1 e^{\beta_2 x})^{2\tau})$. Este resulta como un caso particular del modelo dado en la ecuación (3-4) página 12. El parámetro τ es un valor desconocido y puede ser tanto positivo como negativo, dependiendo de la relación funcional que exista entre la varianza de la respuesta Y y su media. Además, la estimación del vector de parámetros $\theta = (\beta_1, \beta_2, \tau, \sigma)$ se hace por el método de mínimos cuadrados no lineales generalizados. En la figura 4-3 se muestra la gráfica del ajuste obtenido y en la tabla 4-3 se muestra el resultado del ajuste del modelo mediante dicho método. Estos valores serán de gran utilidad para hallar el diseño D óptimo, pues pueden ser usados como un valor local del vector de parámetros.

Tabla 4-3.: Ajuste modelo (4-2)

Parámetro	Estimación
β_1	0.97
β_2	0.29
τ	1.12
σ	0.37

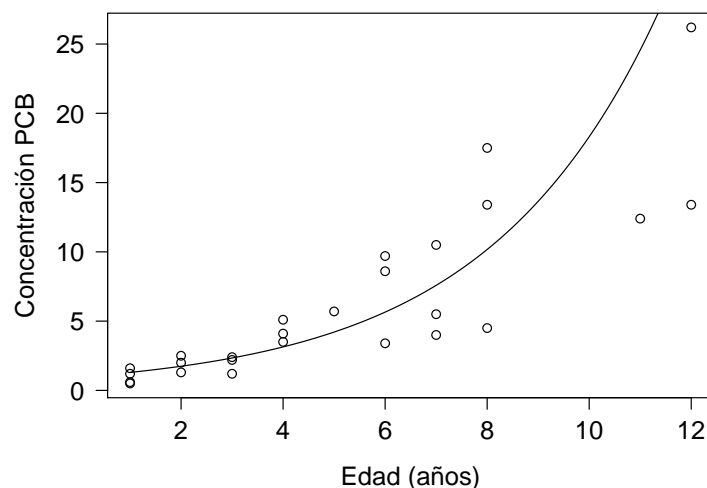


Figura 4-3.: Gráfica del modelo ajustado 4-2

Ahora, con el fin de determinar si el modelo (4-2) es más adecuado que el modelo con varianza constante (4-1), se realizó la prueba de razón de verosimilitudes entre los dos modelos, lo que equivale a realizar la siguiente prueba de hipótesis:

$$H_0 : \tau = 0 \text{ vs } H_1 : \tau \neq 0 \quad (4-3)$$

En la tabla 4-4 se muestra los resultados de la prueba, donde el modelo 1 es el modelo (4-2) donde se modela la varianza como una función potencia de la media y el modelo 2 es el (4-1) con varianza contante:

Tabla 4-4.: Tabla Anova para contrastar el modelo 1 con el 2

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
modelo (4-2)	1	4	114.5069	119.5392	-53.25343			
modelo (4-1)	2	3	142.2971	146.0714	-68.14856	1 vs 2	29.79026	< .0001

La razón de verosimilitud entre los dos modelos es de 29.79026 y el valor p de la prueba es menor que 0.0001, así que se concluye con un nivel de significación del 1%, el rechazo de la hipótesis nula. Luego $\tau \neq 0$ y por lo tanto el modelo 1 es más adecuado para ajustar los datos.

4.2.1. Diseño D óptimo local

Para encontrar el diseño D óptimo local para el modelo en (4-2) con la estructura de varianza se necesita un valor local para el vector de parámetros $\theta = (\beta_1, \beta_2, \tau, \sigma)$. Este valor local se toma como la estimación por mínimos cuadrados no lineales generalizados obtenida en la sección anterior: $\theta_0 = (0.97, 0.29, 1.12, 0.37)$. Luego se encuentra la matriz de información para el modelo usando el corolario 3.1.2 de la sección 3.1.1 del capítulo 3 y se minimiza $-\ln(|M(\xi, \theta)|)$ usando la función *nlinb* de R para el proceso de optimización con rango de regresión para la variable edad $\chi = [1, 12]$ (ver detalles del algoritmo en el anexo A).

En la tabla 4-5 se muestra el diseño D óptimo local, ξ_D , para el modelo con estructura de varianza (4-2) usando como valor local θ_0 . Nótese que aunque se deben estimar cuatro parámetros, el diseño D óptimo local solo consta de dos puntos con igual peso, y estos son precisamente los extremos del intervalo $[1, 12]$ sobre el que se optimizó el funcional de la matriz de información. Ahora, nótese que para haber obtenido este diseño, se necesitaron los datos con los cuales hacer el ajuste para poder obtener un valor local razonable. En este sentido, como si se dispone de los datos, el diseño D-óptimo obtenido nos dice que para estimar con mínima varianza el vector de parámetros θ del modelo bajo estudio, si se volviera a realizar el experimento sería suficiente con medir la concentración de *policlorobifenilos*

(PCB) en truchas con edades de 1 y 12 años respectivamente y con un número igual de réplicas. Así, por ejemplo si se pudieran realizar 10 corridas experimentales, el diseño D óptimo obtenido sugeriría tomar cinco mediciones de *policlorobifenilos* a truchas con un año de edad y otras cinco mediciones en truchas con 12 años de edad. En este sentido los diseños óptimos son una técnica que no solo permite estimar de forma óptima los parámetros del modelo bajo estudio, sino que también implica un ahorro en cuanto a los costos del diseño.

Tabla 4-5.: Diseño D óptimo local

x_i	1.00	12.00
w_i	0.50	0.50

Ahora, se debe verificar que el diseño obtenido es D óptimo mediante el Teorema General de Equivalencia (3.1.3) de la sección 3.1.1. Para ello si $M(\xi_D, \theta)_{4 \times 4}$ denota la matriz de información del diseño encontrado ξ_D , entonces el diseño ξ_D es D óptimo si la derivada direccional de Ψ en ξ_D en la dirección del diseño que asigna toda su masa en x , ξ_x , cumple

$$\phi(M(\xi_D, \theta), M(\xi_x, \theta)) \leq 0 \quad \forall x \in \chi = [1, 12], \quad (4-4)$$

donde $\phi(M(\xi_D, \theta), M(\xi_x, \theta)) = \text{Tr}(M(\xi_x)M^{-1}(\xi_D)) - 4$. Como se observa en la figura 4-4, la derivada direccional siempre está por debajo de cero, así que cumple la condición. Además también se cumple que la derivada direccional $\phi(M(\xi_D), M(\xi_x)) = 0$ en los puntos de soporte del diseño 1 y 12. Por lo tanto se verifica que el diseño ξ_D es D óptimo local.

