

Curso 2005/06
CIENCIAS Y TECNOLOGÍAS/11
I.S.B.N.: 84-7756-753-0

ROBERTO DORTA GUERRA

**Diseños factoriales óptimos para datos
binarios con dos variables de diseño**

Directores

**ENRIQUE F. GONZÁLEZ DÁVILA
JOSEP GINEBRA MOLINS**



SOPORTES AUDIOVISUALES E INFORMÁTICOS
Serie Tesis Doctorales

*A mis padres,
Roberto y Blanca
y a Beatriz.*

Agradecimientos

Sin duda, el escribir esta tesis me ha reportado una gran satisfacción personal y me ha permitido conocer gente que me ha hecho ver la investigación desde diferentes puntos de vista. Sin ellos, este proyecto no hubiese salido adelante.

En primer lugar quisiera agradecer a Enrique González el haberme permitido trabajar con él y la confianza depositada en mí dirigiendo este trabajo de investigación. Las aportaciones tanto humanas como investigadoras han sido innumerables. Sin su apoyo constante y su perseverancia, este trabajo nunca hubiese llegado a su fin. Muchas gracias.

Quiero agradecer también a Josep Ginebra que me ha brindado su apoyo y su ayuda a lo largo de todo el proceso de realización de la tesis, y por sus aportaciones y dedicación inestimables.

No puedo dejar de agradecer a mis compañeros del Departamento de Estadística, Investigación Operativa y Computación por haberme permitido compartir momentos tanto en mi carrera profesional como personal. En especial tengo que agradecer a Luz Marina, mi antigua compañera de despacho, por apoyarme en los comienzos de este trabajo; a Vicky, Isa Sánchez y Macu por esos ratos de esparcimiento que me hacían desconectar momentáneamente del trabajo de investigación; a Marcos Colebrook, por esas charlas tan interesantes, a veces en serio otras en broma, pero siempre enriquecedoras; a Jose Miguel por esos ratos de humor tan necesarios en los momentos duros; a Antonio Sedeño por su predisposición para ayudarme ante cualquier problema y a Sergio Alonso por ayudarme a dar el último empujón para que este proyecto se hiciera realidad.

Pero, esta tesis no hubiese sido posible sin la estabilidad personal que requiere. En ella han intervenido mi familia, a la que quiero agradecer su comprensión y su apoyo incondicional. Especialmente a mi padres Roberto y Blanca, por creer en todo lo que hago y por su cariño constante.

Por último quiero agradecer a la persona que más ha aguantado esas horas interminables de despacho. A Beatriz, por estar siempre ahí y por creer tanto en mí. Tu apoyo y cariño me han permitido seguir adelante con este proyecto en todo momento.

A todos muchas gracias.

La Laguna, Santa cruz de Tenerife
10 Mayo, 2006

Roberto Dorta Guerra

Índice general

Agradecimientos	v
Índice general	vii
1. Introducción y revisión bibliográfica	1
1.1. Antecedentes y motivación	1
1.2. Optimalidad en diseño de experimentos	3
1.3. Experimentos a dos niveles	5
1.3.1. Experimentos factoriales a dos niveles para modelos lineales normales	7
1.4. Experimentos Factoriales Fraccionarios a dos niveles para modelos lineales normales	10
1.5. Diseños óptimos para modelos de respuesta binaria	12
1.5.1. Diseños localmente óptimos	12
1.5.2. Diseños óptimos bayesianos según Fisher	16
1.6. Descripción de contenidos	17
2. Modelo Lineal Generalizado para datos binarios	21
2.1. Introducción	21
2.2. Modelos lineales generalizados	23
2.3. Matriz de Información y su determinante	25
2.4. Modelos de respuesta binaria	28
2.4.1. Funciones link para datos binarios	29
2.4.2. Modelos de respuesta binaria	33
2.4.3. Matriz de Información	36
2.5. El Criterio de la D-optimalidad para modelos de respuesta binaria . .	37
2.6. Determinante de $I(\beta)$ para experimentos a dos niveles	41
2.7. Conclusiones	43

Experimentos a dos niveles para modelos de respuesta binaria	44
3. Experimentos con un solo factor	47
3.1. Experimentos con un solo factor dependientes de z_0	48
4. Experimentos con dos factores	55
4.1. Diseños 2^2 balanceados D-óptimos	58
4.2. Diseños D-óptimos a dos niveles	61
4.3. Diseños factoriales D-óptimos cuando hay restricciones en los factores.	64
4.4. Conclusiones	71
5. Experimentos con 3 factores	73
5.1. Diseños 2^3 balanceados D-óptimos	75
5.2. Diseños D-óptimos a dos niveles	76
5.3. “ 2^{3-1} por defecto” D-óptimos	80
6. Experimentos a dos niveles en presencia de interacciones	87
6.1. Caso de dos factores	88
6.1.1. Diseño D-óptimo general	89
6.1.2. Diseño D-óptimo centrado en un punto x_0	91
6.2. Caso de tres factores	95
7. Casos Particulares	99
7.1. Experimento factorial 2^2 cuadrado	100
7.2. Experimento cuadrado con ángulo de giro θ	108
7.3. Comparación entre diseños D-óptimos	114
7.3.1. Diseño D-óptimo de referencia	115
7.3.2. D-eficiencia de los diseños D-óptimos cuadrados	118
7.3.3. D-eficiencia de un diseño a dos niveles cuadrado rotado un ángulo θ	119
8. Conclusiones y líneas futuras	123
Apéndices	126
A. Demostración de la Proposición 2.1	129
B. Fracciones mitad de diseños 2^3 en orden estándar	131
C. Demostración del Corolario 7.2	147
Bibliografía	149

Capítulo 1

Introducción y revisión bibliográfica

1.1. Antecedentes y motivación

A lo largo de las últimas décadas, se han llevado a cabo de forma creciente en distintas áreas, estudios donde las variables respuesta toman únicamente dos valores. Así por ejemplo, un circuito integrado manufacturado por una compañía electrónica puede ser defectuoso o no; un paciente en un ensayo clínico para comparar distintos tratamientos, puede o no experimentar alivio en los síntomas asociados a la enfermedad que padece; un insecto puede sobrevivir o morir cuando se le expone a una dosis particular de insecticida. En general, lo más significativo en este tipo de investigaciones, es la proporción de individuos o artículos que, o bien verifican una determinada característica de interés, o bien la rechazan. A este tipo de estudios se les denomina, *problemas de respuesta binaria*.

Un estudio de respuesta binaria tiene dos fases. En la primera, se aplican los distintos tratamientos que están siendo investigados a individuos o artículos; y en la segunda, se mide la proporción de individuos que cumplen una determinada característica. Es evidente que este tipo de respuestas no pueden ser modeladas a través de variables aleatorias normales. Los modelos que se utilizan en estos casos, deben proporcionar las herramientas necesarias que permitan el uso de variables de tipo

discreto y de varianza no constante. Los modelos que abarcan estas características son los *modelos lineales generalizados* que sirven para representar el comportamiento de respuestas a través de distribuciones de la familia exponencial.

Los modelos lineales generalizados para datos de tipo binario han sido estudiados ampliamente en la literatura (McCullagh y Nelder (1989); Myers, Montgomery y Vinning (2002)). Recientemente, los diseños de experimentos para modelos de respuesta binaria han tomado especial importancia y muchos artículos han tratado el tema desde diferentes puntos de vista (ver Sección 1.5).

Por otro lado, desde la perspectiva del diseño de experimentos, hay dos enfoques básicos a la hora de planificar un experimento, el primero, usar diseños factoriales a dos niveles y el segundo, utilizar experimentos que resulten de la maximización de alguna medida de información. Los diseños factoriales presentan algunas ventajas. Por ejemplo, utilizados secuencialmente permiten acercarse al óptimo y, como diseños por bloques, permiten ajustar superficies de respuesta. Además, proporcionan estimaciones de las interacciones entre factores y estimaciones con varianza reducida de los efectos de las variables, siendo relativamente sencillos de construir y analizar. Con respecto a la segunda alternativa, está indicada en aspectos estadísticamente más estructurados, pues conlleva un tratamiento matemático-teórico de mayor dificultad. En general, estos diseños tratan de optimizar procesos. En el caso de procesos donde la variable respuesta de interés sea continua, hay una gran cantidad de referencias relacionadas con diseños óptimos para modelos lineales (Atkinson y Donev (1992); Chaloner y Larntz (1989); Pukelsheim (1993)).

Sin duda nos encontramos con dos aspectos estadísticamente relevantes que hasta ahora no habían sido relacionados: modelos lineales generalizados de respuesta binaria y experimentos a dos niveles. *En este trabajo tratamos de llenar este vacío obteniendo diseños a dos niveles eficientes para modelos de respuesta binaria.* De forma más específica, estamos interesados en, por un lado, obtener los diseños a dos niveles D-óptimos para modelos de respuesta binaria y, por otro, comparar las características de estos diseños con las de los diseños óptimos para modelos lineales

generales.

1.2. Optimalidad en diseño de experimentos

En esta sección, aportamos una visión general de los criterios más comunes de la optimalidad en diseño de experimentos, haciendo hincapié en el criterio de la D-optimalidad. Kiefer y Wolfowitz (1959) y Kiefer (1959), definieron los principios teóricos de los criterios de optimalidad de diseños. Un *diseño*, es una medida de probabilidad que representa la distribución en el espacio experimental de las observaciones que se van a llevar a cabo. Desarrollaron la teoría de la optimalidad en diseño de experimentos e introdujeron los criterios de la D- y E-optimalidad para modelos de regresión lineal, sentando las bases para el desarrollo de otros criterios como el de la A-, F- y G-optimalidad. Muchos de ellos fueron desarrollados para modelos lineales de varianza constante, pero la mayoría, han sido adaptados para utilizarlos en situaciones no lineales y con varianza no constante. La idea básica que subyace en la teoría de diseños óptimos es la de mejorar las estimaciones de los parámetros de interés seleccionando de forma óptima los niveles de las variables de control.

Silvey (1980) y Atkinson y Donev (1992) describen los criterios de optimalidad más utilizados en la práctica. En esta memoria, nos hemos concentrado básicamente en los diseños D-óptimos. Veamos una descripción de este criterio.

De entre todos los criterios el más conocido y utilizado es el de la D-optimalidad, desarrollado por Kiefer y Wolfowitz (1959) y ampliamente tratado por otros autores como Wynn (1972a y b), Atkinson y Donev (1992) y Wu y Hamada (2000) entre otros. Si el vector de parámetros del modelo se denota por β , con $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$, la D-optimalidad sugiere que la elección del diseño debería maximizar la información contenida en β por medio de la minimización de la varianza generalizada de su estimador. En general, el estimador utilizado en modelos no lineales es el estimador de máxima verosimilitud (EMV). Si $\hat{\beta}$ es el (EMV) de β ,

entonces la matriz de varianza asintótica de $\hat{\beta}$ es la inversa de la matriz de información de Fisher (ver Lehmann (1983)). Los criterios de optimalidad tradicionales están fundamentados en esta matriz de información que viene definida según (1.1). Si la dimensión de β es $(k + 1) \times 1$, la matriz de información de Fisher $I(\beta)$ es una matriz de dimensión $(k + 1) \times (k + 1)$ tal que

$$I(\beta, x) = -E \left[\frac{\partial^2 \log(L(\beta, x))}{\partial \beta \partial \beta'} \right], \quad (1.1)$$

donde x es un vector de k variables explicativas y $L(\beta, x)$ es la función de verosimilitud de los datos. El criterio de la D-optimalidad maximiza el determinante de la matriz de información de Fisher. El criterio general de la D-optimalidad viene definido por

$$\max_{X \in D} |I(\beta)|, \quad (1.2)$$

donde X es la *matriz del diseño*, de forma que

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1q} & \dots & x_{kq} \end{pmatrix}, \quad (1.3)$$

y D es el conjunto de todos los posibles diseños. De esta manera, el diseño D-óptimo se obtiene maximizando, únicamente, el determinante de la matriz de información de Fisher. El criterio de la D-optimalidad minimiza el volumen de los elipsoides de confianza en los parámetros del modelo basados en el estimador máximo verosímil (Myers y Montgomery (2002)).

Silvey (1980) expuso muchas de las propiedades de los diseños D-óptimos. Una de las más importantes es la que se deduce del teorema de Carathéorodý, según el cual, para un modelo con $(k + 1)$ parámetros existe un diseño D-óptimo con un número de puntos soporte comprendido entre $(k + 1)$ y $(k + 1)(k + 2)/2$. Así, sólo es necesario considerar diseños con un número finito de puntos soporte.

Para modelos lineales con varianza constante y conocida, la matriz de información de Fisher es $X'X$. esta matriz no depende de los parámetros desconocidos del

modelo β , sino de la localización de los puntos del diseño. De esta forma, el diseño D-óptimo se obtiene e implementa independientemente del valor de los parámetros del modelo.

Para modelos lineales generalizados, salvo en el caso de modelos lineales normales, la matriz de información depende de los parámetros del modelo. En el Capítulo 2 hemos desarrollado de forma general esta matriz obteniendo propiedades relacionadas con la aditividad de su determinante y, de forma más exhaustiva, hemos desarrollado el caso para modelos de respuesta binaria. Debido a la dependencia con los parámetros, los diseños óptimos para modelos lineales generalizados únicamente se pueden definir a partir de información a priori. Si dicha información es puntual, estaremos en el caso de diseños localmente óptimos. Cuando se conoce una distribución a priori de los parámetros del modelo, podemos calcular diseños óptimos Bayesianos. El resto de diseños óptimos, como los minimax o los secuenciales, se derivan de los dos anteriores. En la Sección 1.5 hacemos una breve descripción de estos diseños, haciendo hincapié en los diseños construidos a partir de información puntual que son los que vamos a utilizar en la presente memoria.

1.3. Experimentos a dos niveles

Dentro de los diseños de experimentos existen varios casos de especial importancia en relación a su uso en el campo experimental, y a que constituyen la base de otros diseños de gran valor práctico. El más importante de estos casos es el diseño a dos niveles. Denotamos por 2^k a un diseño de k factores a dos niveles. Una réplica completa de este diseño requiere $2 \times 2 \times \dots \times 2 = 2^k$ observaciones. Estos diseños requieren relativamente pocos experimentos elementales por cada factor, y a pesar de que no permiten explorar exhaustivamente una amplia región del espacio de los factores, pueden indicar tendencias y así determinar una dirección prometedora para futuros experimentos. Además son de gran utilidad en las primeras fases de una investigación, donde suele ser aconsejable estudiar, en un primer intento,

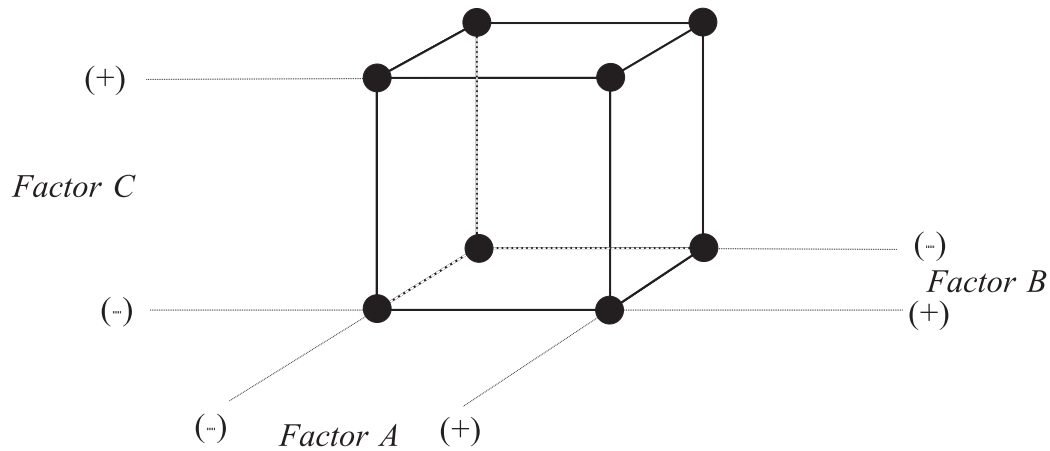


Figura 1.1: Diseño de tres factores a dos niveles (2^3).

un gran número de variables superficialmente en lugar de estudiar intensamente un pequeño número (que puede o no incluir las variables importantes). Por último la interpretación de las observaciones producidas por estos diseños se puede realizar en gran parte a base de sentido común y aritmética elemental. Esto hace que los diseños a dos niveles sean utilizados ampliamente en diversos campos tales como la industria, medicina o biología entre otros, incluso sin ser los más eficientes.

De forma gráfica, entendemos por experimento a dos niveles de k -factores, a aquel que tiene sus puntos soporte en los vértices de un hiper-rectángulo en \mathbb{R}^k , orientado de tal forma que sus lados sean paralelos a los ejes cartesianos y, en el cual, se puedan considerar tanto el caso de asignaciones balanceadas como no balanceadas. Una asignación balanceada asigna el mismo número de observaciones a cada uno de los puntos donde se va a muestrear. En la Figura 1.1 se representa un diseño 2^3 cuyos factores son A , B y C . Los niveles de los factores se encuentran codificados de manera que, si se consideran factores cuantitativos (en esta memoria no trabajaremos con factores cualitativos), un menos representa el nivel bajo y un mas el nivel alto del factor. En el caso de factores cualitativos los niveles se pueden codificar también por signos menos y mas. No importa en este caso cual sea el asociado al mas mientras

la asignación sea coherente. Una forma de representar estos diseños de forma clara y precisa es mediante una matriz donde queden recogidas todas y cada una de las combinaciones de factores de que consta el diseño. Esta matriz es lo que se denomina como *matriz del diseño*.

En modelos lineales normales los diseños a dos niveles son de gran importancia ya que además de las características mencionadas anteriormente, la clase de los diseños *factoriales* a dos niveles, son óptimos, o casi-óptimos bajo un gran número de regiones de experimentación y de medidas de información (ver, por ejemplo, Pukelsheim (1993) o Goel y Ginebra (2003)). Los diseños factoriales a dos niveles son aquellos en los que se investigan todas las posibles combinaciones de los niveles de los factores en cada ensayo completo o réplica del experimento. Esto es, un experimento a dos niveles completo y balanceado. Así, los experimentos factoriales a dos niveles son herramientas de gran utilidad a la hora de identificar las variables explicativas más relevantes, y de explorar la superficie respuesta en dichos modelos (Box y Draper (1987); Myers y Montgomery (2002)).

Por otro lado, hasta el momento, el estudio de los experimentos a dos niveles en modelos lineales generalizados y, más concretamente en modelos de respuesta binaria, no había sido desarrollado. Nuestro interés es obtener las propiedades de estos diseños en modelos de respuesta binaria y compararlas con las propiedades que poseen los diseños factoriales a dos niveles en el caso de modelos lineales normales. En la siguiente sección enumeramos algunas de las propiedades más importantes de los diseños factoriales a dos niveles para modelos lineales normales.

1.3.1. Experimentos factoriales a dos niveles para modelos lineales normales

Fisher comenzó a trabajar en Rothamsted, un centro de investigación agrícola, en 1919 y en un período de 10 años sentó las bases del diseño de experimentos, un área que desarrolló posteriormente su sucesor Yates. La principal aportación de Fisher,

entre otras muchas, fue la de definir los diseños factoriales. Un *diseño factorial* es aquel en el que se investigan todas las posibles combinaciones de los niveles de los factores en cada ensayo completo o réplica del experimento. Por ejemplo, si consideramos un problema donde están involucrados dos factores, A , con a niveles y B , con b niveles, entonces cada ensayo completo o réplica del experimento contiene todas las $a \times b$ combinaciones de los tratamientos.

Las ventajas de los diseños factoriales se pueden resumir en que son más eficientes que los experimentos de un factor a la vez y permiten estimar los efectos de un factor en diversos niveles de los otros factores, aportando propiedades que son válidas sobre toda la extensión de las condiciones experimentales.

Supongamos un estudio donde la respuesta de interés puede ser modelada convenientemente a través de un modelo de regresión lineal normal de primer orden o con interacciones. Para determinar de forma óptima los parámetros del modelo normalmente hay dos vías. La primera, elegir un diseño factorial a dos niveles y la segunda, obtener un diseño por maximización de una medida de información. En el caso de modelos lineales normales estas dos líneas de acción acostumbran a tener como resultado el diseño factorial.

En general, lo que se desea es determinar qué variables o factores influyen en una variable respuesta de interés. Además, tratamos de obtener un modelo matemático que relacione los factores con la variable en estudio. En base a esto, siempre podemos encontrar, como una buena aproximación, un modelo lineal de primer orden, para una región de interés predeterminada, que ajuste bien los datos observados. En este proceso surgen algunas cuestiones tales como dónde observar y cuántas observaciones tomar para que el modelo que obtengamos sea el óptimo. En general, tal y como hemos comentado en la Sección 1.3, los diseños factoriales a dos niveles son los más apropiados para definir estos modelos. La propiedad más importante de estos experimentos es la ortogonalidad (Box y Hunter (1961a y b) y ; Myers

y Montgomery (2002) y Montgomery (2001)). Se dice que un diseño es *ortogonal*, cuando las columnas de la matriz del diseño X , son ortogonales dos a dos y verifican que los elementos fuera de la diagonal de su matriz de información ($X'X$), son cero. A partir de la propiedad de ortogonalidad y de la propia estructura de los diseños factoriales, podemos sacar las siguientes conclusiones o propiedades que verifican los diseños factoriales cuando tratan de ajustar modelos lineales normales de primer orden.

1. **A mayor rango de variación de los factores involucrados, mayor es el determinante y, por tanto, el diseño aporta más información. Esto hace que las regiones experimentales para un modelo lineal normal deban estar acotados.**
 2. **El valor del determinante no depende de donde esté centrado el experimento, ni de su orientación con respecto a las líneas de contorno de la superficie respuesta.**
 3. **Los experimentos factoriales balanceados siguen siendo óptimos cuando se añaden términos de interacción al modelo.**
 4. **Considerando réplicas, asignaciones balanceadas que distribuyen de forma equitativa el número de réplicas entre todas las combinaciones de factores y cuyas matrices de diseño estandarizadas son ortogonales, tienen siempre un valor del determinante mayor que asignaciones no balanceadas con el mismo número total de observaciones.**
 5. **El determinante de un experimento donde se modifican los factores uno a uno, es menor que el determinante de un factorial completo.**
-

Como consecuencia, al planificar un experimento factorial a dos niveles para el conjunto de modelos lineales normales de primer orden, lo verdaderamente trascendente es el rango de variación de los factores involucrados y la ortogonalidad de la matriz de diseño. En presencia de interacciones, solamente son significativos varios experimentos factoriales fraccionales, evaluados en términos de qué interacciones se pueden estimar y cuáles están confundidas.

1.4. Experimentos Factoriales Fraccionarios a dos niveles para modelos lineales normales

A medida que el número de factores en un diseño factorial 2^k aumenta, el número de ensayos necesarios para obtener una réplica completa sobrepasa rápidamente los recursos de la mayoría de los experimentadores. Nótese que, por ejemplo, un diseño factorial 2^6 requiere 64 ensayos. Para evitar este sobre-costos en la experimentación aparecen los diseños factoriales fraccionales que reducen considerablemente el número de ensayos a realizar (ver Box y Hunter (1961a y b)).

Si el experimentador puede suponer razonablemente que algunas interacciones de orden superior son despreciables, la información sobre los efectos principales y las interacciones de menor orden pueden obtenerse realizando sólo una fracción del experimento factorial completo. Estos diseños se denominan *factoriales fraccionarios* y se encuentran entre los tipos más ampliamente usados para la detección y solución de problemas en el diseño de productos y procesos. Un uso importante de los factoriales fraccionarios se da en los experimentos de *escrutinio*, también llamados de tamizado o cribado. Éstos son experimentos en los cuales se consideran muchos factores con el fin de identificar aquellos que tienen efectos importantes, de haberlos. Los experimentos de escrutinio suelen realizarse en las primeras fases de un proyecto cuando es probable que muchos de los factores inicialmente considerados tengan poco o ningún efecto sobre la respuesta. Los factores que se identifican como

importantes se investigan entonces con mayor detalle en experimentos posteriores.

Algunos ejemplos de diseños fraccionales son: para dos factores los diseños Plackett and Burman (1946) que consisten en diseños altamente fraccionalizados con el máximo número de efectos principales que tienen efecto en la variable respuesta con el menor número de observaciones experimentales; y para tres factores, el equivalente al de Plackett and Burman es el denominado diseño Box-Behnken (Box y Behnken (1960)).

Las propiedades más importantes relacionadas con los diseños factoriales fraccionales a dos niveles para modelos lineales normales podemos resumirlas en tres:

- 1. Cuando se fraccionan experimentos factoriales, las asignaciones balanceadas para factores simétricos, fijando el mismo número de combinaciones experimentales a cada uno de sus niveles y cuyas matrices de diseño son ortogonales, son siempre al menos tan buenas como asignaciones cuyas matrices de diseño son no ortogonales con el mismo número de observaciones total.**
- 2. Dos fracciones complementarias de un factorial completo a dos niveles son estadísticamente equivalentes.**
- 3. Considerando asignaciones balanceadas, el valor del determinante de cualquier fracción de un factorial a dos niveles siempre es menor que el valor del determinante de un experimento factorial completo con el mismo número de observaciones total.**

En esta tesis, aparte de hacer un estudio de diseño de experimentos para los diseños a dos niveles, también se han estudiado los diseños factoriales fraccionarios.

Veremos que las propiedades para estos diseños en los modelos lineales normales no se verifican en el caso de modelos de datos binarios.

1.5. Diseños óptimos para modelos de respuesta binaria

Los diseños óptimos para modelos lineales generalizados, y en particular para modelos de respuesta binaria, dependen de los parámetros desconocidos del modelo. Esto hace que los diseños óptimos para modelos de respuesta binaria sean más difíciles de obtener si los comparamos con los diseños óptimos para modelos lineales clásicos donde la matriz de información no depende de los parámetros. En esta sección introduciremos algunos métodos propuestos en la literatura, que tratan de solventar el problema de la dependencia de los parámetros. La mayor parte de la investigación que hasta el momento se ha desarrollado en diseños de experimentos para modelos de respuesta binaria se limita al modelo de regresión logístico (link canónico), y casi no hay referencias de diseños óptimos para modelos con links no canónicos, (probit, doble exponencial, doble recíproco y complementario log-log).

1.5.1. Diseños localmente óptimos

El primer método está basado en considerar buenas estimaciones de los parámetros al inicio del experimento. Los diseños óptimos se calculan a partir de la maximización de una función asociada a uno de los criterios de optimalidad vistos en la Sección 1.2, evaluada en las estimaciones de los parámetros. A este tipo de diseño se le denomina *diseño óptimo local*, término introducido por Chernoff (1953). Las estimaciones de los parámetros pueden ser obtenidas a partir de un experimento previo, o de un experimento piloto llevado a cabo expresamente con este propósito, o simplemente a partir de la experiencia o intuición del investigador. A estas aproximaciones las denominaremos estimaciones iniciales o aproximaciones iniciales sin importar

cómo se hayan obtenido. Muchos autores han considerado este método para obtener los diseños D-óptimos en el caso de modelos de respuesta binaria (Finney (1971) y (1978); Meeker y Hahn (1977); Maxim, Hendrickson y Cullen (1977)).

A continuación, presentamos un breve resumen de los diseños localmente óptimos para el modelo de regresión logístico. La mayoría de los trabajos consideran modelos de regresión logístico de primer orden con una variable de diseño definido por

$$y \sim \text{Bernoulli}(p(x; \beta)) \quad \text{con} \quad p(x; \beta) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x))} \quad (1.4)$$

donde, y es la respuesta binaria (0 ó 1) y p es la probabilidad de tomar el valor 1 que depende de la variable independiente x . Para este modelo, Kalish y Rosenberg (1978) desarrollaron un diseño D-óptimo de dos puntos simétricos respecto al $ED_{.50}$ (ED es la abreviatura de “Effective dose” (dosis efectiva), donde ED_p representa el valor de x que tiene probabilidad p cuando la respuesta es 1). El diseño D-óptimo de dos puntos distribuye la mitad de las observaciones en el $ED_{.176}$ y la otra mitad en el $ED_{.824}$. Para implementar el diseño D-óptimo, debemos conocer los valores de $ED_{.176}$ y $ED_{.824}$, o, de forma equivalente, conocer los valores para β_0 y β_1 . En la práctica, los parámetros son desconocidos y es necesario dar estimaciones iniciales. Nos encontramos en el caso de un diseño localmente D-óptimo.

Abdelasit y Plackett (1983) llevaron a cabo una revisión de los resultados que se habían obtenido hasta ese momento y aportaron un estudio de la robustez del criterio de la D-optimalidad, entre otros, a estimaciones pobres de los parámetros del modelo (1.4). Para ello, consideraron un diseño de dos puntos, el obtenido por Kalish y Rosenberg (1978), cuando el número de observaciones total era par, y otro de tres puntos cuando era impar. El diseño localmente D-óptimo para el caso impar situaba proporcionalmente las observaciones sobre los ED 's del 86.4%, 50.0% y 13.6%. Este diseño tenía un determinante de la matriz de información un 13% menor que el obtenido para dos puntos. Minkin (1987) demuestra que el diseño de dos puntos se puede generalizar al caso impar. De esta forma, si denotamos por n al número total de observaciones, en el caso impar, el diseño localmente D-óptimo sitúa $(n-1)/2$ observaciones sobre uno de los puntos y $(n+1)/2$ observaciones sobre

el otro.

Ford, Torsney y Wu (1992) demostraron que para ciertas clases de modelos lineales generalizados el problema de determinar los diseños localmente D-óptimos se podía reducir a una forma canónica. Aportaron los diseños localmente D-óptimos para los modelos de respuesta binaria con una variable regresora y para varios link. En el caso del modelo (1.4) el diseño localmente D-óptimo sitúa sus puntos sobre el $ED_{.176}$ y el $ED_{.864}$, demostrando así que éste método funcionaba correctamente, al coincidir con el obtenido por Kalish y Rosenberg (1978), Abdelasit y Plackett (1983) y Minkin (1987). Además, este trabajo es la base para obtener diseños localmente D-óptimos y sirvió para construir diseños en casos más complejos donde había más de una variable regresora.

Por otro lado, en un gran número de aplicaciones, la estimación de un determinado ED , en particular del $ED_{.50}$, es fundamental. Por ejemplo, en ensayos biológicos, es importante conocer la concentración de un medicamento que produzca una respuesta del 50 % de individuos (dosis mediana efectiva). Así, se han sugerido varios procedimientos para construir diseños óptimos que estimen el $ED_{.50}$. El primero consiste en minimizar la varianza asintótica del estimador del $ED_{.50}$, el segundo, está basado en el criterio de la F-optimalidad, mediante el cual se minimiza la longitud del intervalo Fieller en el $ED_{.50}$, y el tercero, minimiza la longitud del intervalo de confianza basado en la verosimilitud. Algunos autores que trataron estos problemas fueron, Kalish (1990), Finney (1971), Abdelasit y Plackett (1983), Sitter y Wu (1993), Williams (1986) Minkin y Kundhal (1999) y Myers, Myers y Carter (1994), entre otros.