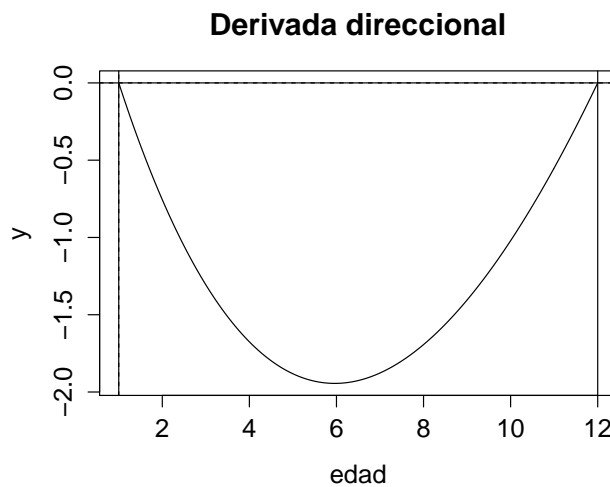


Figura 4-4.: Gráfica de la derivada direccional del diseño

4.2.2. Diseños D óptimos variando el parámetro τ

El diseño D óptimo local ξ_D depende del valor local para el vector de parámetros. En especial interesa determinar qué tan sensible es el diseño cuando se varía el parámetro τ . Para ello se fijaron los parámetros β_1, β_2 y σ en los valores locales de la sección anterior y se varió el parámetro τ en un rango plausible de valores. Luego, para cada valor local se encontró el diseño D óptimo. En la tabla 4-6 se muestra los distintos diseños D óptimos para valores de τ entre 0.1 y 2 y se observa que:

1. Para $0.1 \leq \tau \leq 0.4$ el diseño D óptimo tiene tres puntos de soporte, donde siempre se encuentran los extremos (1 y 12) y el punto intermedio varía desde 8.28 (para $\tau = 0.1$) hasta 6.11 para $\tau = 0.4$.
2. Para $0.5 \leq \tau \leq 1.5$ el diseño consta de solo dos puntos de soporte (los extremos 1 y 12).
3. Para $1.6 \leq \tau \leq 2$ el diseño D óptimo tiene tres puntos de soporte, dos de ellos son los extremos del intervalo $[1, 12]$ y el punto intermedio varía desde 6.89 hasta 4.32.

Tabla 4-6.: Tabla de diseños D-óptimos variando τ

τ	x_1	x_2	x_3	w_1	w_2	w_3
0.10	1.00	8.28	12.00	0.27	0.28	0.45
0.20	1.00	7.76	12.00	0.29	0.25	0.46
0.30	1.00	7.08	12.00	0.33	0.19	0.47
0.40	1.00	6.11	12.00	0.42	0.09	0.49
0.50	1.00		12.00	0.50		0.50
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
1.50	1.00		12.00	0.50		0.50
1.60	1.00	6.89	12.00	0.49	0.09	0.42
1.70	1.00	5.92	12.00	0.47	0.19	0.33
1.80	1.00	5.24	12.00	0.46	0.25	0.29
1.90	1.00	4.72	12.00	0.45	0.28	0.27
2.00	1.00	4.32	12.00	0.44	0.30	0.26

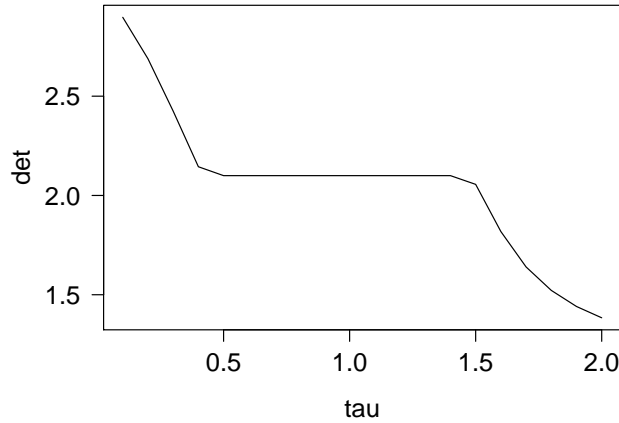


Figura 4-5.: Razón entre los determinantes de $M(\xi_1, \beta, \tau, \sigma)$ y $M(\xi_2, \beta, \tau + 0.1, \sigma)$

En particular se observa que para $0.5 \leq \tau \leq 1.5$, el diseño D óptimo local consta de los mismos dos puntos de soporte, esto se debe en parte a que como se muestra en la figura (4-5), el cociente:

$$\frac{|M(\xi_1, \beta, \tau, \sigma)|}{|M(\xi_2, \beta, \tau + 0.1, \sigma)|}, \quad (4-5)$$

entre el determinante de la matriz de información de un diseño específico para un valor de τ y el respectivo determinante cuando se incrementa en 0.1 el parámetro τ , es constante y por lo tanto la relación entre las matrices de información para los dos diseños es de la forma:

$$M(\xi_1, \beta, \tau, \sigma) = M(\xi_2, \beta, \tau + 0.1, \sigma)C \quad (4-6)$$

donde C es una matriz de constantes (no depende de x ni del vector de parámetros) y por lo tanto no influye en el proceso de optimización del determinante de la matriz de información. Por el contrario, para los valores de τ con los que el diseño tiene tres puntos de soporte, dicho cociente no es constante y por lo tanto los diseños si son diferentes.

4.2.3. Simulación

Con el fin de evaluar la bondad del ajuste para el vector de parámetros θ en el modelo

$$Y_i = \beta_1 e^{\beta_2 x_i} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2 (\beta_1 e^{\beta_2})^{2\tau}), \quad \text{para } i = 1, 2, \dots, N, \quad (4-7)$$

mediante el diseño D óptimo local mostrado en la tabla 4-7, se propone la siguiente estrategia de simulación:

Tabla 4-7.: Diseño D óptimo local: Modelando la varianza

x_i	1.00	12.00
w_i	0.50	0.50

1. Fijar el número de observaciones N que se van a simular en el modelo 4-7 y el número de simulaciones $Nsim$ que se van a hacer.
2. Usando como valor verdadero del vector de parámetros $\theta = (\beta_1, \beta_2, \tau, \sigma) = (0.97, 0.29, 1.12, 0.37)$ (ver tabla 4-3), evaluar la parte determinística $\eta(x_i, \theta) = \beta_1 e^{\beta_2 x_i}$ en los puntos de soporte x_i del diseño D óptimo local ξ_D con el número de réplicas w_i , dado por los pesos del diseño.
3. Simular los errores $\epsilon_i \sim N(0, \sigma^2 (\beta_1 e^{\beta_2})^{2\tau})$ para $i = 1, 2, \dots, N$.
4. Con los resultados de los pasos 2 y 3, hallar la respuesta y_i para $i = 1, 2, \dots, N$ mediante el modelo (4-7).
5. Ajustar el modelo no lineal (4-7) con los datos simulados y estimar el vector de parámetros θ , denotado por $\hat{\theta}$.
6. Calcular el error relativo, ER , para el vector de parámetros:

$$ER = \frac{\|\theta - \hat{\theta}\|}{\|\theta\|} \times 100\%. \quad (4-8)$$

7. Calcular el error cuadrático medio (ECM) para cada parámetro individual y para el vector de parámetros θ se propone una medida de discrepancia global (MDG) entre los vectores θ y $\hat{\theta}$ dada por $MDG(\hat{\theta}) = \|\theta - \hat{\theta}\|^2$, donde la norma es la norma Euclídeana.
8. Realizar $Nsim$ simulaciones de los pasos anteriores y calcular algunas medidas descriptivas (media, mediana, el rango y el coeficiente de variación) tanto para el error relativo ER como para el error cuadrático medio.