Cuando se consideran dos variables regresoras, el problema se complica. Sitter y Torsney (1995a y b) estudiaron los diseños D-óptimos para modelos de respuesta binaria de primer orden con dos factores, considerando links canónicos y no canónicos. Demostraron que para regiones experimentales no acotadas, el determinante de la matriz de información de Fisher era no acotado. Para este caso sería imposible obtener diseños D-óptimos sin considerar regiones experimentales finitas.

Propusieron espacios acotados en forma de paralelogramo y orientados con los ED . Demostraron que cualquier paralelogramo acotado por los $ED_{.227}$ y $ED_{.773}$ era un diseño D-óptimo para el modelo logístico. Jia y Myers (2001) establecen que el criterio de la D-optimalidad es invariante a la localización y al ángulo de este paralelogramo. Sin embargo, al aumentar el diseño, debido al aumento de los rangos de variación de los factores, el determinante aumenta y, por lo tanto, el diseño es más eficiente. Nótese que esta propiedad es común para los diseños D-óptimos en el caso de tener dos variables regresoras tanto para modelos de respuesta binaria como para modelos lineales normales.

Como podemos observar, los diseños localmente D-óptimos para modelos de respuesta binaria, se calculan de forma general y, en ningún caso, se hace referencia a experimentos a dos niveles. Esto hizo que nos preguntáramos cuáles serían las características de estos diseños que aún cuando no fueran óptimos generales, sí son interesantes por las propiedades práctico-teóricas que presentaban.

Intuitivamente, podríamos esperar que los diseños localmente óptimos fuesen relativamente eficientes si las predicciones están próximas a los verdaderos valores de los parámetros. Sin embargo, se ha demostrado repetidas veces, que los diseños localmente óptimos no son eficientes cuando se consideran estimaciones iniciales pobres. Básicamente este es el problema con el que nos encontramos cuando trabajamos con experimentos no lineales. Ahora bien, aunque no siempre se pueden obtener buenas estimaciones de los parámetros, los diseños localmente óptimos son de gran importancia por dos razones: (1) existen algunas situaciones donde los investigadores poseen estimaciones fidedignas de los parámetros; (2) más importante, los diseños localmente óptimos aportan un punto de referencia útil a partir del cual somos capaces de calibrar cualquier otro diseño. Uno de los inconvenientes que presenta esta opción, está relacionada con la robustez de los diseños obtenidos. Aún así, y sabiendo que nuestro trabajo está basado en determinar las características de los experimentos a dos niveles para los modelos de respuesta binaria, el problema de la robustez es secundario, pues tiene connotaciones que se acercan más al análisis del

experimento que al estudio de las propiedades del diseño en sí mismo. Por todo ello, hemos considerado los diseños localmente D-óptimos a lo largo de esta memoria.

A continuación damos una breve descripción del método Bayesiano para solventar el problema de la dependencia de los parámetros. Aún cuando dicho método no se utilizará en este trabajo, es interesante mencionarlo como una línea de acción futura a la hora de determinar las propiedades de los diseños a dos niveles para modelos de respuesta binaria desde todos los puntos de vista.

1.5.2. Diseños óptimos bayesianos según Fisher

Un procedimiento alternativo para obtener diseños óptimos de modelos binarios, considera distribuciones a priori de los parámetros del modelo. Un diseño óptimo Bayesiano es aquel que maximiza la media de una función asociada a un criterio de optimalidad, sobre una distribución a priori de los parámetros. Por ejemplo, si denotamos la distribución a priori $\pi(\beta)$, el criterio Bayesiano D-óptimo viene dado por

$$\max_{X \in D} \int |I(X, \beta)| \pi(\beta) d\beta \quad (1.5)$$

donde D es el conjunto de los diseños posibles. El diseño localmente óptimo puede ser considerado como un caso particular de un diseño óptimo Bayesiano, para el cual, la distribución a priori degenera en un punto. Dicho punto no son mas que las estimaciones iniciales de los parámetros.

Para el modelo de regresión logística, Charloner y Larntz (1989) desarrollaron formalmente los diseños óptimos de tipo Bayesiano. Tanto el criterio de la D-optimalidad Bayesiana como el de optimalidad Bayesiana para la estimación de un ED en particular, están fundamentados en las mismas premisas. Charloner y Larntz obtuvieron diseños óptimos Bayesianos usando distribuciones a priori uniformes e independientes para los parámetros. Además, llevaron a cabo un estudio comparativo entre los diseños localmente óptimos y los diseños óptimos Bayesianos, calculando su eficiencia. En particular, demostraron que los diseños Bayesianos óptimos eran

más robustos que los diseños localmente óptimos, cuando se toman estimaciones iniciales de los parámetros poco precisas.

Básicamente, un diseño óptimo Bayesiano optimiza la media pesada de una función para un criterio de optimalidad, sobre una distribución a priori. Estos diseños son más robustos que los diseños localmente óptimos, que maximizan la función del criterio de optimalidad elegido para una estimación de los parámetros. No obstante, los diseños Bayesianos presentan algunas desventajas en relación a los diseños localmente óptimos. Por un lado, en este tipo de diseños, los investigadores trasladan la incertidumbre de las estimaciones de los parámetros, a una función de densidad de probabilidad sobre ellos. En algunos casos, este cambio es criticado por no ser del todo objetivo. Otro inconveniente que presentan estos diseños es el esfuerzo computacional necesario para construirlos, debido principalmente, al factor de integración asociado al método. En general, salvo que se desee obtener un diseño óptimo robusto, se suelen utilizar los diseños localmente óptimos, tal y como hemos comprobado en la literatura.

1.6. Descripción de contenidos

Como ya hemos comentado, el cálculo de los diseños a dos niveles D-óptimos para modelos de respuesta binaria depende de los parámetros desconocidos del modelo. Bajo esta premisa, y considerando estimaciones puntuales de los parámetros, hemos desarrollado el estudio de dichos diseños y de sus propiedades. El problema en sí es bastante amplio ya que hay muchos casos que tratar y considerar. Nos hemos basado en aquellos diseños y modelos más comunes. Nótese que estamos considerando dos campos de acción. Por un lado, los modelos de respuesta binaria y por otro, los diseños a dos niveles. Cada uno de ellos presenta diferentes alternativas. Por ejemplo, en el caso del tipo de diseño a tratar, pueden surgir preguntas como cuántos factores tomar, o si hay o no restricciones sobre los factores considerados. Cada capítulo considera un tipo de diseño que por su interés práctico creemos que debía ser tenido

en cuenta. A continuación exponemos de forma pormenorizada las cuestiones que se han tratado en cada uno de los capítulos de la presente memoria.

En el Capítulo 2 se presentan los modelos de respuesta binaria desde la perspectiva de modelos lineales generalizados. Hemos obtenido la expresión general para la matriz de información de Fisher y el determinante de la misma en el caso de modelos lineales generalizados, herramientas que desarrollaremos, en este mismo capítulo, para el caso particular de respuesta binaria. Esto servirá para que en los capítulos posteriores podamos calcular los diseños D-óptimos para diseños a dos niveles. Además se presentan los modelos de respuesta binaria más comunes, como son el logit y probit, y se nombran algunos otros, que aun cuando no suelen ser utilizados en la práctica, en casos puntuales pueden ser de gran necesidad.

En el Capítulo 3 se aportan los resultados computacionales y propiedades de los diseños localmente D-óptimos a dos niveles con un factor, cuando se consideran varias funciones link y para un predictor lineal de primer orden. Para ello previamente se obtuvieron las expresiones de la matriz de información y su determinante. Además se presenta un ejemplo donde se establecen los pasos a seguir para implementar uno de estos diseños .

En el Capítulo 4 obtenemos los diseños a dos niveles localmente D-óptimos centrados en un determinado punto de la región experimental, en el caso de dos factores. Aquí el estudio fue más exhaustivo puesto que consideramos tanto el caso balanceado como no balanceado. También se obtuvieron los diseños D-óptimos para el caso de regiones acotadas. Nótese que en determinadas circunstancias los investigadores tienen restricciones sobre los factores en uso. Esto hizo que planteáramos este problema, y en concreto en los diseños con dos factores puesto que son los que habitualmente se utilizan en las primeras fases del desarrollo de un problema donde intervienen varios factores. Finalmente se obtuvieron los diseños localmente D-óptimos generales para cada uno de los link definidos en el Capítulo 2.

El caso con tres factores se trata en el Capítulo 5 donde aportamos la matriz de información y el determinante así como los diseños localmente D-óptimos para

modelos de respuesta binaria tanto cuando se centran en un punto de la región experimental como para el caso general. Además se realiza un estudio de los diseños factoriales fraccionales, y se comprueba que dichos diseños no son D-óptimos dentro de la clase de diseños de cuatro puntos.

El Capítulo 6 añadimos interacciones al predictor lineal del modelo de respuesta binaria cuando se toman dos y tres factores. En este caso el cálculo se complicaba debido principalmente a problemas computacionales al aumentar el número de parámetros y con ello el número de variables que debemos optimizar. Además los lugares geométricos de las líneas ED cambian y algunas propiedades que se dan para los diseños D-óptimos en los capítulos anteriores ya no son válidas, lo que implica que las propiedades en este caso no se conserven como ocurría en los modelos lineales normales.

En el Capítulo 7 hemos considerado el caso de dos factores donde el rango para el diseño en los dos factores coincide. Este capítulo está íntimamente relacionado con los diseños D-óptimos para el modelo de respuesta binaria con dos variables de diseño del Capítulo 4. Las aportaciones se basan en el estudio de los diseños en base a los parámetros del modelo y al estudio del comportamiento del diseño cuando se rota un determinado ángulo con el fin de compararlos con los diseños D-óptimos generales obtenidos por Sitter y Torsney.

En el Capítulo 8 se recogen de forma resumida las conclusiones más importantes de este trabajo y se plantean las líneas futuras de investigación.

Por último aparecen tres apéndices. En dos de ellos se recogen las demostraciones de los resultados más importantes de la tesis. En el tercero se analizan las fracciones mitad de los diseños de dos factores a dos niveles D-óptimos para modelos de respuesta binaria con predictor lineal sin interacciones.

Capítulo 2

Modelo Lineal Generalizado para datos binarios

2.1. Introducción

La mayor parte de los modelos usados para describir fenómenos y procesos en disciplinas tan diversas como la socio-economía, las ciencias naturales, la tecnología y la industria, están fundamentados en la teoría de modelos lineales clásicos. En general, el desarrollo teórico de estos modelos está basado en tres premisas. La primera, según la cual, la parte aleatoria del modelo, el error aleatorio, se distribuye normalmente. La segunda, establece que las varianzas del error son constantes e independientes de la media, y la tercera, apunta la aditividad de la parte sistemática del modelo. Esto es, normalidad, homocedasticidad y linealidad, tres conceptos bien definidos dentro del entorno estadístico. Cuando una de estas tres propiedades falla, los modelos lineales clásicos dejan de tener sentido y se han de considerar alternativas tales como la transformación de la variable respuesta. Mediante estas transformaciones se resuelven los problemas derivados de la falta de normalidad de los errores o la heterocedasticidad. Hamada y Nelder (1997) enumeran los problemas que se plantean a la hora de elegir estas transformaciones y exponen la necesidad de optar por otros procedimientos para poder modelar variables que no cumplan alguna de las tres propiedades anteriormente expuestas.

Aparte de los problemas descritos en el párrafo anterior, cuando se usan modelos lineales clásicos no se pueden modelizar variables respuesta, por ejemplo, de tipo binario o de conteo, esto es, variables que siguen una distribución binomial o de Poisson respectivamente. Con el fin de modelar este tipo de variables, Nelder y Wedderburn (1972) desarrollan la teoría de modelos lineales generalizados. Estos modelos generalizan, como su propio nombre indica, el concepto de modelo lineal, englobando variables que siguen distribuciones tan dispares como la Normal, Poisson o binomial, pero que pertenecen todas a la misma familia, tal y como apuntamos en la Sección 2.2 del presente capítulo. Hamada y Nelder (1997) muestran cómo los problemas asociados a variables que no verifican las propiedades de los modelos lineales clásicos, desaparecen cuando los datos se ajustan por modelos lineales generalizados. A lo largo de los años, los modelos lineales generalizados se han aplicado en diferentes áreas como la medicina, biología, o la industria, siendo de vital importancia a la hora de obtener modelos que controlan una variable que por su naturaleza no puede ser modelada con los modelos lineales clásicos. Algunos ejemplos los podemos encontrar en McCullagh y Nelder (1989), Aitkin, Anderson, Francis, y Hinde (1989), Dobson (1990), Francis, Green y Payne (1993), Myers y Montgomery (1997), Lee y Nelder (1998) y Lewis, Montgomery y Myers (2001a).

Los modelos de respuesta binaria son un caso particular de los modelos lineales generalizados. En esta memoria, nos hemos concentrado en este tipo de modelos, con el fin de obtener los experimentos a dos niveles D-óptimos asociados al mismo. En particular, estudiamos las propiedades de los experimentos D-óptimos a dos niveles para los modelos de tipo binario. En este capítulo introducimos los conceptos básicos de los modelos lineales generalizados, particularizándolos al caso de modelos de respuesta binaria.

Uno de los conceptos que más utilizaremos a lo largo de esta disertación es el de matriz de información, por su importancia a la hora de evaluar la información contenida en un diseño. En las Secciones 2.3 y 2.4.3 definimos las matrices de Información de Fisher tanto para los modelos lineales generalizados, como para el caso

particular de los modelos de respuesta binaria.

2.2. Modelos lineales generalizados

Consideremos que observamos un conjunto de respuestas, y_{ij} , con $j = 1, \dots, n_i$, bajo un conjunto de q condiciones experimentales distintas, cada una caracterizada por los valores de $x_i = (x_{1i}, \dots, x_{ki})$, para $i = 1, \dots, q$. En modelos lineales generalizados, las distribuciones de $y_{ij}|x_i$ se asumen independientes y de la familia exponencial. La función de densidad de probabilidad para una variable respuesta cuya distribución pertenece a la familia exponencial, viene dada por

$$f(y_i|x_i; \theta, \psi) = \exp \left\{ \frac{y_i \theta - b(\theta)}{a(\psi)} + c(y_i, \psi) \right\} \quad (2.1)$$

donde $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$ son funciones conocidas. El parámetro θ es el parámetro de localización y ψ el de dispersion. En la tabla 2.2 se definen las funciones $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$ para algunas distribuciones de la familia exponencial. El modelo lineal generalizado viene definido por

$$E[y_{ij}|x_i] = F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}) = F(z_i), \quad (2.2)$$

donde $F(\cdot)$ es una función monótona conocida que satisface las propiedades de una función de distribución acumulativa, y donde $z_i = F^{-1}(E[y_{ij}|x_i])$ se denomina *predictor lineal*, que podría ser un polinomio de un conjunto de p variables regresoras. Nótese que el predictor lineal es de tipo general. Puede contener interacciones, términos cuadráticos, cúbicos, etc. Veamos un ejemplo donde están involucradas como variables regresoras la *altura* y el *peso* de un conjunto de individuos de forma que el predictor lineal sea

$$z_i = \beta_0 + \beta_1 \text{ altura}_i + \beta_2 \text{ peso}_i + \beta_3 \text{ altura}_i \text{ peso}_i$$

Si siguiendo la notación (2.2) tenemos que $p = 2$ y $k = 3$ con $x_{1i} = \text{altura}_i$, $x_{2i} = \text{peso}_i$ y $x_{3i} = \text{altura}_i \text{ peso}_i$ como variables regresoras. Así, el predictor lineal quedaría de

Distribución	θ	ψ	$a(\psi)$	$b(\theta)$	$c(y, \psi)$
Normal	μ	σ^2	ψ	$\mu^2/2$	$-\frac{1}{2} \left(\frac{y^2}{\psi} + \ln(2\pi\psi) \right)$
Poisson	$\ln \mu$	1.0	1.0	e^θ	$-\ln y!$
Binomial	$\ln \left(\frac{p}{(1-p)} \right)$	1.0	1.0	$n \ln(1 + e^\theta)$	$\ln \binom{n}{y}$

Tabla 2.1: Características de algunas distribuciones de la familia exponencial

la forma

$$z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$$

De esta manera hemos conseguido simplificar de forma significativa la notación del problema.