Se implementó la estrategia anterior con $Nsim = 1000$ simulaciones. Como la base de datos depurada tiene 26 observaciones, se tomó éste como el número de corridas experimentales. Los resultados se resumen en la tabla 4-8, donde se muestra las medidas descriptivas (media, mediana, rango y desviación estándar) del error relativo (en porcentaje) y los errores cuadráticos medios tanto para el vector de parámetros θ como para cada parámetro individual. Como se observa en la tabla, en cuanto a las medidas de tendencia central, la media y la mediana del error cuadrático medio ECM para cada parámetro individual oscilan en orden de 10^{-4} hasta 10^{-2} . Para la medida de discrepancia global (MDG), la media y la mediana son también pequeñas (del orden de 10^{-2}). En cuanto al error relativo, la media y mediana es de 10.2% y 9.52% respectivamente. En cuanto a las medidas de variabilidad, en especial si se observa la desviación estándar, en general se obtuvo poca variabilidad, siendo la desviación más pequeña la obtenida para el error cuadrático medio del parámetro β_2 . En resumen todas estas medidas son una buena medida en cuanto a la bondad del ajuste dada por el diseño D óptimo encontrado.

Tabla 4-8.: Resultados simulaciones modelando la varianza

	$ECM(\beta_1)$	$ECM(\beta_2)$	$ECM(\tau)$	$ECM(\sigma)$	$MDG(\theta)$	$ER(\theta)\%$
Media	1.19e-02	2.64e-04	1.05e-02	7.90e-03	3.05e-02	10.2
Mediana	5.06e-03	1.23e-04	5.32e-03	4.36e-03	2.19e-02	9.52
Rango	1.33e-01	3.04e-03	1.14e-01	8.72e-02	2.20e-01	29.7
Desviación	1.64e-02	3.65e-04	1.39e-02	1.02e-02	2.83e-02	4.81

4.2.4. Eficiencias

Con el fin de determinar qué tan robusto es el diseño D óptimo en cuanto a la elección del valor local θ_0 , se calcularon las D-eficiencias del diseño para cada combinación de los parámetros perturbando un porcentaje Δ el vector de parámetros:

$$\theta \pm \Delta \times \theta. \quad (4-9)$$

Como se tienen cuatro parámetros y cada uno se puede perturbar un $\Delta\%$ a la izquierda; es decir $\theta_i - \Delta \times \theta_i, i = 1, 2, 3, 4$, un $\Delta\%$ a la derecha; es decir $\theta_i + \Delta \times \theta_i, i = 1, 2, 3, 4$ o no perturbarlo ($\Delta = 0$), se tienen $3^4 = 81$ combinaciones en total. Sin embargo, muchos de los diseños obtenidos coinciden con el diseño D óptimo. Así por ejemplo, para un porcentaje $\Delta = 60\%$ (no se consideran porcentajes menores a 50% ya que para esos diseños la D eficiencia es casi 1), en la tabla 4-9 se muestra los 36 diseños que tienen una D eficiencia menor a 1. Para cada parámetro, se usan los símbolos $-$, $+$ o 0 para indicar si el parámetro se perturba un 60% a la izquierda, a la derecha o no se perturba respectivamente. Así, por ejemplo, en el primer diseño, se perturbaron los parámetros β_1, τ y σ un 60% a la

izquierda, mientras que β_2 no se varió. El diseño correspondiente a esta combinación consta de 3 puntos de soporte: 1, 5.49 y 12 y la D eficiencia del diseño es bastante alta (0.9964). Como se observa, los 36 diseños constan de 3 puntos de soporte, dos de ellos siempre son los extremos del intervalo para la variable edad (recordar que estos fueron los puntos de soporte para el diseño D óptimo) y el punto intermedio varía para cada diseño desde 3.6 hasta 8.2, aunque como se observa en la tabla 4-9, los diseños están agrupados consecutivamente de a 3 de acuerdo al punto de soporte intermedio. También se calcularon las D eficiencias para un porcentaje $\Delta = 80\%$, aunque no se muestran los resultados, las D eficiencias que se obtuvieron fueron de al menos 0.8.

Tabla 4-9.: Tabla de D eficiencias y puntos de soporte perturbando un 60% el vector de parámetros: perturbar a la izquierda (-), a la derecha (+) o sin perturbar (0)

Diseño	β_1	β_2	τ	σ	x_1	x_2	x_3	D_{ef}	Diseño	β_1	β_2	τ	σ	x_1	x_2	x_3	D_{ef}
1	-	0	-	-	1	5.49	12	0.996	19	-	0	+	0	1	5.28	12	0.851
2	0	0	-	-	1	5.49	12	0.996	20	0	0	+	0	1	5.28	12	0.851
3	+	0	-	-	1	5.49	12	0.996	21	+	0	+	0	1	5.28	12	0.851
4	-	+	-	-	1	8.2	12	0.829	22	-	+	+	0	1	3.6	12	0.796
5	0	+	-	-	1	8.2	12	0.829	23	0	+	+	0	1	3.6	12	0.796
6	+	+	-	-	1	8.2	12	0.829	24	+	+	+	0	1	3.6	12	0.796
7	-	0	+	-	1	5.28	12	0.851	25	-	0	-	+	1	5.49	12	0.996
8	0	0	+	-	1	5.28	12	0.851	26	0	0	-	+	1	5.49	12	0.996
9	+	0	+	-	1	5.28	12	0.851	27	+	0	-	+	1	5.49	12	0.996
10	-	+	+	-	1	3.6	12	0.796	28	-	+	-	+	1	8.2	12	0.829
11	0	+	+	-	1	3.6	12	0.796	29	0	+	-	+	1	8.2	12	0.829
12	+	+	+	-	1	3.6	12	0.796	30	+	+	-	+	1	8.2	12	0.829
13	-	0	-	0	1	5.49	12	0.996	31	-	0	+	+	1	5.28	12	0.851
14	0	0	-	0	1	5.49	12	0.996	32	0	0	+	+	1	5.28	12	0.851
15	+	0	-	0	1	5.49	12	0.996	33	+	0	+	+	1	5.28	12	0.851
16	-	+	-	0	1	8.2	12	0.829	34	-	+	+	+	1	3.6	12	0.796
17	0	+	-	0	1	8.2	12	0.829	35	0	+	+	+	1	3.6	12	0.796
18	+	+	-	0	1	8.2	12	0.829	36	+	+	+	+	1	3.6	12	0.796

En la figura 4-6 se resumen los resultados de la tabla en cuanto a las eficiencias de los diseños obtenidos. Como se observa las D eficiencias de los diseños oscilan entre 0.8 y 1 aproximadamente, lo que muestra que aunque el diseño D-óptimo ξ_D es un diseño local, este es bastante robusto frente a la elección del valor local.

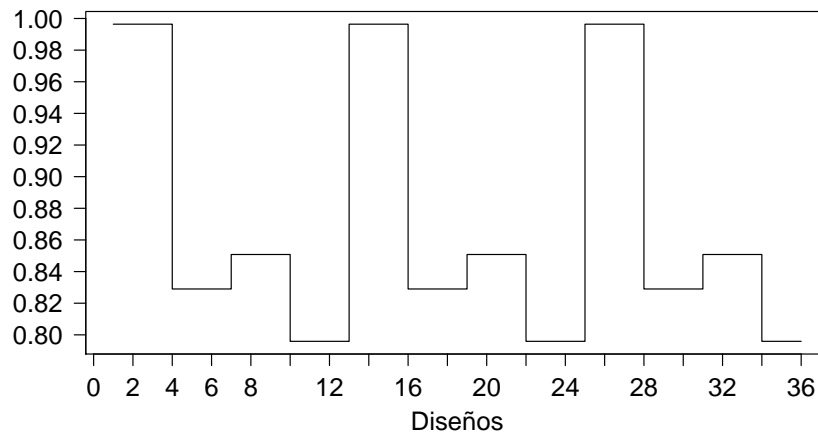


Figura 4-6.: D eficiencias perturbando el vector de parámetros θ un 60%

4.3. Segunda metodología: Transformación del modelo

4.3.1. Ajuste del modelo transformado

En la sección anterior se modeló la varianza de los errores y se encontró el diseño D óptimo local. Una segunda metodología es la de aplicar alguna transformación sobre el modelo para homogenizar la varianza y lograr la normalidad de los errores. En esta sección se aplicarán los métodos descritos en la sección 3.2 para la parte determinística del modelo 4-1:

$$Y = \beta_1 e^{\beta_2 x} + \epsilon. \quad (4-10)$$

Como el modelo transformado debe ser lineal en los parámetros, es claro que la transformación adecuada para este modelo es la logarítmica. Luego, aplicando logaritmo a ambos lados de (4-10) y suponiendo errores aditivos se tiene el modelo transformado:

$$\log Y = \log \beta_1 + \beta_2 x + \epsilon^*, \quad (4-11)$$

el cual se puede expresar como un modelo lineal de la forma:

$$Y^* = \beta_1^* + \beta_2 x + \epsilon^*, \quad (4-12)$$

donde $Y^* = \log Y$, $\beta_1^* = \log \beta_1$ y los nuevos errores ϵ_i^* son aproximadamente normales con varianza constante. En la figura 4-7 se muestra la gráfica de los datos transformados y los

gráficos de los residuales. Como se observa, comparado con los residuales del modelo original (figura 4-2), la transformación logra homogenizar la varianza, y al realizar la prueba de Levene el valor p fue de 0.3163 y al mismo tiempo la transformación hace que los errores cumplan el supuesto de normalidad (valor p prueba de Shapiro-Wilk 0.8442).

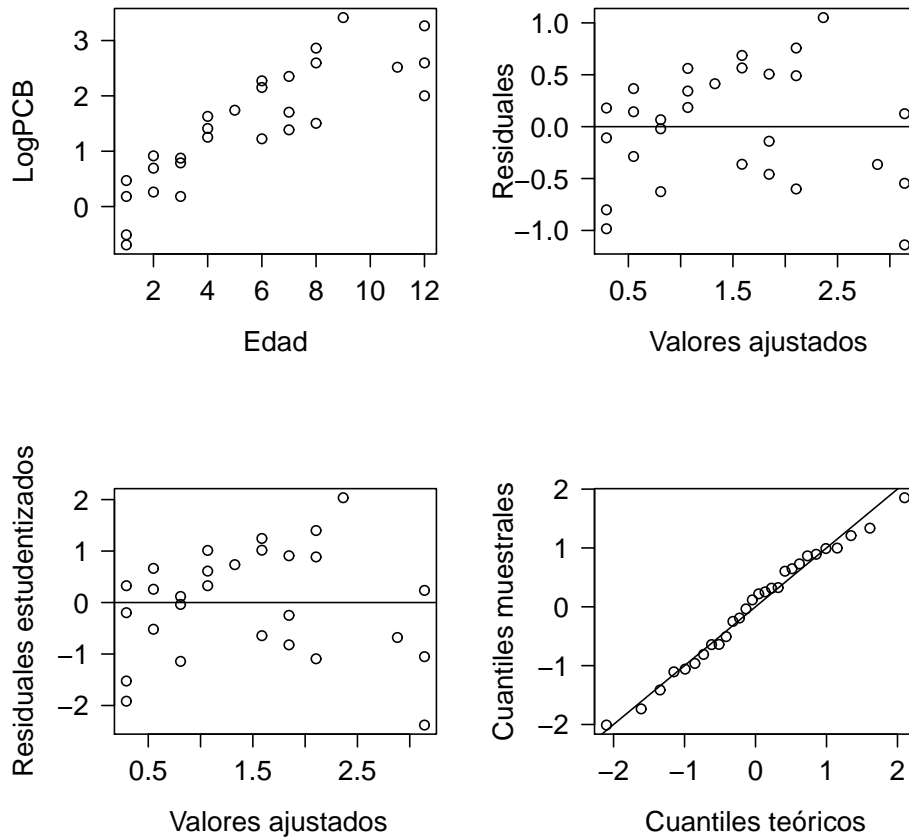


Figura 4-7.: Gráfico de dispersión de los datos transformados y gráficos de residuales

4.3.2. Diseño D óptimo para el modelo transformado

Con el fin de obtener un valor local para el vector de parámetros θ , se ajustó el modelo lineal en (4-12) y se obtuvo $\hat{\beta}^* = (0.03, 0.26)^T$ y $\hat{\sigma} = 0.57$, de donde la estimación para el vector de parámetros original es $\hat{\beta} = (e^{0.03}, 0.26)^T = (1.03, 0.26)^T$. Luego, usando la metodología

descrita en la sección 3.2.2 para el modelo transformado con $\lambda = 0$ se minimizó el logaritmo de la matriz de información sobre el rango de regresión $\chi = [1, 12]$ y en la tabla 4-10 se muestra el diseño obtenido. Como se observa se obtiene el mismo diseño que con la metodología anterior. Sin embargo, debe tenerse en cuenta que los diseños que se están obteniendo son locales, esto es dependen del valor local para el vector de parámetros. En este caso si bien en cada metodología se usaron distintos valores locales para el vector de parámetros θ , de manera independiente con cada una se obtuvo el mismo diseño de dos puntos. Así, como se vió en los resultados de la primera metodología, si no se tuviera un buen conocimiento del valor local para el vector de parámetros (en especial en cuanto al valor de la potencia τ) el diseño obtenido podría ser de 3 puntos y en ese caso no coincidiría con el diseño para el modelo transformado. En este caso como se tiene un buen conocimiento del valor local (pues se dispone de los datos completos), con cada metodología se obtuvo que el diseño D óptimo es un diseño de 2 puntos, pero nuevamente como se mostrará a continuación, con esta segunda metodología también podrían obtenerse diseños de 3 puntos de soporte dependiendo del valor local usado para el vector de parámetros. (se verificó que el diseño obtenido en la tabla 4-10 es D óptimo mediante el Teorema General de Equivalencia, obteniéndose una gráfica de sensibilidad igual a la obtenida con la primera metodología).