La varianza de $y_{ij}|x_i$, es igual a $\psi V[E[y_{ij}|x_i]]$, donde $V[\cdot]$ es una función de la varianza que depende de la distribución.

A la función $F^{-1}(\cdot)$ que relaciona el predictor lineal y el valor medio de la variable respuesta es lo que se conoce como función link. El término link está relacionado con el hecho de que esta función es el “enlace” que hace que el valor medio de la variable respuesta se exprese como una función lineal. De hecho, para el modelo de regresión lineal múltiple

$$\mu_i = z_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}, \quad i = 1, \dots, q$$

el link es la función identidad. La elección apropiada de la función link en un modelo lineal generalizado es muy importante ya que está íntimamente relacionada con la idoneidad del modelo para ajustar una serie de datos. Nótese que en modelos lineales clásicos, únicamente teníamos en cuenta el predictor lineal. En modelos lineales generalizados tenemos que considerar tanto el predictor lineal como la función link. La función link con mejores propiedades matemáticas es el link canónico que es el que hace que $z_i = \theta_i$. En el caso de modelos de datos binarios hay varias funciones link que se suelen utilizar en la práctica. Las más importantes están recogidas en la Sección 2.4.1, entre ellas el link canónico.

Una vez decidido el tipo de función link y el predictor lineal que se va a utilizar para describir el comportamiento de la variable respuesta en estudio, debemos estimar los parámetros del modelo. Es aquí donde se aplica la optimalidad de diseños. Así, tratamos de calcular cuál es la condición experimental idónea que debemos reproducir para obtener la estimación de los parámetros de forma óptima.

La matriz de información de Fisher describe la información sobre los parámetros desconocidos de un modelo, contenida en un experimento. En el caso de modelos lineales clásicos dicha matriz se simplificaba enormemente por las características propias de estos modelos mientras que en el caso de modelos lineales generalizados su estructura se complica. En la siguiente sección se obtiene la matriz de Información de Fisher para el modelo lineal generalizado y de forma particular para el modelo de respuesta binaria. Además se calcula su determinante y se estudian sus propiedades.

2.3. Matriz de Información y su determinante

La matriz de información de Fisher para el experimento lineal generalizado descrito en la sección anterior, es una matriz de dimensión $(k + 1) \times (k + 1)$ tal que

$$I(\beta) = \sum_{i=1}^q \begin{pmatrix} n_i h(z_i) & n_i h(z_i) x_{1i} & \dots & n_i h(z_i) x_{ki} \\ n_i h(z_i) x_{1i} & n_i h(z_i) x_{1i}^2 & \dots & n_i h(z_i) x_{1i} x_{ki} \\ \dots & \dots & \dots & \dots \\ n_i h(z_i) x_{ki} & n_i h(z_i) x_{1i} x_{ki} & \dots & n_i h(z_i) x_{ki}^2 \end{pmatrix}, \quad (2.3)$$

donde $h(z_i) = F'(z_i)^2 / (\psi V[F(z_i)])$, con $F'(z_i)$ la derivada de la función $F(\cdot)$ respecto del predictor lineal z_i . La inversa de $I(\beta)$ es la matriz de varianza-covarianza asintótica del estimador máximo verosímil de β . Esta matriz determina el tamaño y forma de las regiones de confianza aproximadas para β , basadas en estas estimaciones. Sin embargo, no podemos comparar experimentos únicamente basándonos en $I(\beta)$, ya que las matrices no se pueden ordenar y por ello $I(\beta)$ no genera una ordenación total en el espacio de los experimentos. Debemos recurrir a criterios de valores reales basados en $I(\beta)$ que están relacionados con la forma y el tamaño de

dichas regiones. Así, cuando la finalidad de un experimento está basada en estimar los parámetros β de forma individual, o cualquier función real de ellos, se trata de elegir el valor de q y los (x_i, n_i) que minimizan la varianza asintótica de los estimadores máximo verosímiles para dicha componente, ver por ejemplo, Wu (1988) o Mathew y Sinha (2001). Por otro lado, cuando la finalidad es estimar de forma conjunta todas las componentes de β , la forma de hacerlo es maximizando el determinante de $I(\beta)$, y por consiguiente, minimizando el volumen de los elipsoides de confianza aproximados para β , obteniendo los denominados diseños D-óptimos. Otros criterios de optimalidad estudian la traza o el mayor autovalor de $I(\beta)$, aunque el determinante está considerado como el criterio por defecto cuando uno trata de determinar las propiedades de los experimentos de forma más detallada. Para modelos lineales generalizados, el determinante de (2.3) es

$$\det(I(\beta)) = \det \left(\sum_{i=1}^q n_i h(z_i) (1, x_{1i}, \dots, x_{ki})' (1, x_{1i}, \dots, x_{ki}) \right) \quad (2.4)$$

Una de las propiedades más importantes que aportamos en esta memoria, está relacionada con la aditividad del determinante de la matriz de información de Fisher de una secuencia de experimentos independientes. Sea X la matriz del diseño de dimensión $q \times (k + 1)$ cuya i -ésima fila es $(1, x_{1i}, \dots, x_{ki})$,

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1q} & \dots & x_{kq} \end{pmatrix}, \quad (2.5)$$

y denotemos por $X_{i_1, \dots, i_{k+1}}$ a una sub-matriz de X de dimensión $(k + 1) \times (k + 1)$ formada por sus filas i_1, i_2, \dots, i_{k+1} . El determinante de aquellos experimentos cuya matriz X tenga un rango inferior a $k + 1$, es igual a 0 y por lo tanto no se pueden estimar las componentes de β . Esto se da en casos en los que, por ejemplo, tenemos más parámetros que datos. Así, debemos de considerar experimentos cuyo rango de X es mayor o igual a $k + 1$.

La matriz de información de una secuencia de experimentos independientes es la suma de sus matrices de información, pero el determinante de $I(\beta)$ para dicha secuencia no es la suma de los determinantes de los experimentos individuales. Sin embargo, el siguiente resultado, que aparece para modelos lineales normales en Fedorov (1972, p.84) y que ahora nosotros hemos desarrollado para modelos lineales generalizados, indica que cuando la información se mide a través del determinante de $I(\beta)$, se mantiene un tipo de aditividad especial.

Proposición 2.1. *Bajo modelos lineales generalizados tal y como se han definido en esta sección, con $q \geq k + 1$,*

$$\det(I(\beta)) = \sum_{\{i_1, \dots, i_{k+1}\}} \det(I_{\{i_1, \dots, i_{k+1}\}}(\beta)), \quad (2.6)$$

donde la sumatoria es sobre todos los $\binom{q}{k+1}$ subconjuntos de $\{1, \dots, q\}$ de tamaño $k + 1$, y donde $I_{\{i_1, \dots, i_{k+1}\}}(\beta)$ es la matriz de información al restringir el experimento original a las observaciones asociadas a las filas i_1, \dots, i_{k+1} de la matriz del modelo X . Además,

$$\det(I_{\{i_1, \dots, i_{k+1}\}}(\beta)) = (\det(X_{i_1, \dots, i_{k+1}}))^2 n_{i_1} h(z_{i_1}) \dots n_{i_{k+1}} h(z_{i_{k+1}}). \quad (2.7)$$

Esta proposición nos muestra que se puede descomponer la información de cualquier experimento lineal generalizado de cualquier orden, en la suma de todos los $\binom{q}{k+1}$ sub-experimentos que se pueden formar cuando tomamos $k + 1$ observaciones de las q posibles que están recogidas en las filas de la matriz X . Su demostración queda recogida en el Apéndice A.

Este resultado es la base de los experimentos que hemos obtenido en esta memoria. Lo aplicaremos para modelos de respuesta binaria con distintos predictores lineales y para distintos link. A continuación hacemos una breve descripción de dichos modelos así como el desarrollo teórico de los determinantes para los distintos casos que desarrollaremos en capítulos posteriores.

2.4. Modelos de respuesta binaria

Si consideramos y la respuesta de un individuo o unidad experimental en un experimento de respuesta binaria, los valores que puede tomar dicha variable serán únicamente dos. Hemos considerado por conveniencia que los valores sean 0 y 1 respectivamente. Por ejemplo, supongamos un ensayo médico donde, al final del período de ensayo, el paciente se ha recuperado ($y = 1$) o no ($y = 0$). Por lo tanto, y se distribuye como una Bernoulli de parámetro $p = P(y = 1)$. En muchas investigaciones de diseño de experimentos o estudios observacionales, cada individuo o unidad experimental, tiene asociado un vector de covariables o variables explicativas $x = (x_1, \dots, x_k)$, que definen las condiciones bajo las que se lleva a cabo el experimento en cuestión. De forma análoga al caso lineal, el objetivo principal es investigar la relación entre la respuesta y y las variables explicativas x_1, \dots, x_k , con la salvedad de que en esta ocasión la respuesta es binaria.

Supongamos un experimento de tipo binario donde la i -ésima combinación de condiciones experimentales $x_i = (x_{1i}, \dots, x_{ki})$, $i = 1, \dots, q$, se repite n_i veces. En este caso se dice que las observaciones se encuentran agrupadas, siendo los tamaños de los grupos n_1, \dots, n_q . Así, y_i es el número de respuestas positivas en n_i pruebas o ensayos. En general, el número total de individuos, ensayos o pruebas, $n = \sum_{i=1}^q n_i$, viene fijado a priori.

Si las observaciones realizadas en los individuos son independientes y la probabilidad de una respuesta positiva es constante para todos los individuos del mismo grupo, entonces y_i es una variable aleatoria binomial de parámetros n_i y $P(y_i = 1|x_i) = p(x_i)$ tal que,

$$y_i|x_i \sim \text{Binomial}(n_i; p(x_i)).$$

Bajo este modelo, $E[y_i] = n_i p(x_i)$ y $\text{var}(y_i) = n_i p(x_i)(1 - p(x_i))$. En vez de modelar directamente la dependencia de y_i con las variables explicativas, es más natural explorar cómo la probabilidad de éxito $p(x_i) = E[y_i/n_i]$ puede ser descrita por dichas variables. Una aproximación para modelar datos binarios es hacerlo directamente, tal y como lo hacíamos para el modelo lineal general. Adoptamos el modelo donde

$p(x_i; \beta) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$. Bajo este supuesto, y sabiendo que los β_j , $j = 1, \dots, q$, no poseen ningún tipo de restricción, una combinación lineal de ellos, en principio, puede tomar cualquier valor en el intervalo $(-\infty, \infty)$. Por lo tanto, no habría garantías de que el valor de $p(x_i; \beta)$ tome valores en el intervalo $(0, 1)$.

En vez de usar un modelo lineal para definir la dependencia de la probabilidad de éxito y las variables explicativas, lo primero que se hará es transformar el intervalo $(0, 1)$ en $(-\infty, \infty)$. Estas transformaciones son lo que hemos definido en la Sección 2.3 como funciones link.

2.4.1. Funciones link para datos binarios

En la mayoría de la literatura especializada sobre datos binarios, aparecen 5 links como los más utilizados en modelos de respuesta binaria, el logit, el probit, doble exponencial, doble recíproco y complementario log-log. El link viene definido por la función $F^{-1}(\cdot)$ de forma que

$$F^{-1}(p(x_i; \beta)) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} = z_i.$$

A continuación definimos estos links y damos algunas de sus propiedades.

El link logit

El link *logit* de una probabilidad de éxito $p(x_i; \beta)$ viene definido por la expresión

$$\log \left(\frac{p(x_i; \beta)}{1 - p(x_i; \beta)} \right) = z_i.$$

Este logaritmo es lo que se conoce como $\text{logit}(p(x_i; \beta))$. Nótese que $p(x_i; \beta)/(1 - p(x_i; \beta))$ es el odds de un éxito y de esta forma la transformación logística de $p(x_i; \beta)$ es el log odds de un éxito. Se puede comprobar fácilmente que cualquier valor de $p(x_i; \beta)$ en el rango $(0, 1)$ se corresponde a un valor del $\text{logit}(p(x_i; \beta))$ en $(-\infty, \infty)$. Cuando $p(x_i; \beta) \rightarrow 0$, $\text{logit}(p(x_i; \beta)) \rightarrow -\infty$, y si $p(x_i; \beta) \rightarrow 1$, $\text{logit}(p(x_i; \beta)) \rightarrow \infty$. Además es el link canónico para modelos de respuesta binaria, y por lo tanto, el link más utilizado en este tipo de modelos.

El link probit

El link *probit* de una probabilidad $p(x_i; \beta)$ viene definido por la expresión

$$\Phi^{-1}(p(x_i; \beta)) = z_i,$$

donde $\Phi(\cdot)$ es la función de distribución de la variable normal estándar Z , de forma que $\Phi(z_i) = P(Z \leq z_i)$. Se cumple también que para cualquier valor de $p(x_i; \beta)$ en el rango $(0, 1)$, el valor correspondiente del probit de $p(x_i; \beta)$ tomará valores entre $-\infty$ e ∞ . Este link junto con el logit, son los que habitualmente se utilizan para modelar variables respuesta de tipo binario.

El link doble exponencial

El link *doble exponencial* de una probabilidad $p(x_i; \beta)$ viene dado por

$$\log(2 p(x_i; \beta)) = z_i \text{ si } p(x_i; \beta) < 0.5$$

$$-\log(2 (1 - p(x_i; \beta))) = z_i \text{ si } p(x_i; \beta) \geq 0.5$$

Nótese que en este caso no podemos dar una expresión cerrada para este link. Normalmente sólo es necesario conocer la función $F(\cdot)$ para obtener el modelo de respuesta binaria. Esta función viene dada por

$$p(x_i; \beta) = F(z_i) = \frac{(1 + s)}{2} - \frac{s}{2} e^{-|z_i|}$$

donde $s = \text{sign}(z_i)$. La función $\text{sign}(\cdot)$ tiene como dominio de definición \mathbb{R} y como conjunto imagen $\{-1; 0; 1\}$ de forma que si $z_i < 0 \Rightarrow \text{sign}(z_i) = -1$, si $z_i = 0 \Rightarrow \text{sign}(z_i) = 0$ y si $z_i > 0 \Rightarrow \text{sign}(z_i) = 1$.

El link doble recíproco

El link *doble recíproco* de una probabilidad $p(x_i; \beta)$ viene dado por

$$1 - (2 p(x_i; \beta))^{-1} = z_i \text{ si } p(x_i; \beta) < 0.5$$

$$((2(1 - p(x_i; \beta)))^{-1} - 1) = z_i \text{ si } p(x_i; \beta) \geq 0.5$$

Este link tampoco tiene una expresión cerrada. En este caso la función $F(\cdot)$ viene dado por

$$p(x_i; \beta) = F(z_i) = \frac{(1 + s)}{2} - \frac{s}{2} (1 + |z_i|)^{-1}$$

donde $s = \text{sign}(z_i)$. Al igual que el anterior, este link es menos utilizado que el logit o el probit.