Tabla 4-10.: Diseño D óptimo para el modelo transformado

i	1	2
x_i	1.00	12.00
w_i	0.50	0.50

D eficiencias

Con el fin de determinar que porcentaje de eficiencia se pierde en un diseño ξ cuando se usa un valor equivocado del vector de parámetros como valor local, se calculan las D eficiencias como se definió en la sección 3.2.2:

$$D_{eff}(\xi, \theta) = \left(\frac{|M(\xi, \theta)|}{|M(\xi_D, \theta)|} \right)^{1/4}. \quad (4-13)$$

El procedimiento empleado para obtener las combinaciones posibles para el vector de parámetros es muy similar al usado para la primera metodología, excepto que en este caso solo se varían 3 parámetros: β_1, β_2 y σ , ya que el modelo transformado es lineal, y por lo tanto el parámetro $\lambda = 0$. Como son tres parámetros los que hay que variar y cada uno tiene 3 posibilidades (izquierda, derecha o no perturbar), se obtuvieron $3^3 = 27$ combinaciones en total, variando un $\Delta\%$ a la izquierda o la derecha cada posible combinación. Se encontró que

con $\Delta \leq 20\%$ todos los diseños coinciden con el diseño D óptimo de 2 puntos. Con $\Delta = 30\%$ solo se obtuvieron 3 diseños de 3 puntos, para $\Delta = 40\%$, 60% se obtienen exactamente los mismos 6 diseños de 3 puntos (variando la D eficiencia) y para $\Delta = 70, 80\%$ se obtienen exactamente los mismos 9 diseños (variando la D eficiencia) de 3 puntos. Todos estos diseños son de 3 puntos, dos de ellos los extremos del intervalo de regresión y el punto intermedio es 6.5. En la tabla 4-11 se muestran los 6 diseños, sus puntos de soporte (los pesos w_i se muestran en paréntesis) y sus respectivas eficiencias para un porcentaje de perturbación $\Delta = 60\%$. Como se observa, los 3 primeros diseños son iguales y por lo tanto tienen la misma eficiencia (0.929). Análogamente los 3 últimos diseños. En ambos diseños la eficiencia es bastante alta, indicando nuevamente que el diseño obtenido es bastante robusto en cuanto al valor local del vector de parámetros.

Tabla 4-11.: Tabla de puntos de soporte y D-eficiencias con un 60% de perturbación a la izquierda (-), a la derecha (+) o sin perturbar (0)

Diseño	β_1	β_2	σ	x_1	x_2	x_3	D_{ef}
1	-	0	-	1(0.401)	6.5(0.198)	12(0.401)	0.929
2	0	0	-	1(0.401)	6.5(0.198)	12(0.401)	0.929
3	+	0	-	1(0.401)	6.5(0.198)	12(0.401)	0.929
4	-	+	-	1(0.360)	6.5(0.280)	12(0.360)	0.892
5	0	+	-	1(0.360)	6.5(0.280)	12(0.360)	0.892
6	+	+	-	1(0.360)	6.5(0.280)	12(0.360)	0.892

4.3.3. Simulación

Con el fin de evaluar la bondad del ajuste para el vector de parámetros θ mediante el diseño D óptimo local se implementa una estrategia de simulación análoga a la usada para la metodología de modelar la varianza:

- Fijar el número de corridas experimentales N y el número de simulaciones N_{sim} que se van a hacer.
- Usando como valor verdadero del vector de parámetros $\theta^T = (\beta_1, \beta_2, \sigma) = (1.03, 0.26, 0.57)$ (obtenido con el ajuste para el modelo transformado), evaluar la parte determinística del modelo transformado en los puntos de soporte del diseño D óptimo local ξ_D con el número de réplicas dado por los pesos del diseño.
- Simular los errores $\epsilon_i \sim N(0, \sigma^2)$ para $i = 1, 2, \dots, N$ y hallar la respuesta y_i mediante el modelo (4-12).

- Ajustar el modelo lineal con los datos simulados y estimar el vector de parámetros θ , denotado por $\hat{\theta}$.
- Calcular el error relativo ER para $\hat{\theta}$: $ER(\hat{\theta}) = \frac{\|\theta - \hat{\theta}\|}{\|\theta\|} \times 100\%$, el error cuadrático medio, ECM, para β_1, β_2 y σ y análogamente a como se hizo con la primera metodología, se usa como una medida de discrepancia global: $MDG(\hat{\theta}) = \|\theta - \hat{\theta}\|^2$.
- Realizar $Nsim$ simulaciones de los pasos anteriores y calcular algunas medidas descriptivas para el error relativo ER , el error cuadrático medio ECM y la medida de discrepancia global, MDG , como la media, mediana, rango y desviación estándar.

Tabla 4-12.: Resultados simulaciones transformando el modelo

	$ECM(\beta_1)$	$ECM(\beta_2)$	$ECM(\sigma)$	$MDG(\theta)$	$ER(\theta)\%$
Media	1.83e-01	9.58e-04	1.30e-02	1.97e-01	34.1
Mediana	1.38e-01	6.35e-04	8.44e-03	1.55e-01	32.6
Rango	1.18e+00	6.66e-03	1.23e-01	1.17e+00	84.3
Desviación estándar	1.65e-01	1.02e-03	1.41e-02	1.64e-01	13.8

La tabla 4-12 muestra los resultados de las $Nsim = 1000$ simulaciones realizadas usando el procedimiento antes descrito. Así como para la metodología anterior, se muestran las medidas descriptivas (media, mediana, rango y desviación estándar) del error relativo (en porcentaje), de la medida de discrepancia global para el vector de parámetros θ y de los errores cuadráticos medios para cada parámetro individual. Como se observa en la tabla, en cuanto a las medidas de tendencia central, la media y la mediana del error cuadrático medio ECM para cada parámetro individual y el MDG para θ oscilan en orden de 10^{-4} hasta 10^{-1} e igual que en los resultados de la tabla 4-8, los errores más pequeños se obtuvieron para el parámetro β_2 . En cuanto al error relativo, la media y mediana (en porcentaje) fueron casi tres veces más grandes que los errores obtenidos con la metodología anterior (34.1 % y 32.6 % respectivamente). Por otro lado, la variabilidad de los errores cuadráticos fue pequeña, aunque para el error relativo se obtuvo una desviación estándar de 13.8, nuevamente casi tres veces que la obtenida con la primera metodología (4.81) En resumen todas estas medidas son una buena medida en cuanto a la bondad del ajuste dada por el diseño D óptimo encontrado.

5. Conclusiones y recomendaciones

5.1. Conclusiones

- Los diseños D óptimos locales dependen del modelo trabajado y en el caso no lineal del vector de parámetros. Además se basan en los supuestos de normalidad, homocedasticidad e independencia. En este trabajo se abordó el problema de encontrar diseños D óptimos locales cuando no se cumple el supuesto de homogeneidad de varianza.
- El problema de hallar diseños D óptimos locales cuando la varianza del error no es constante se diferencia del caso homogéneo principalmente en la estructura de la matriz de información. Esto se debe principalmente a la forma de la función de verosimilitud cuando la varianza no es constante, ya que en ese caso no solo se involucran los parámetros de la parte determinística sino otros adicionales. En este caso la matriz de información se deberá obtener a partir de las derivadas parciales de la función de verosimilitud que pueden tener unas expresiones bastante complejas.
- Para encontrar la estructura de la matriz de información usando la primera metodología que consiste en modelar la varianza del error, se debe tener claro el modelo y conocimiento de cuál es la función que mejor modela la varianza. En el teorema 3.1.1 se demostró cual es la estructura de la matriz de información para el caso general en que la función que modela la varianza es una función que depende de x y del vector de parámetros θ . En particular, en el ejemplo usado en este trabajo se modeló la varianza de la respuesta como una función potencia de la media. Para obtener un buen valor local del vector de parámetros se ajustó el modelo propuesto mediante el método de mínimos cuadrados no lineales generalizados y usando la estimación de θ como valor local se encontró que el diseño D óptimo local consiste de dos puntos de soporte. Aunque se deben estimar cuatro parámetros, el diseño D óptimo local solo consta de dos puntos con igual peso, y estos son precisamente los extremos del intervalo $[1, 12]$ sobre el que se optimizó el funcional de la matriz de información.
- Para la segunda metodología de transformar el modelo en homocedástico, se utilizó la transformación de Box y Cox y usando los resultados de Atkinson y Cook (1997), se encontró la estructura de la matriz de información cuando se aplica la transformación logarítmica al modelo original. En este caso la transformación vuelve el modelo lineal en los parámetros transformados y logra homogenizar la varianza y al mismo tiempo hacer que los errores del modelo transformado sean aproximadamente normales. Luego

usando la estimación obtenida con el ajuste del modelo transformado como valor local del vector de parámetros, se encontró que el diseño D óptimo cuando se transforma el modelo da exactamente el mismo que en la metodología anterior.

- Con ambas metodologías se obtuvo el mismo diseño de dos puntos de soporte, 1 y 12, con igual peso (0.5), aunque esto se debe al hecho de que los diseños encontrados son diseños locales y en este caso como se dispone de los datos completos, en cada metodología se usó el mejor valor local posible y con cada una de manera independiente se llega al mismo diseño. Sin embargo, como se mostró con cada metodología al calcular las D eficiencias para varios diseños, en ciertos casos se tienen diseños de 3 puntos y en ese caso no coincidirían los diseños entre ambos métodos.
- Una característica que se destaca en todos los diseños de 3 puntos es que siempre los extremos son los puntos de soporte del diseño D óptimo y el punto intermedio varía entre los extremos de acuerdo al porcentaje de variación y la combinación de los parámetros. En este sentido, para la segunda metodología se está cumpliendo el hecho que en un modelo lineal con dos parámetros los puntos de soporte del diseño D óptimo sobre un región de diseño $[a, b]$, son precisamente los extremos de dicho intervalo.
- En cuanto a las D eficiencias de los diseños obtenidos con las dos metodologías, se puede decir que no hay mucha diferencia, ya que en ambos casos la eficiencia obtenidas están por encima de 0.80, aún para porcentajes de perturbación del valor local θ_0 del 80 %. En este sentido el diseño D óptimo es bastante robusto en cuanto a la elección del valor local.
- De los resultados de las simulaciones cuando se modela la varianza se observa que en general los errores cuadráticos medios con esa metodología son más pequeños que los obtenidos transformando el modelo, aunque la diferencia no es muy grande. Sin embargo al comparar el error relativo, si se nota una gran diferencia, pues mientras que con la primera metodología se obtuvo un error relativo promedio del 10.2 %, transformando el modelo se obtiene un error relativo promedio 3 veces mayor (34.1 %). En este sentido la primera metodología sería mucho mejor que la segunda.
- Una desventaja de la primera metodología es que basa en el supuesto de normalidad, mientras que con la segunda metodología dicho supuesto no es necesario aunque la transformación adecuada debe ser tal que los errores del modelo transformado sean aproximadamente normales. En el ejemplo la transformación logarítmica logró homogenizar la varianza y hacer que la distribución de los errores fuera aproximadamente normal. Esta puede ser una ventaja de transformar el modelo frente a modelar la varianza.
- Como se observa las dos metodologías presentan similitud en el sentido que con ambas se obtiene el mismo diseño D óptimo. La diferencia es que con la primera metodología

se trabaja con el modelo original y se modela la varianza del error. En cambio, con la segunda metodología se trabaja con el modelo transformado y es necesario incorporar el jacobiano de la transformación en la función de verosimilitud de la nueva variable respuesta, aunque la variable regresora x no sufre transformación, por lo que el rango de regresión no se ve afectado.

- Como conclusión general, no puede decirse que una u otra metodología sea mejor, ya que ambos métodos dependen del modelo utilizado, de los datos que se tengan y sobre todo de la estructura de la varianza del error.

5.2. Recomendaciones

- Los diseños D óptimos locales son modelo dependientes, así que lo primero que se recomienda es hacer un muy buen análisis de los datos (si se dispone de ellos) con el fin de ajustar el modelo adecuado. Además, una vez se encuentra el modelo adecuado se debe hacer un buen análisis de los residuales, ya que estos son la clave para identificar la estructura de varianza de los errores.
- Además del modelo, se recomienda tener un buen valor local para el vector de parámetros, ya que el proceso de optimización del funcional de la matriz de información dependerá de dicho valor. Si no se dispone de los datos se deberá usar el conocimiento de un estudio previo o similar.
- Se recomienda el uso del software R, ya que permite implementar los algoritmos necesarios para encontrar diseños D óptimos locales con cada una de las metodologías. En especial las funciones *nlminb* y *optim* para el proceso de optimización y las librerías *nlrwr* (Ritz y Streibig, 2008) y *nlme* (Pinheiro et al., 2012) son muy útiles para el uso de las transformaciones de Box Cox y la estimación por mínimos cuadrados generalizados, respectivamente.
- En este trabajo se encontraron diseños D óptimos locales para un modelo no lineal heterocedástico. Queda como trabajo futuro abordar el problema para otros criterios de optimalidad como el de G optimalidad, aunque ya se ha encontrado ciertos avances pero para modelos lineales. También queda trabajo futuro abordar el problema de encontrar diseños D óptimos bayesianos para modelos no lineales heterocedásticos.