El link complementario log-log

El link *complementario log-log* de una probabilidad $p(x_i; \beta)$, está definido por

$$\log\{-\log(1 - p(x_i; \beta))\} = z_i.$$

Como las anteriores, esta función transforma una probabilidad en el rango $(0, 1)$, en un valor en el intervalo $(-\infty, \infty)$.

Los link logit, probit, doble exponencial y doble recíproco son simétricos respecto del $ED_{.50}$. El complementario log-log es el único asimétrico. Tal y como expone Collet (2003), este link está limitado a aquellas situaciones donde es apropiado tratar con probabilidades de éxito de forma asimétrica. Por ejemplo, cuando se quiere modelar la probabilidad de que un material se rompa al aplicar una determinada fuerza. En este caso, conforme aumenta la fuerza aplicada a un componente, la probabilidad de rotura va aumentando de forma lenta hasta llegar a lo que se denomina tolerancia de la sección más débil del componente. A partir de aquí, la probabilidad de rotura aumenta de forma más rápida conforme aumenta la fuerza ejercida. De esta forma, el comportamiento de la probabilidad es distinta en ambos casos, por lo que haríamos uso del link complementario log-log. En la Figura 2.1 se observa dicha situación cuando usamos un modelo de respuesta binaria con predictor lineal logit (1) y el complementario log-log (2). El comportamiento de la curva (2) es más apropiado que el de la curva (1) ya que en el caso del logit, tal y como podemos observar, la probabilidad de rotura es igual para fuerzas de baja y alta intensidad

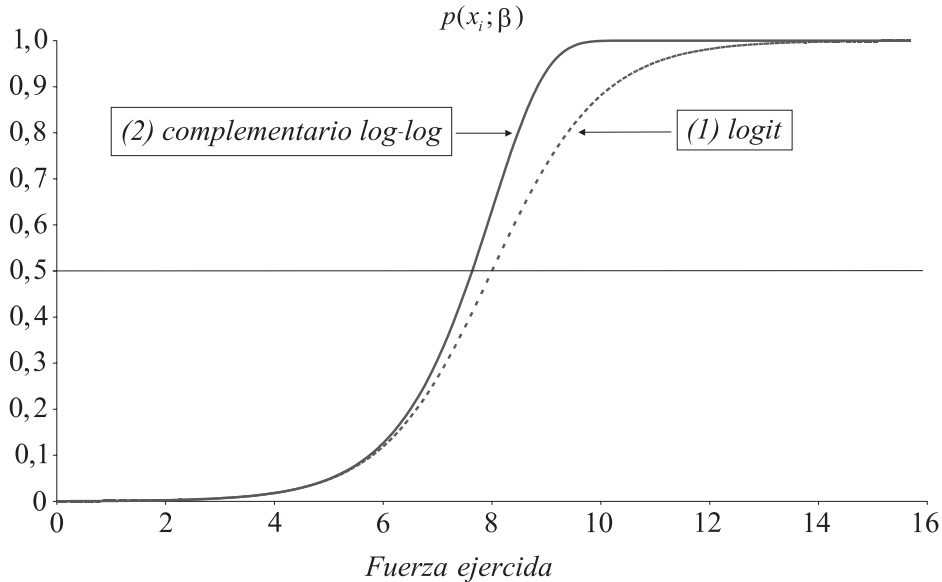


Figura 2.1: Dos posible modelos que definen el comportamiento de la probabilidad de rotura de un determinado componente. La curva (1) representa un modelo basado en el link logit y la (2) un modelo basado en el link complementario log-log.

ya que se trata de una curva simétrica respecto al $ED_{.50}$. Los link simétricos son apropiados en situaciones donde las probabilidades se comportan de forma simétrica. Collet (2003) expone algunos estudios epidemiológicos dónde se aplican este tipo de modelos, como pueden ser los estudios etiológicos que tratan de determinar si hay asociación entre uno o más factores de exposición y la ocurrencia de una determinada enfermedad. Esto se lleva a cabo realizando un experimento donde se aplican dichos factores a una serie de individuos y se calcula mediante un modelo de respuesta binaria la probabilidad de contraer una determinada enfermedad asociada a la exposición a unos niveles determinados de dichos factores. Los modelos basados en links simétricos son los más apropiados a la hora de plasmar el comportamiento de estas probabilidades, y dentro de éstos, son los link logit y probit los más utilizados en la literatura relacionada con modelos de respuesta binaria tanto en este área como en otras.

2.4.2. Modelos de respuesta binaria

El primer paso para definir un modelo de respuesta binaria es elegir el predictor lineal. En esta memoria hemos considerado los siguientes:

1. $z_i = \beta_0 + \beta_1 x_i$
2. $z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$
3. $z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$
4. $z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}$
5. $z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_{12} x_{1i} x_{2i} + \beta_{13} x_{1i} x_{3i} + \beta_{23} x_{2i} x_{3i} + \beta_{123} x_{1i} x_{2i} x_{3i}$

En la literatura especializada de problemas de respuesta binaria (Sitter y Torsney (1995a y b); McCullagh y Nelder (1989); Collett (2003)), estos modelos juegan un papel importantísimo tanto desde el punto de vista práctico como teórico. En esta memoria no se han tratado modelos de más de tres variables regresoras por su complejidad computacional, si bien en el caso balanceado los resultados de la optimización se obtenían en tiempo relativamente bajos, en el caso no balanceado dichos tiempos se disparaban. Nótese que en un problema de tres factores donde se trata de determinar el diseño localmente D-óptimo, tenemos que obtener 3 valores asociados a las x_i y $(8 - 1)$ λ_i 's. Esto es, debíamos optimizar en 10 variables para poder obtener el óptimo, algo que resultaba bastante complicado de llevar a cabo en cuanto al tiempo de ejecución requerido de los programas construidos al efecto. A veces para obtener el resultado debíamos esperar varios días ya que al ser un problema multivariante, calculábamos los diseños a través de rejillas en las variables a optimizar, con una precisión suficientemente alta para poder obtener resultados fiables, aumentando con esto la complejidad del problema.

Otra componente vital en este tipo de modelos son las curvas de nivel de las superficies a ajustar, definidas por las dosis efectivas, o curvas *ED*. Si consideramos un predictor lineal de primer orden sin interacciones, el conjunto de puntos

x_i que verifican la expresión $F^{-1}(p(x_i; \beta)) = z_i$, es un hiperplano en el espacio \mathbb{R}^k . A este conjunto se le denota por $ED_{F(z_i)}$. Por ejemplo, si $k = 2$, el conjunto $\{ED_{F(z_i)}/z_i \in \mathbb{R}\}$ es una sucesión de líneas rectas que definen las curvas de nivel de la superficie $p(x; \beta)$. Por otro lado, si consideramos un predictor lineal de primer orden con interacciones, el conjunto de puntos x_i que verifican la expresión $F^{-1}(p(x_i; \beta)) = z_i$, varía en función de k . Si $k = 2$ este conjunto de puntos definen una hipérbola mientras que para $k = 3$ definen un hiperboloide. Para valores de k superiores a tres, las superficies son cuádricas.

Para cada una de las funciones link definidas en la Sección 2.4.1 podemos obtener un modelo de respuesta binaria. Los modelos tratados en esta memoria han sido principalmente dos, el logístico y el probit, aunque de forma más resumida hemos tratado modelos basados en los link doble exponencial, doble recíproco y complementario log-log.

Un modelo de respuesta binaria se denomina modelo logístico si $F(\cdot)$ es la función de distribución logística acumulada, $e^{z_i}/(1 + e^{z_i})$. Los primeros trabajos relacionados con estos modelos se remontan a los años cuarenta y cincuenta. Así, Dyke y Patterson (1952) publicaron un análisis de una muestra sobre la proporción de individuos que tenían buen conocimiento del cáncer. Se propuso un modelo factorial donde el logit o log-odd de éxito (buen conocimiento de) se expresaba como una combinación lineal de las fuentes de información y sus interacciones, esto es, un modelo logit. Este tipo de modelos se ha utilizado usualmente en bioensayos (ver por ejemplo, Berkson (1944) y (1951)).

Por otro lado, un modelo probit es aquel cuya función $F(\cdot)$ es la función de distribución normal estándar acumulada, $\Phi(z_i)$. Bliss (1935), uno de los primeros en utilizar estos modelos en sus investigaciones, empleó la técnica de análisis probit en el campo de bioensayos dentro de un problema de toxicología . Dividió una población de estudio de insectos en grupos de tamaño no forzosamente idéntico. A cada grupo se le aplicó una cierta dosis de toxina (nivel de la toxina), guardando registro del número de individuos que sobrevivieron de cada grupo. El objetivo consistió en

Link	$F(z_i) = p(x_i; \beta)$	$h(z_i)$
Logit	$e^{z_i}/(1 + e^{z_i})$	$e^{z_i}/(1 + e^{z_i})^2$
Probit	$\Phi(z_i)$	$e^{-z_i^2}/(2\pi\Phi(z_i)\Phi(-z_i))$
D. Exponencial	$(1 + s_i - s_i e^{- z_i })/2$	$(2e^{ z_i } - 1)^{-1}$
D. recíproco	$(1 + s_i - s_i(1 + z_i)^{-1})/2$	$((1 + z_i)^2(2(1 + z_i) - 1))^{-1}$
Comp. log-log	$1 - e^{-e^{z_i}}$	$e^{2z_i}/(e^{e^{z_i}} - 1)$

Tabla 2.2: Función de distribución acumulada $F(\cdot)$ para los link que definen modelos de respuesta binaria y función $h(z_i) = F'(z_i)^2/(F(z_i)(1 - F(z_i)))$ para cada uno de ellos. $\Phi(\cdot)$ es la función de distribución de la normal estándar y $s_i = \text{sign}(z_i)$.

modelizar la proporción de individuos que sobrevivían en función de la dosis de toxina (en unidades logarítmicas).

Las distribuciones normales y logísticas son simétricas a cero y su representación gráfica, cuando es posible obtenerla, es muy similar, excepto que la logística tiene más ancha las colas. Por lo tanto, las funciones de probabilidad condicional son muy parecidas para los dos modelos, y así tanto el modelo logístico como el probit producen predicciones similares, excepto para valores extremos de las condiciones experimentales x_i .

Otros modelos utilizados son el doble exponencial, el doble recíproco y el complementario log-log. Los dos primeros pertenecen a modelos con links simétricos mientras que el tercero es un modelo con link no simétrico. Las funciones de distribución acumulada $F(\cdot)$ que definen dichos modelos, están recogidas en la Tabla 2.4.2. Estos modelos son menos utilizados en la práctica aunque son apropiados para describir fenómenos que no pueden ser modelados a través de los modelos logístico o probit.

Una vez decidido cuál es el modelo más apropiado para describir el fenómeno en estudio, se debe presentar un experimento que determine cuáles son las condiciones experimentales óptimas bajo las que se debe observar para poder estimar el modelo que acumule el máximo de información del proceso. En este trabajo nos hemos decantado por los diseños a dos niveles tal y como quedó reflejado en el Capítulo

1. Para determinar la información que contiene un experimento debemos hallar en primera instancia la matriz de información de Fisher y calcular su determinante. En la siguiente sección se obtiene la expresión de la matriz de información de Fisher y su determinante para los modelos de respuesta binaria. Nótese que dicha matriz se deriva de forma directa de la matriz de información para el caso de modelos lineales generalizados (2.3).

2.4.3. Matriz de Información

La matriz de información para un modelo de respuesta binaria en el caso de dos variables de diseño, esto es $k = 2$, de un experimento cuyos puntos soporte son x_1, \dots, x_q tal que $x_i \in \mathbb{R}^2$ es

$$I(\beta) = n \begin{pmatrix} \sum_i^q \lambda_i h(z_i) & \sum_i^q \lambda_i x_{1i} h(z_i) & \sum_i^q \lambda_i x_{2i} h(z_i) \\ \sum_i^q \lambda_i x_{1i} h(z_i) & \sum_i^q \lambda_i x_{1i}^2 h(z_i) & \sum_i^q \lambda_i x_{1i} x_{2i} h(z_i) \\ \sum_i^q \lambda_i x_{2i} h(z_i) & \sum_i^q \lambda_i x_{1i} x_{2i} h(z_i) & \sum_i^q \lambda_i x_{2i}^2 h(z_i) \end{pmatrix} \quad (2.8)$$

donde $z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$, $\lambda_i = n_i/n$ y $h(z_i) = F'(z_i)^2 / (F(z_i)(1 - F(z_i)))$. En la Tabla 2.4.2 se muestran los valores de $h(z_i)$ correspondientes a los distintos links para datos binarios. Si $k > 2$, la matriz de información de Fisher de un experimento con puntos soporte x_1, \dots, x_q en \mathbb{R}^k , tiene dimensión $(k+1) \times (k+1)$ y viene definida por la expresión

$$I(\beta) = n \sum_i^q \lambda_i h(z_i) (1, x_{1i}, \dots, x_{ki})' (1, x_{1i}, \dots, x_{ki}). \quad (2.9)$$

La diferencia con la matriz de información para modelos lineales generalizados (2.3) radica en el término λ_i . Para obtener los diseños D-óptimos necesitamos maximizar el determinante de la matriz de información eligiendo de forma óptima los valores de q , x_i y n_i . Debido a que n_i debe ser entero, el problema de optimización se complica, lo que hace que debamos considerar el espacio contínuo en el cual n_i/n y $n = \sum_i n_i$ son reemplazados por un real λ_i con $0 < \lambda_i < 1$ y $\sum_i \lambda_i = 1$. Además, en respuesta binaria el modelo es de la forma $E[y_i/n] = z_i$, por lo que la definición

del determinante y la matriz de información en función de λ_i también es coherente desde este punto de vista.

En esta memoria hemos estudiado principalmente los diseños localmente D-óptimos para 2 y 3 factores ya que para un factor Kalish y Rosenberg (1978) obtuvieron el diseño a dos niveles localmente D-óptimo tal y como apuntamos en la Sección 1.5.1. En este caso nos dedicamos a calcular los diseños localmente D-óptimos en distintas regiones del espacio experimental, sabiendo que el mejor de todos era el obtenido por Kalish y Rosenberg.

2.5. El Criterio de la D-optimalidad para modelos de respuesta binaria

Una de las características más importantes del criterio de la D-optimalidad es que induce un orden de los experimentos que es invariante bajo una reparametrización, algo especialmente deseable cuando los parámetros no tienen un significado físico definido. Por lo tanto, los experimentos D-óptimos se obtienen por maximización del determinante de (2.9).

Ahora bien, en la literatura especializada hemos comprobado que los predictores lineales varían en su forma. Sitter (1992), Sitter y Wu (1993) o Sitter y Torsney (1995b) consideran el predictor lineal de la forma

$$z_i = \beta_1(x_{i1} - \mu) + \beta_2(x_{i2} - \mu), \quad (2.10)$$

mientras que otros autores tales como Myers, Myers y Carter (1994) o Ford, Torsney y Wu (1992) se decantaron por el predictor $z_i = \beta_0 + \beta_1 x_{i2} + \beta_2 x_{i1}$. En esta memoria sólo trabajaremos con la expresión (2.9) ya que en el caso de tomar el predictor lineal (2.10) los cálculos son análogos. Para modelos de respuesta binaria, el determinante de (2.9) es

$$\det(I(\beta)) = n^{k+1} \det \left(\sum_{i=1}^q \lambda_i h(z_i) (1, x_{1i}, \dots, x_{ki})' (1, x_{1i}, \dots, x_{ki}) \right) \quad (2.11)$$

Sea X la matriz de diseño definida en (2.5) de dimensión $q \times (k + 1)$, cuya i -ésima fila es $(1|x_i) = (1, x_{1i}, \dots, x_{ki})$, y sea $X_{i_1, \dots, i_{k+1}}$ la submatriz $(k + 1) \times (k + 1)$ de X formada por sus filas i_1, i_2, \dots, i_{k+1} .

Nótese que a diferencia de lo que ocurre en modelos lineales normales, bajo modelos binomiales, la ortogonalidad de las columnas de X no implica que la matriz de información sea diagonal y, por lo tanto, incluso en un contexto restringido de diseños ortogonales, las cuestiones a tener en cuenta no se restringen a que puede ser estimado y con qué precisión, cosa que sí ocurría para los modelos normales de primer orden. Por lo tanto, conceptos tales como la resolución o aberración de un experimento factorial fraccional pierden su significado fuera del conjunto de respuestas normales.

El siguiente corolario se deduce directamente de la Proposición 2.1 y se aplicará en capítulos sucesivos para verificar muchas de las propiedades de los diseños localmente D-óptimos a dos niveles cuando se consideran varios factores.

Corolario 2.1. *Bajo el modelo binomial con $p(x_i, \beta) = F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki})$, el determinante de $I(\beta)$ para cualquier experimento soportado en x_1, \dots, x_q en \mathbb{R}^k es*

$$\det(I(\beta)) = \sum_{\{i_1, \dots, i_{k+1}\}} \det(I_{\{i_1, \dots, i_{k+1}\}}(\beta)) = n^{k+1} \sum_{\{i_1, \dots, i_{k+1}\}} (\det(X_{i_1, \dots, i_{k+1}}))^2 \lambda_{i_1} h(z_{i_1}) \dots \lambda_{i_{k+1}} h(z_{i_{k+1}}), \quad (2.12)$$

donde $z_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$, $\lambda_i = n_i/n$ y donde la sumatoria es sobre todos los $\binom{q}{k+1}$ subconjuntos de $\{1, \dots, q\}$ de tamaño $k + 1$, $\{i_1, \dots, i_{k+1}\}$. Se denota por $I_{\{i_1, \dots, i_{k+1}\}}(\beta)$ a la matriz de información del experimento restringido del original cuando únicamente consideramos los puntos soporte $x_{i_1}, \dots, x_{i_{k+1}}$.

Esto implica que la información del experimento, medida a través de su matriz de información, es la suma de la información de cada uno de los sub-experimentos soportados en $k + 1$ puntos, cuando el número total de puntos es q .

El determinante de $I(\beta)$ depende de los parámetros desconocidos (β) a través de los z_i , y sólo a través de ellos. Un diseño localmente D-óptimo, ver Sección 1.5.1, parte de una estimación inicial de los parámetros β , a partir de la cual se calculan los valores de q y (x_i, λ_i) que maximizan el determinante de la matriz de información para los parámetros iniciales elegidos, algunas veces con la restricción de que los x_i 's han de pertenecer a una región de interés previamente especificada.

En el caso de un único factor, $k = 1$, la solución de este problema de diseños D-óptimos para una región experimental no acotada es finito, y tiene únicamente dos puntos soporte, ver, por ejemplo, Abdelbasit y Plackett (1983), Minkin (1987), Khan y Yadzi (1988), Ford, Torsney y Wu (1992), Sitter y Wu (1993), Myers, Myers, y Carter (1994), Sitter y Fainaru (1997) y Mathew y Sinha (2001). En particular, es bien conocido que para el modelo logístico la solución D-óptima está soportada en los puntos $ED_{.176}$ y $ED_{.824}$, con $\lambda_i = .5$. En el Capítulo 3 desarrollamos este caso, calculando el experimento D-óptimo de dos puntos centrado en un punto x_0 definido a priori.

En el caso de dos factores, Sitter y Torsney (1995a) comprueban que si la región experimental es no acotada, el determinante (2.9) es no acotado, esto es, se puede hacer tan grande como se quiera. Esta situación es similar a la que surge en el contexto de modelos lineales normales. Así, para poder calcular los diseños D-óptimos apuntan la necesidad de considerar espacios de diseño acotados, por razones tanto prácticas como teóricas. Las teóricas parten del hecho de que si no son consideradas dichas acotaciones no se podrían calcular los diseños D-óptimos. Las razones prácticas surgen, por ejemplo, en problemas de dosis-respuesta cuando un investigador desea que todos los niveles de dosis tengan una probabilidad de respuesta moderada, para evitar así el problema de no respuesta o respuesta completa en un nivel de dosis particular. Esto implicaría que no se pudiese llevar a cabo el experimento para ese nivel de dosis y contribuiría a tener poca información del comportamiento de la curva. Otro caso podría darse cuando el investigador no desea usar dosis demasiado elevadas debido a los posibles efectos tóxicos de un determinado fármaco o que no

es factible producir niveles bajos de dosis de un determinado agente. En base a esto restringieron el espacio del diseño a una clase de regiones que fueran manejables en diferentes situaciones prácticas y además que permitieran obtener fácilmente los diseños óptimos. Nótese que los autores en este caso tomaron un predictor lineal de la forma $z_i = \beta_1(x_{1i} - \mu) + \beta_2(x_{2i} - \mu)$ con $i = 1, \dots, q$. Las consideraciones prácticas enumeradas anteriormente los llevaron a considerar acotaciones de la forma:

$$b_1 \leq \beta_1(x_1 - \mu) + \beta_2(x_2 - \mu) \leq b_2 \quad (2.13)$$

$$-1 \leq a_1\beta_1(x_1 - \mu) + a_2\beta_2(x_2 - \mu) \leq 1 \quad (2.14)$$

donde a_1 , a_2 , b_1 y b_2 , son constantes elegidas por el investigador. La expresión (2.13) determina la restricción sobre las líneas ED , mientras que la (2.14) establece dónde situar los puntos sobre las líneas de dosis efectivas (ver Figura 2.2). Bajo dichas restricciones obtuvieron los diseños D-óptimos para cada uno de los link asociados a modelos de respuesta binaria. Para el link logit el diseño D-óptimo está formado por cuatro puntos en los que se tomarán el mismo número de observaciones y se sitúan sobre las líneas de dosis efectivas del 77.2% y 22.7%. Para el link Probit el diseño D-óptimo también está formado por cuatro puntos situados dos a dos sobre las líneas del 82,5% y 17.4% tomándose en dichos puntos el 25% de las observaciones. Para el link Doble exponencial el diseño está formado por seis puntos. Cuatro de ellos sobre las líneas del 89.8% y 10.1%. En cada uno de estos puntos se toman el 9.4% de las observaciones totales. Los otros dos puntos se sitúan sobre la línea del 50.0% tomándose en cada uno de ellos el 31.2% de las observaciones totales. Para el doble recíproco ocurre algo parecido; su diseño D-óptimo está formado por seis puntos, cuatro de ellos situados dos a dos sobre las líneas del 79.2% y 20.7% tomándose en cada uno de ellos el 8.7% de las observaciones totales. Los otros dos se sitúan sobre la línea del 50,0% de forma que el 32.6% de los observaciones totales se tomarán sobre cada uno de ellos. Por último, en el caso del complementario log-log hay que recordar que dicho link es asimétrico y que esto quedará reflejado en su diseño D-óptimo. Éste se compone de cuatro puntos, dos de ellos situados sobre la línea del 28.7% de forma que en cada uno de ellos se toma el 21.1% de las

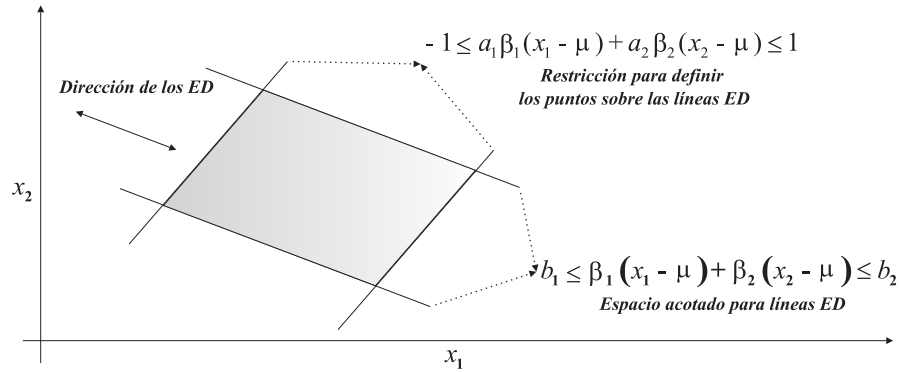


Figura 2.2: Región acotada para calcular diseños D-óptimos de dos factores, para un modelo de respuesta binaria según Sitter y Torsney (1993a).

observaciones totales. Los otros dos puntos se sitúan sobre la línea del 90.4% y el porcentaje de observaciones tomadas en cada uno de ellos será del 28.9%. Nótese que aquí los puntos no están soportados sobre líneas de probabilidad simétricas respecto del 50.0%. Así, ha quedado demostrado que los diseños D-óptimos para el caso de modelos de respuesta binaria con dos factores y con predictor lineal sin interacciones únicamente se pueden obtener bajo restricciones de la región de experimentación. Sin embargo, en el Capítulo 3 demostramos que en el caso de experimentos a dos niveles, a menos que $\beta_1 = 0$ ó $\beta_2 = 0$, el experimento D-óptimo es finito incluso para regiones no acotadas. En el Capítulo 5 vemos que cuando se consideran tres factores ocurre lo mismo, al igual que cuando se consideran interacciones, Capítulo 6.

2.6. Determinante de $I(\beta)$ para experimentos a dos niveles

Sea un experimento soportado en q de los 2^k puntos de la forma $(x_{10} + a_{i1}R_1, x_{20} + a_{i2}R_2, \dots, x_{k0} + a_{ik}R_k)$, donde los a_{ij} 's pueden tomar los valores -1 ó 1 , R_1, \dots, R_k son las mitades de los rangos y $x_0 = (x_{10}, \dots, x_{k0})$ su punto central perteneciente al

hiperplano $ED_{F(z_0)}$ con $z_0 = \beta_0 + \beta_1 x_{10} + \dots + \beta_k x_{k0}$. Sea A la $q \times (k+1)$ matriz de diseño estandarizada, cuya i -ésima fila es el vector $(1|a_i) = (1, a_{i1}, a_{i2}, \dots, a_{ik})$,

$$A = \begin{pmatrix} 1 & a_{11} & \dots & a_{1k} \\ 1 & a_{21} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & a_{q1} & \dots & a_{qk} \end{pmatrix}, \quad (2.15)$$

y sea $A_{i_1, \dots, i_{k+1}}$ la $(k+1) \times (k+1)$ sub-matriz de A formada por las filas i_1, i_2, \dots, i_{k+1} .

Incluso aunque en los diseños factoriales y factoriales fraccionales a dos niveles que normalmente se utilizan en variables respuesta de tipo continuo, los diseños se restringen a aquellos cuyas matrices de diseño son ortogonales, A , no necesariamente lo es, ya que para modelos de respuesta binaria estos experimentos dejan de ser óptimos. En la Proposition (2.2) aportamos una expresión cerrada del determinante de $I(\beta)$ para cualquier experimento a dos niveles.

Proposición 2.2. *Bajo un modelo binomial con $p(x_i, \beta) = F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki})$, el determinante de $I(\beta)$ para cualquier experimento soportado en los $q = 2^k$ vértices de un hiper-rectángulo en \mathbb{R}^k , de la forma $x_i = (x_{10} + a_{i1}R_1, \dots, x_{k0} + a_{ik}R_k)$, es tal que*

$$(\beta_1 \dots \beta_k)^2 \det(I(\beta)) = n^{k+1} \prod_{j=1}^k (\beta_j R_j)^2 \sum_{\{i_1, \dots, i_{k+1}\}} (\det(A_{i_1, \dots, i_{k+1}}))^2 \lambda_{i_1} h(z_{i_1}) \dots \lambda_{i_{k+1}} h(z_{i_{k+1}}), \quad (2.16)$$

donde $z_i = z_0 + a_{i1}|\beta_1|R_1 + \dots + a_{ik}|\beta_k|R_k$ indican los $ED_{F(z_i)}$'s de los puntos soporte, donde $\lambda_i = n_i/n$, y donde la sumatoria varía en los $\binom{2^k}{k+1}$ subconjuntos de $\{1, 2, 3, \dots, 2^k\}$ de tamaño $k+1$.

Este resultado engloba casos especiales tales como experimentos factoriales completos a dos niveles o cualquier fracción de ellos. Cuando no deseamos que una de las 2^k combinaciones experimentales posibles esté incluida en el experimento, únicamente debemos ajustar el correspondiente λ_i a 0.

El tomar valor absoluto en los parámetros del modelo, nos ha permitido trabajar de forma más ordenada con los resultados obtenidos en la experiencias computacionales, cuando se maximizaba el determinante de la matriz de información. En capítulos sucesivos veremos qué implicaciones tiene este valor absoluto y cuáles son sus beneficios a la hora de calcular el diseño D-óptimo.

Nótese que $(\beta_1 \dots \beta_k)^2 \det(I(\beta)) / n^{k+1}$ en (2.16) es una función de z_0 y de los vectores $(|\beta_1| R_1, \dots, |\beta_k| R_k)$, y $(\lambda_1, \dots, \lambda_{2k})$. Por lo tanto, todos los experimentos centrados en un punto situado sobre el $ED_{F(z_0)}$ con los mismos $|\beta_i| R_i$'s y λ_j 's, tienen el mismo valor de $(\beta_1 \dots \beta_k)^2 \det(I(\beta)) / n^{k+1}$. Calculando los $|\beta_i| R_i$ (y λ_j) que maximizan (2.16) dado un z_0 , obtenemos las mitades de los rangos R_1, \dots, R_k (y los $\lambda_1, \dots, \lambda_{2k}$) que determinan el diseño D-óptimo local a dos niveles centrado en el $ED_{F(z_0)}$, para cualquier β .

2.7. Conclusiones

En este capítulo se presenta una descripción general de los modelos lineales generalizados y de forma particular, de los modelos de respuesta binaria. En ambos casos se aporta la matriz de información de Fisher así como su determinante. Nótese que el criterio de optimalidad utilizado es el de la D-optimalidad. Las propiedades más importantes que se aportan están basadas en el hecho de que en ambos casos la información contenida en un experimento puede descomponerse en la suma de sub-experimentos que se pueden formar a partir de las filas de la matriz del diseño X . Además, se obtiene la expresión general del determinante de la matriz de información para un modelo de respuesta binaria cuando se considera un diseño a dos niveles con k factores. Con todo ello hemos definido el problema en conjunto y a partir de aquí vamos a desarrollar algunos casos en función del número de factores a tomar, si hay o no interacciones y si existen restricciones en el espacio del diseño.