Bibliografía

- [1] Atkinson, A.C.; Cook, R.D. (1995). 'D-Optimum Designs for Heteroscedastic Linear Models' *Journal of the American Statistical Association* Vol. 90, No. 429, 204-212.
- [2] Atkinson, A.C. ; Cook, R.D. (1997). 'Designing for a Response Transformation Parameter' *Journal of the Royal Statistical Society* Vol 59, No. 1, 111-124.
- [3] Atkinson, A. C. (2003). 'Horwitz's rule, transforming both sides and the design of experiments for mechanistic models ' *Applied Statistics* Vol 52, 3, 261–278.
- [4] Atkinson, A. C ; Donev, A, N ; Tobias, R. D. (2007). *Optimum Experimental Designs with SAS*. Oxford Science Publications, New York.
- [5] Bates, D.M; Watts D.G. (1988). *Nonlinear Regression Analysis and its Applications*. John Wiley and Sons, USA.
- [6] Beal, S. L ; Sheiner, L. B. (1988). 'Heteroscedastic Nonlinear Regression ' *Technometrics* Vol. 30, No. 3, 327-338.
- [7] Box, G. E ; Cox, D. R. (1964). 'An Analysis of Transformations ' *Journal of the Royal Statistical Society. Series B (Methodological)* Vol.26, No. 2, 211-252.
- [8] Carrol, R.J; Ruppert, D. (1988). *Transforming and weighting in regression* . Chapman y Hall, New York.
- [9] Casella, G; Berger, R.L. (2002). *Statistical inference*. Duxbury, USA.
- [10] Dette, H; Wong, K. (1999). 'Optimal Designs When the Variance Is a Function of the Mean'. *Biometrics*.Vol 55 (3). 925-929.
- [11] Dette, H; Muller, W.G . (2012). 'Optimal designs for regression models with a constant coefficient of variation'. *AMS*.
- [12] Downing, D; Fedorov, V. V; Leonov,S (2001). 'Extracting Information from the Variance Function: Optimal Design'. *mODa6 Physica*, Heidelberg.
- [13] Fedorov, V. V; Hackl, P. (1997) *Model-oriented design of experiments*. Springer-Verlag, New York. Lecture notes in statistics, Vol 125.

- [14] Huet, S; Bouvier, A; Poursat, M ; Jolivet, E. (2004). *Statistical Tools for Nonlinear Regression: A Practical Guide With S-PLUS and R Examples* . Springer-Verlag, New York.
- [15] Kimura, D.K. (1990). ‘Testing Nonlinear Regression Parameters under Heteroscedastic, Normally Distributed Errors’ *Biometrics* Vol. 46, No. 3, 697-708.
- [16] López-Fidalgo, H; Wong, K. (2002). ‘Design issues for the Michaelis-Menten model’. *J.theo.biol.* Vol 215. 1-11.
- [17] López-Ríos, V; Ramos - Quiroga, R. (2007). ‘Introducción a los diseños óptimos’. *Revista Colombiana de Estadística.* Vol 30 (1). 37-51.
- [18] López-Ríos, V (2008). *Diseños óptimos para discriminación y estimación en modelos no lineales.* Tesis de Doctorado, Centro de investigación en matemáticas (CIMAT), México.
- [19] O’Brien, T. E; Funk, G. M. (2003). ‘A gentle introduction to optimal design for regression models’. *Journal of the American Statistical Association.* Vol 57 (4). 265-267.
- [20] Pinheiro, J ; Bates, D ; DebRoy, S ; Sarkar; D and the R Development Core Team (2012). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-103.
- [21] Ritz, C; Streibig, J.C. (2008) *Nonlinear Regression with R.* Springer, New York.
- [22] Steinberg, D.D ; Hunter, W.G. (1984). ‘Experimental Design: review and comment ’. *Technometrics.* Vol 26 (2). 71-97.
- [23] R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [24] Wong, K. (1992). ‘A unified approach to the construction of minimax designs’. *Biometrika.* Vol 79 (3), 611-9.
- [25] Wong, K. (1995). ‘On the equivalence of D and G-optimal designs in heteroscedastic models’. *Statistics and Probability Letters.* Vol 25, 317-321.