Experimentos a dos niveles para modelos de respuesta binaria

Capítulo 3

Experimentos con un solo factor

Basándonos en los resultados del capítulo anterior, sabemos que el determinante de la matriz de información de Fisher para modelos de respuesta binaria cuando se consideran experimentos a dos niveles, depende de la localización del centro del diseño. Así, todos aquellos diseños centrados sobre un mismo z_0 , o región ED , tienen el mismo determinante de la matriz de información. En otras palabras, si queremos calcular un diseño D-óptimo en un entorno de condiciones experimentales, debemos calcular previamente el ED asociado a dichas condiciones, que es donde vamos a centrar nuestro diseño, para posteriormente maximizar el determinante de la matriz de información. Por otro lado, nos puede interesar cuáles son las condiciones experimentales óptimas para estimar nuestro diseño. Esto lo haríamos estudiando el determinante de los diseños D-óptimos para los distintos ED 's y extrayendo el máximo. En este caso tendríamos el diseño D-óptimo general y con ello las condiciones óptimas donde observar. Tanto en este capítulo como en el 4 y el 5 llevaremos a cabo ambos estudios. De esta forma caracterizamos, para el predictor lineal sin interacciones, los diseños D-óptimos a dos niveles con un factor (en este capítulo), con dos factores (Capítulo 4) y con tres factores (Capítulo 5). En el Capítulo 6 se introducen interacciones en el predictor lineal. Esto hace que las propiedades de los diseños D-óptimos cambien y por ello se ha considerado un capítulo para este caso.

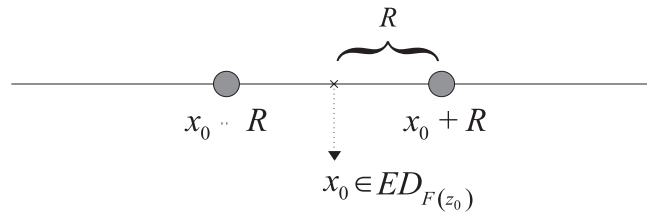


Figura 3.1: Experimento a dos niveles con un factor.

3.1. Experimentos con un solo factor dependientes de z_0

En esta sección, calculamos los experimentos D-óptimos, dentro de la clase de diseños de dos puntos simétricos con respecto a un punto central que se sitúa sobre un ED fijado a priori. Esto es, consideramos experimentos soportados en (x_1, x_2) , de forma que $x_i = x_0 + a_{i1}R$, donde a_{i1} toma valores 1 ó -1 y x_0 es un punto situado en la dosis efectiva $ED_{F(z_0)}$ con $z_0 = \beta_0 + \beta_1 x_0$, tal y como se muestra en la Figura 3.1. Nótese que la región ED es un punto y por lo tanto coincide con el centro del diseño. Nuestro cometido es calcular el valor de R que maximiza el determinante de $I(\beta)$, para posteriormente buscar, de entre todos los diseños obtenidos, aquel cuyo determinante sea máximo. Esto es, buscaremos la dosis efectiva donde debe estar centrado el diseño D-óptimo a dos niveles con un factor. Nótese que en modelos de respuesta binaria con link logit y un factor, el diseño D-óptimo obtenido por Kalish y Rosenberg (1978) es un diseño de dos puntos simétricos respecto al $ED_{.50}$. El diseño D-óptimo que aportemos va a coincidir con el diseño obtenido por Kalish y Rosenberg ya que en ambos casos son diseños a dos niveles.

En el siguiente corolario, el cual se obtiene de la aplicación directa de la Proposición 2.2, se muestra la expresión del determinante de la matriz de información de Fisher para este diseño.

Corolario 3.1. *Bajo un modelo binomial con $p(x_i, \beta) = F(\beta_0 + \beta_1 x_i)$, el determinante de $I(\beta)$ para cualquier experimento soportado en los puntos $x_0 - R$ y $x_0 + R$ es tal que*

$$\beta_1^2 \det(I(\beta)) = 4n^2 (\beta_1 R)^2 \lambda_1 \lambda_2 h(z_1) h(z_2), \quad (3.1)$$

donde $z_1 = z_0 - |\beta_1| R$, $z_2 = z_0 + |\beta_1| R$, $\lambda_2 = n_2/n = 1 - \lambda_1$ y $h(z_i) = F'(z_i)^2 / (F(z_i)(1 - F(z_i)))$.

Para un factor, las asignaciones D-óptimas para cualquier z_0 dado, son balanceadas (con $\lambda_i = .5$) puesto que el número de puntos soporte coincide con el número de parámetros. En la Tabla 3.1 se recogen los valores de $|\beta_1| R$, $F(z_1)$, $F(z_2)$, y $\beta_1^2 \det(I(\beta))$ para un experimento D-óptimo de dos puntos centrado en varios $F(z_0)$'s y que permiten obtener los valores localmente D-óptimos de R , z_1 y z_2 para cualquier valor de $\beta = (\beta_0, \beta_1)$, tanto para el link logit como para el probit. Los experimentos D-óptimos para valores negativos de los z_0 's, pueden ser calculados a partir de esta tabla por simetría, sin más que considerar $F(z_1) = 1 - F(z_2)$ y $F(z_2) = 1 - F(z_1)$.

Tal y como ya comentamos en el Capítulo 2, el valor absoluto de los parámetros del modelo se introduce para simplificar la interpretación de los resultados que se obtienen al maximizar el determinante de la matriz de información. En definitiva lo que se está haciendo, cuando se introduce el valor absoluto, es calcular el diseño D-óptimo cuando los parámetros del modelo son positivos. Veamos un ejemplo que ilustra el comportamiento del diseño D-óptimo cuando los parámetros del predictor lineal cambian de signo.

Ejemplo 3.1: Una empresa farmacéutica desea obtener un modelo que estime el porcentaje de pacientes recuperados cuando se les administra un fármaco en una determinada dosis. Se desea obtener un modelo de respuesta binaria con link logit de un factor. Dicho modelo deberá estimar de forma fiable el número de pacientes que se recuperan al administrarles una dosis determinada del fármaco. De todas las posibles dosis, al investigador le interesan aquellas que estén entorno al 30%

Link Logit

$F(z_0)$	z_0	$ \beta_1 R$	$F(z_1)$	$F(z_2)$	$\lambda_{1,2}$	$\beta_1^2 \det(I(\beta))$
.50	.000	1.5434	.1760	.8240	.5	.0501 n^2
.55	.201	1.5487	.2062	.8519	.5	.0495 n^2
.60	.405	1.5652	.2387	.8777	.5	.0478 n^2
.65	.619	1.5937	.2739	.9014	.5	.0449 n^2
.70	.847	1.6371	.3122	.9230	.5	.0409 n^2
.75	1.099	1.6995	.3541	.9426	.5	.0358 n^2
.80	1.386	1.7885	.4008	.9599	.5	.0296 n^2
.85	1.735	1.9196	.4539	.9748	.5	.0225 n^2
.90	2.197	2.1286	.5171	.9870	.5	.0146 n^2
.95	2.944	2.5339	.6012	.9958	.5	.0064 n^2

Link Probit

$F(z_0)$	z_0	$ \beta_1 R$	$F(z_1)$	$F(z_2)$	$\lambda_{1,2}$	$\beta_1^2 \det(I(\beta))$
.50	.000	1.1382	.1275	.8725	.5	.1987 n^2
.55	.126	1.1372	.1559	.8967	.5	.1960 n^2
.60	.253	1.1345	.1891	.9174	.5	.1881 n^2
.65	.385	1.1298	.2283	.9351	.5	.1750 n^2
.70	.524	1.1230	.2747	.9503	.5	.1571 n^2
.75	.674	1.1138	.3302	.9631	.5	.1348 n^2
.80	.842	1.1018	.3974	.9740	.5	.1086 n^2
.85	1.036	1.0860	.4802	.9831	.5	.0795 n^2
.90	1.282	1.0650	.5857	.9905	.5	.0489 n^2
.95	1.645	1.0351	.7290	.9963	.5	.0196 n^2

Tabla 3.1: Experimentos D-óptimos de dos puntos para un factor, centrados en el $ED_{F(z_0)}$, bajo los modelos logísticos y probit. El mejor experimento de dos puntos se encuentra resaltado en negrita.

de enfermos recuperados, ya que a partir de este porcentaje se considera que el fármaco cumple su finalidad. El simétrico de .30 en la Tabla 3.1 es el .70, de forma que para $F(z_0) = .70$, $|\beta_1|R = 1.6371$, $F(z_1) = 1 - .9503$, $F(z_2) = 1 - .2747$, y el determinante de su $I(\beta)$ es $.1571n^2/\beta_1^2$. Así, para llevar a cabo el experimento el investigador tomará dos grupos de pacientes. Al primer grupo se le administra una dosis de $x_0 - 1.1230/|\beta_1|$ unidades y al segundo una de $x_0 + 1.1230/|\beta_1|$ unidades. Este resultado depende de los parámetros del modelo. Para resolver el problema debemos conocer una estimación de dichos parámetros como ya comentamos en el

Capítulo 1. Supongamos que $\beta_0 = 0$ y $\beta_1 = -1$. En este caso, tendríamos que el predictor lineal es de la forma $z = -x$. Sabemos que $x_0 \in ED_{.3}$ y por lo tanto

$$F(z_0) = 0.3 = \frac{e^{-x_0}}{1 + e^{-x_0}}$$

de lo que deducimos que $x_0 = 0.84729$ y los puntos del diseño son $x_1 = x_0 - 1.1230 = -0.27571$ y $x_2 = x_0 + 1.1230 = 1.97029$. Por otro lado, si la estimación inicial de los parámetros fuese $\beta_0 = 0$ y $\beta_1 = 1$ entonces obtenemos que $x_0 = -0.84729$ y los puntos del diseño son $x_1 = x_0 - 1.1230 = -1.97029$ y $x_2 = x_0 + 1.1230 = 0.27571$. Cuando $\beta_1 = -1$ el orden de los puntos de menor a mayor en la recta real sería x_1, x_2 mientras que para $\beta_1 = 1$ el orden no cambia. Así concluimos que el signo de los parámetros del modelo no influye en el determinante de la matriz de información ni en los ED 's sobre los que debemos situar los puntos del diseño. Por lo tanto, podemos considerar sin pérdida de generalidad el caso en el que dichos parámetros sean positivos, esto es, queda justificado el uso del valor absoluto a la hora de maximizar el determinante de la matriz de información.

Para finalizar con el ejemplo se mide la variable respuesta "Número de pacientes que se han recuperado" en los puntos obtenidos y con este resultado y las dosis administradas podremos determinar el mejor modelo de respuesta binaria.

En la Tabla 3.1, para el modelo logístico, conforme aumenta $F(z_0)$, más allá del $ED_{.50}$, también aumenta el valor óptimo de $|\beta_1|R$, mientras que en el caso del modelo probit ocurre todo lo contrario. Así mismo, se observa que, cuando usamos el link probit, los determinantes de $I(\beta)$ para los experimentos D-óptimos son considerablemente mayores, y los valores de R son considerablemente más pequeños que para el caso del link logit. Estas propiedades permiten al investigador decantarse por uno u otro diseño según sus necesidades.

Después de haber estudiado cada uno de los diseños D-óptimos de dos puntos centrados en un determinado x_0 , hemos extraído los diseños con mayor determinante. En la Tabla 3.1 se recogen los experimentos D-óptimos de dos puntos para algunas funciones link en modelos de respuesta binaria. Nótese que el determinante de los experimentos depende de la función link, de forma que, se puede observar

Link	$F(z_0)$	z_0	$ \beta_1 R$	$F(z_1)$	$F(z_2)$	$\lambda_{1,2}$	$\beta_1^2 \det(I(\beta))$
Logit	.50	.000	1.5434	.1760	.8240	.5	.0501 n^2
Probit	.50	.000	1.1382	.1275	.8725	.5	.1987 n^2
Doble exponencial	.8009	.921	.92080	.5000	.9272	.5	.0730 n^2
Doble Recíproco	.7235	.808	.80832	.5000	.8089	.5	.0225 n^2
Compl. log-log	.5666	-.179	1.1587	.2308	.9303	.5	.1638 n^2

Tabla 3.2: Experimentos D-óptimos de un factor de dos puntos. Para el logit, probit y complementario log-log, estos experimentos de dos puntos son los experimentos D-óptimos globales. Para el doble exponencial y doble recíproco, los experimentos D-óptimos globales son de tres puntos con $\beta_1^2 \det(I(\beta))$ igual a $.081n^2$ y $.023n^2$, respectivamente.

cómo los link probit y complementario log-log son potencialmente más informativos respecto de β que los link logit y doble recíproco. Tanto para el logit como para el probit, el experimento de dos puntos D-óptimo está centrado en el $ED_{.50}$, y para el complementario log-log, está centrado en el $ED_{.566}$, siendo dichos experimentos D-óptimos globales, de forma que maximizan el determinante de $I(\beta)$ entre todos los posibles experimentos, (ver White, (1975); Ford, Torsney y Wu, (1992)). Para la función link doble exponencial, el experimento D-óptimo de dos puntos está soportado en el $ED_{.50}$ y el $ED_{.9272}$, (o en el $ED_{.0728}$ y el $ED_{.50}$), con determinante de su matriz de información, un 37.5% más grande que el determinante del diseño centrado en el $ED_{.50}$ y simétrico, que Ford, Torsney y Wu (1992) definen en su artículo, en la Tabla 4, como el mejor diseño de dos puntos para dicho link. El experimento D-óptimo global para este link (ver por ejemplo, Sitter y Wu (1993); Torsney y Musrati (1993)), es un experimento de tres puntos, soportados en los $ED_{.10}$, $ED_{.50}$ y $ED_{.90}$.

Para la función link doble recíproco, el experimento D-óptimo de dos puntos está soportado en el $ED_{.50}$ y el $ED_{.8089}$, (o en el $ED_{.1911}$ y el $ED_{.50}$), con determinante de su matriz de información, un 75.3% mayor que el determinante del diseño centrado en el $ED_{.50}$ y simétrico, que Ford, Torsney y Wu (1992) definen en su

artículo, en la Tabla 4, como el mejor diseño de dos puntos para dicho link. El experimento D-óptimo global para este link, es un experimento de tres puntos que se encuentran en el $ED_{.21}$, $ED_{.50}$ y $ED_{.79}$.

En la Figura 3.2 están representados los diseños D-óptimos globales para los cinco links que hemos utilizado. Para el logit, probit y complementario log-log, los diseños coinciden con los de la Tabla 3.1, mientras que para el doble exponencial y doble recíproco los diseños D-óptimos globales están soportados en tres puntos.

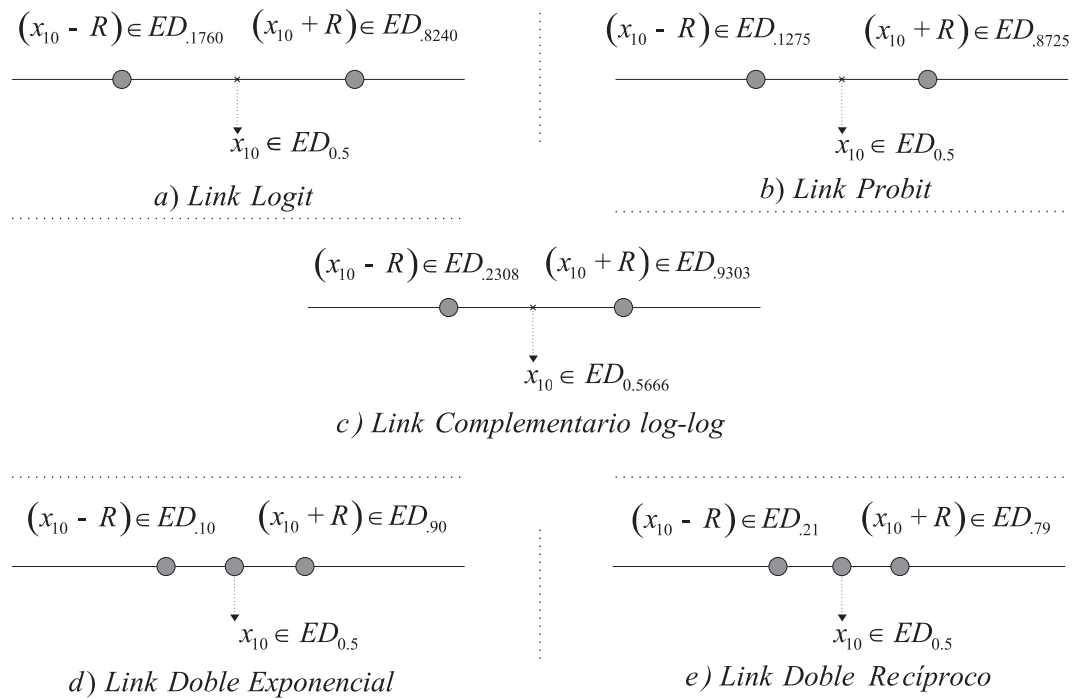


Figura 3.2: Experimentos D-óptimos globales de un factor, para cada uno de los cinco links de un modelo de respuesta binaria de primer orden. Nótese que en función del link el diseño es de dos o tres puntos.