A. Algoritmos

```
## Algoritmo para encontrar el Diseño D óptimo modelando la varianza:
library(xtable)
library(matrixcalc) # Para la traza de una matriz.
eta<-function(t,b1,b2) b1*exp(b2*t)
#1) vector de derivadas con respecto a los parámetros
eta.deriv <- function(t,b1,b2){
  f1 <- exp(b2*t)
  f2 <- b1*t*f1
  return(as.matrix(c(f1,f2),ncol=1))}
u<- function(t,b1,b2,tau,sigma){
  u1<-eta.deriv(t,b1,b2)/(sigma*(eta(t,b1,b2)^tau))
  return(as.matrix(c(u1,rep(0,2)),ncol=1))}
v<- function(t,b1,b2,tau,sigma){
  v1<-(sqrt(2)*tau*eta.deriv(t,b1,b2))/(eta(t,b1,b2))
  v2<-c(sqrt(2)*log(eta(t,b1,b2)),1/(sqrt(2)*sigma^2))
  return(as.matrix(c(v1,v2),ncol=1))}

# 2) Matriz de información(el argumento xi debe ser una matriz)
Mom <-function(xi,b1,b2,tau,sigma){
# convierte la primera fila de la matriz xi en un vector columna
x <- as.matrix(xi[1,]);w <- xi[2,];
# aplicar por filas la funcion u en los puntos de soporte del diseño.
U <- apply(x,1,u,b1,b2,tau,sigma);
# aplicar por filas la funcion v en los puntos de soporte del diseño.
V <- apply(x,1,v,b1,b2,tau,sigma);
W <- diag(w);# matriz diagonal con los pesos del diseño.
# matriz de información del diseño
Mom <- U%*%W%*%t(U) + V%*%W%*%t(V);
return(Mom)}

#funcional de la matriz de información a minimizar
G<-function(xi0,b1,b2,tau,sigma){
  ss<-length(xi0);m<-ss/2
  #xi<-matrix(c(xi0[1:m],xi0[5:ss])/sum(xi0[5:ss]),ncol=m,byrow=T)
```

```

xi<-matrix(c(xi0[1:m],xi0[(m+1):ss]/sum(xi0[(m+1):ss])),ncol=m,byrow=T)
M=Mom(xi,b1,b2,tau,sigma)
VC<- -log(det(M));
return(VC)}

#Derivada direccional:
phi<-function(x,xioptimo,b1,b2,tau,sigma){
  xi1<-matrix(c(x,1),ncol=1) # Diseño que asigna todo el peso en x.
  #p<-length(theta0) # Número de parámetros a estimar.
  p<-4
  matrix.trace(Mom(xi1,b1,b2,tau,sigma)%*%solve(Mom(xioptimo,b1,b2,tau,sigma)))-p }

##### Función que calcula el diseño D óptimo:

Doptimo<-function(xi0,theta0){
b1<-theta0[1];b2<-theta0[2]
tau<-theta0[3]; sigma<-theta0[4]
ss<-length(xi0);m<-ss/2
#función para calcular el mínimo teniendo en cuenta xi0
res2=nlminb(xi0,G,lower=c(rep(1,m),rep(0,m)),
upper=c(rep(12,m),rep(1,m)),b1=b1,b2=b2,tau=tau,sigma=sigma)
soportexicand<-res2$par[1:m]
pesosxicand<-res2$par[(m+1):ss]/sum(res2$par[(m+1):ss])
## Eliminar los puntos de soporte si el peso es cero:
which(pesosxicand!=0) # Posición de los pesos diferentes de cero.
soporte<-soportexicand[which(pesosxicand!=0)]
pesos<-pesosxicand[which(pesosxicand!=0)]
# Obtenga el diseño que minimiza el funcional de la matriz de información
xioptimo<-matrix(c(soporte,pesos),ncol=length(soporte),byrow=T)
#Gráfica de la derivada direccional de phi en xiD (diseño D óptimo )
en la dirección de xi1(diseño que asigna todo el peso en x)
x<-seq(1,12,0.01)
y<-apply(matrix(x),1,phi,xioptimo,b1,b2,tau,sigma)
plot(x,y,"l", xlab="edad",main='Derivada direccional')
abline(h=0,v=xioptimo[1,-c(2:3)],lty=2)
abline(h=0,v=xioptimo[1,],lty=1)
return(xioptimo)}

xi0=c(2,5,7,10,0.25,0.25,0.25,0.25);beta<-round(c(0.9687562,0.2939104),2)
b1<-beta[1]; b2<-beta[2];sigma<-round(0.3733506,2)
tau<-round(1.115644,2),theta0<-c(b1,b2,tau,sigma)

```

```

xiD<-Doptimo(xi0,theta0)

##Algoritmo D optimo modelo lineal transformando la respuesta y
### Modelo lineal: logy=logb1 + b2*x
fx <- function(t){
  f1 <- 1
  f2 <- t
  return(as.matrix(c(f1,f2),ncol=1))}
#nota: b1 = logb1_0:
#media<-function(t,b1,b2) t(fx(t))%*%c(log(b1),b2)
media<-function(t,b1,b2) t(fx(t))%*%c(b1,b2)

#### Caso 1: lambda=0 (Transformación logaritimica)
#Definir los vectores z1 y z2:
z1<- function(x,b1,b2,sigma){
  z11<-fx(x)/sigma
  z12<-0
  z13<--(media(x,b1,b2)^2)/(2*sigma)
  return(as.matrix(c(z11,z12,z13),ncol=1))}

z2<- function(x,b1,b2,sigma){
  z21<-rep(0,2)
  z22<-1/(sqrt(2)*(sigma^2))
  z23<--(sqrt(2))*media(x,b1,b2)
  return(as.matrix(c(z21,z22,z23),ncol=1))}

#Matriz de información (el argumento xi debe ser una matriz)
Mom <-function(xi,b1,b2,sigma){
# convierte la primera fila de la matriz xi en un vector columna
  x <- as.matrix(xi[1,]);w <- xi[2,];
  K <- apply(x,1,z1,b1,b2,sigma);
  L <- apply(x,1,z2,b1,b2,sigma);
  W <- diag(w);
  Mom <- K%*%W%*%t(K) + L%*%W%*%t(L);
  return(Mom)}

#funcional de la matriz de información a minimizar
G<-function(xi0,b1,b2,sigma){
  ss<-length(xi0);m<-ss/2
  #xi<-matrix(c(xi0[1:m],xi0[5:ss]/sum(xi0[5:ss])),ncol=m,byrow=T)
  xi<-matrix(c(xi0[1:m],xi0[(m+1):ss]/sum(xi0[(m+1):ss])),ncol=m,byrow=T)

```

```

M=Mom(xi,b1,b2,sigma)
VC<- -log(det(M));
return(VC)}

# Función de sensibilidad
d<-function(x,xi,b1,b2,sigma)
t(z1(x,b1,b2,sigma))%% solve(Mom(xi,b1,b2,sigma))%%z1(x,b1,b2,sigma)+
  t(z2(x,b1,b2,sigma))%% solve(Mom(xi,b1,b2,sigma))%%z2(x,b1,b2,sigma)

##### Función que calcula el diseño D óptimo:
Doptimo<-function(xi0,theta0){
  b1<-theta0[1];b2<-theta0[2]; sigma<-theta0[3]
  ss<-length(xi0);m<-ss/2
  #función para calcular el mínimo teniendo en cuenta xi0
  res=nlminb(xi0,G,lower=c(rep(1,m),rep(0,m)),
  upper=c(rep(12,m),rep(1,m)),b1=b1,b2=b2,sigma=sigma)
  soportexicand<-res$par[1:m]
  pesosxicand<-res$par[(m+1):ss]/sum(res$par[(m+1):ss])
  ## Eliminar los puntos de soporte si el peso es cero:
  which(pesosxicand!=0) # Posición de los pesos diferentes de cero.
  soporte<-soportexicand[which(pesosxicand!=0)]
  pesos<-pesosxicand[which(pesosxicand!=0)]
  # Obtenga el diseño que minimiza el funcional de la matriz de información
  xioptimo<-matrix(c(soporte,pesos),ncol=length(soporte),byrow=T)

  #Gráfica de la función de sensibilidad en xiD
  (diseño D óptimo ) en la dirección de xi1(diseño que asigna todo el peso en x):
  x<-seq(1,12,0.01)
  y<-apply(matrix(x),1,d,xioptimo,b1,b2,sigma)
  plot(x,y,"l", xlab="edad",main='Función de sensibilidad',las=1)
  #abline(h=0,v=xioptimo[1,-c(2:3)],lty=2)
  abline(h=4,v=xioptimo[1,],lty=2)
  return(xioptimo)}

#Valor local (Ajuste obtenido con la transformación logarítmica)
beta0<-c(0.03,0.26);sigma<-0.57
#beta0<-c(1.03,0.26);sigma<-0.57
theta0<-c(beta0,sigma)
xi0<-c(4,6,9,0.3,0.3,0.3)
xiD<-Doptimo(xi0,theta0)

```