Capítulo 4

Experimentos con dos factores

En este capítulo calculamos tanto los diseños D-óptimos centrados en un cierto $ED_{F(z_0)}$ como los diseños D-óptimos generales dentro de la clase de los experimentos soportados en los vértices de un rectángulo, (x_1, x_2, x_3, x_4) , de forma que $x_i = (x_{10} + a_{i1}R_1, x_{20} + a_{i2}R_2)$, donde a_{i1} y a_{i2} toman valores 1 ó -1 y $x_0 = (x_{10}, x_{20})$ es cualquier punto perteneciente a la línea de dosis efectiva $ED_{F(z_0)}$, con $z_0 = \beta_0 + \beta_1x_{10} + \beta_2x_{20}$. Nótese que en el caso de dos factores las dosis efectivas, o regiones ED , son líneas rectas que vienen definidas por la expresión $z = \beta_0 + \beta_1x_1 + \beta_2x_2$ tal y como se muestra en la Figura 4.1.

Veamos cuál es la expresión del determinante de la matriz de información de Fisher que debemos maximizar para obtener los diseños D-óptimos. De la Proposición 2.2 se deduce que el determinante de $I(\beta)$ para un diseño de dos factores a dos niveles en un modelo de respuesta binaria, viene dado a través del siguiente corolario:

Corolario 4.1. *Bajo el modelo binomial con $p(x_i, \beta) = F(\beta_0 + \beta_1x_{1i} + \beta_2x_{2i})$, el determinante de $I(\beta)$ para cualquier experimento soportado en los vértices de un rectángulo, centrado en la línea $ED_{F(z_0)}$, y con la mitad de los rangos R_1 y R_2 , es de la forma*

$$(\beta_1\beta_2)^2 \det(I(\beta)) = 4^2 n^3 (\beta_1 R_1)^2 (\beta_2 R_2)^2 \sum_{i < j < r} \lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r), \quad (4.1)$$

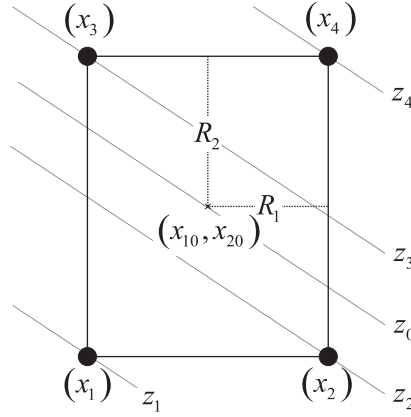


Figura 4.1: Experimento a dos niveles de dos factores.

donde $z_1 = z_0 - |\beta_1|R_1 - |\beta_2|R_2$, $z_2 = z_0 + |\beta_1|R_1 - |\beta_2|R_2$, $z_3 = z_0 - |\beta_1|R_1 + |\beta_2|R_2$ y $z_4 = z_0 + |\beta_1|R_1 + |\beta_2|R_2$ y $h(z_i) = F'(z_i)^2 / (F(z_i)(1 - F(z_i)))$.

Nótese que el determinante de $I(\beta)$ es igual a la suma de los determinantes de las matrices de información de cuatro sub-experimentos obtenidos al quitar cada uno de los cuatro puntos, uno cada vez. Por lo tanto, la información en el experimento es la suma de la información de cada uno de estos cuatro sub-experimentos.

La expresión (4.1) tiene varias particularidades que debemos comentar:

Valor absoluto de los parámetros: Su uso está justificado por las mismas razones expuestas en el caso de un factor.

Valores λ_i : Como ya vimos, es el término que define el porcentaje de observaciones que debemos tomar en cada uno de los puntos del diseño. En el caso balanceado si el diseño está formado por p puntos tendríamos que $\lambda_i = 1/p$. Cuando un diseño es no balanceado los términos λ_i no están definidos a priori y entrarían en el proceso de optimización.

Términos $|\beta_i|R_i$: Se han agrupado los parámetros del modelo con su correspondiente rango, esto es $|\beta_1|R_1$ y $|\beta_2|R_2$. De esta forma se simplifican las operaciones

y se puede maximizar el determinante considerando las expresiones $|\beta_i|R_i$ como variables. Así, una vez definida la recta de dosis efectiva donde queremos centrar el diseño, maximizamos la expresión (4.1) obteniendo como resultado un valor para $|\beta_1|R_1$ y otro para $|\beta_2|R_2$. Con estos valores y una estimación inicial de los parámetros del predictor lineal calculamos los rangos del diseño. Nótese que hemos conseguido reducir el problema en gran medida ya que el resultado de la optimización no depende de los parámetros desconocidos del modelo.

Valor del determinante: Si analizamos la expresión (4.1) podemos observar dos partes diferenciadas. La primera estaría formada por la sumatoria

$$\sum_{i < j < r} \lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r) \quad (4.2)$$

y la segunda por la expresión

$$4^2 n^3 (\beta_1 R_1)^2 (\beta_2 R_2)^2. \quad (4.3)$$

Al aumentar los rangos del diseño, la expresión (4.3) tiende a infinito. Esto nos podría llevar a la conclusión de que el determinante fuese no acotado. Ahora bien, al estudiar la expresión (4.2), y concretamente las funciones $h(\cdot)$, hemos podido constatar que dicha expresión tiende a cero al aumentar los rangos del diseño (salvo en el caso del link doble recíproco o si $\beta_1 = 0$ ó $\beta_2 = 0$). Además, la tendencia de (4.2) a cero es más rápida que la de (4.3) a infinito. Esto hace que una expresión se compense con la otra de lo que se deduce que el determinante máximo de $I(\beta)$ está acotado. Nótese, como dato importante, que Sitter y Torsney (1995a) muestran como el determinante de $I(\beta)$ se puede hacer tan grande como se quiera cuando se sitúan los lados del rectángulo paralelamente a los ED_F 's, y se aumenta el rango del diseño. En este caso, abandonamos el marco de los diseños a dos niveles tal y como los hemos definido en esta memoria y tendríamos que acotar la región de experimentación para obtener el diseño D-óptimo.

Para modelos lineales normales, asignaciones balanceadas, con $\lambda_i = 1/4$, son siempre mejores que asignaciones no balanceadas para el mismo valor de n . En modelos de respuesta binaria, esta propiedad no se verifica. A continuación calculamos (R_1, R_2) , (y los valores de $\lambda_1, \lambda_2, \lambda_3$ y λ_4), que maximizan (4.1) dado un $F(z_0)$, cuando consideramos asignaciones balanceadas y no balanceadas de forma separada.

4.1. Diseños 2^2 balanceados D-óptimos

En la la Tabla 4.1 quedan reflejados tanto los valores de $(|\beta_1|R_1, |\beta_2|R_2)$ como de $(F(z_1), \dots, F(z_4))$, para el experimento D-óptimo que asigna un cuarto de los ensayos a cada uno de los vértices de un rectángulo centrado en la recta $ED_{F(z_0)}$, para algunos valores no negativos de z_0 . Dicha tabla permite obtener los valores localmente D-óptimos de (R_1, R_2) y de $(F(z_1), F(z_2), F(z_3), F(z_4))$ para cualquier valor de $\beta = (\beta_0, \beta_1, \beta_2)$.

Los valores para los z_0 's negativos, pueden ser obtenidos a partir de la Tabla 4.1 por simetría; por ejemplo, bajo el link logit, el experimento D-óptimo centrado en la $ED_{.40}$, tiene $|\beta_i|R_i = 1.9643$, $F(z_1) = 1 - .9871$, $F(z_{2,3}) = .4$ y $F(z_4) = 1 - .0287$.

Nótese que en este caso, la configuración D-óptima se produce cuando $|\beta_1|R_1 = |\beta_2|R_2$, haciendo que $z_2 = z_3 = z_0$, y por lo tanto, dos de los puntos soporte D-óptimos, x_2 y x_3 , están en la misma recta $ED_{F(z_0)}$. Además se cumple que, conforme $F(z_0)$ aumenta, partiendo de 0.5, los valores óptimos de R_1 y R_2 también aumentan. Para el link probit, los valores D-óptimos de R_1 y R_2 son considerablemente mas pequeños, y el determinante de $I(\beta)$ es considerablemente más grande, que para el link logit. Estas dos propiedades son compartidas por todos los experimentos D-óptimos calculados en esta memoria.

Estudiando los resultados para cada z_0 , hemos obtenido los experimentos 2^2 balanceados D-óptimos. En la Tabla 4.2 se muestran los diseños 2^2 balanceados D-óptimos para los link considerados en esta memoria asociados a un modelo de

Link Logit

$F(z_0)$	z_0	$ \beta_1 R_1, \beta_2 R_2$	$F(z_1)$	$F(z_{2,3})$	$F(z_4)$	$(\beta_1\beta_2)^2 \det(I(\beta))$
.50	.000	1.9418	.0202	.50	.9798	.00947 n^3
.55	.201	1.9475	.0243	.55	.9836	.00944 n^3
.60	.405	1.9643	.0287	.60	.9871	.00934 n^3
.65	.619	1.9914	.0334	.65	.9901	.00915 n^3
.70	.847	2.0278	.0388	.70	.9926	.00881 n^3
.75	1.099	2.0737	.0453	.75	.9948	.00826 n^3
.80	1.386	2.1310	.0534	.80	.9965	.00742 n^3
.85	1.735	2.2060	.0643	.85	.9979	.00619 n^3
.90	2.197	2.3150	.0807	.90	.9989	.00448 n^3
.95	2.944	2.5168	.1101	.95	.9997	.00226 n^3

Link Probit

$F(z_0)$	z_0	$ \beta_1 R_1, \beta_2 R_2$	$F(z_1)$	$F(z_{2,3})$	$F(z_4)$	$(\beta_1\beta_2)^2 \det(I(\beta))$
.50	.000	1.0087	.0218	.50	.9782	.03205 n^3
.55	.126	1.0165	.0282	.55	.9846	.03203 n^3
.60	.253	1.0411	.0337	.60	.9902	.03201 n^3
.65	.385	1.0853	.0371	.65	.9947	.03213 n^3
.70	.524	1.1494	.0380	.70	.9976	.03255 n^3
.75	.674	1.2256	.0378	.75	.9991	.03331 n^3
.80	.842	1.3036	.0387	.80	.9997	.03398 n^3
.8130	.889	1.3238	.0393	.8130	.9998	.03402 n^3
.85	1.036	1.3822	.0420	.85	.9999	.03352 n^3
.90	1.282	1.4713	.0483	.90	1.000	.03014 n^3
.95	1.645	1.5992	.0601	.95	1.000	.02067 n^3

Tabla 4.1: Experimentos 2^2 D-óptimos balanceados, con $\lambda_{1,2,3,4} = 1/4$, y centrados en cualquier punto de la recta $ED_{F(z_0)}$, para modelos de primer orden. Los mejores experimentos 2^2 balanceados son aquellos en negrita.

	Link			
	Logit	Probit	D.Exp.	C.log-log
$F(z_0)$	0.5000	0.8130	0.5000	0.8914
z_0	0.0000	0.8890	0.0000	0.7980
$ \beta_i R_i$	1.9418	1.3238	1.9608	2.0386
$F(z_1)$	0.0202	0.0393	0.0099	0.0369
$F(z_{2,3})$	0.5000	0.8130	0.5000	0.8914
$F(z_4)$	0.9798	0.9998	0.9901	1.0000
$(\beta_1\beta_2)^2 \det(I(\beta))$.00947 n^3	.03402 n^3	.07468 n^3	.05751 n^3

Tabla 4.2: Experimentos D-óptimos 2^2 balanceados, con $\lambda_{1,2,3,4} = 1/4$, para modelos de primer orden. Para el link doble recíproco, el determinante de $I(\beta)$ no está acotado superiormente.

respuesta binaria. En el caso del modelo logístico se encuentra centrado en el $ED_{.50}$ y tiene dos puntos soporte, uno en el $ED_{.9798}$ y otro en el $ED_{.0202}$. Si por ejemplo asumimos que $\beta_1 = 2$ y $\beta_2 = 3$, entonces el experimento 2^2 localmente D-óptimo es tal que $R_1 = .9709$, $R_2 = .6473$ y donde el determinante de su $I(\beta)$ es $2.63 \cdot 10^{-4}n^3$. Para el caso del link doble exponencial, el experimento D-óptimo 2^2 balanceado también se encuentra centrado y tiene dos puntos soporte en el $ED_{.50}$.

Por otro lado, bajo el modelo probit el mejor experimento 2^2 balanceado está centrado en el $ED_{.8130}$, y tiene dos puntos soporte sobre las rectas $ED_{.9998}$ y $ED_{.0393}$; si $\beta_1 = 2$ y $\beta_2 = 3$, entonces $R_1 = .7619$, $R_2 = .4412$ y el determinante de su $I(\beta)$ es $9.45 \cdot 10^{-4}n^3$. Para el link complementario log-log, el mejor experimento 2^2 balanceado es muy similar al que se obtuvo para el link probit.

Por último, si consideramos el link doble recíproco, el determinante de $I(\beta)$ puede hacerse tan grande como se quiera al aumentar los valores R_i , $i = 1, 2$, y de esta manera, si no se consideran restricciones sobre la región experimental, no existe el diseño 2^2 D-óptimo balanceado para este link.

En la columna de la izquierda de la Figura 4.2 están representados a escala los diseños 2^2 balanceados D-óptimos para los distintos link de un modelo de respuesta binaria con dos factores sin interacciones.

4.2. Diseños D-óptimos a dos niveles

Considerando asignaciones no balanceadas, dado un z_0 , debemos maximizar la expresión (4.1) respecto de $(|\beta_1|R_1, |\beta_2|R_2)$ y de $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. La Tabla 4.3 caracteriza a los experimentos D-óptimos entre todos aquellos soportados en los vértices de los rectángulos centrados en el $ED_{F(z_0)}$, para varios z_0 's. Los valores correspondientes a $z_0 < 0$ pueden, una vez más, ser calculados por simetría.

Bajo el modelo logístico, si $.5 < F(z_0) < .51593$ el experimento D-óptimo es un diseño no balanceado de cuatro puntos y cuando $F(z_0) > .51593$, es un diseño balanceado de tres puntos. De forma similar, bajo el modelo probit, si $.5 < F(z_0) < .5297$ el experimento D-óptimo es un diseño no balanceado de cuatro puntos, y si $F(z_0) > .5297$ se convierte en un diseño balanceado de tres puntos. Para estos dos links, la configuración D-óptima se alcanza cuando $|\beta_1|R_1 = |\beta_2|R_2$, y, por consiguiente, $z_2 = z_3 = z_0$. Partiendo de $F(z_0) = 0.5$, conforme $F(z_0)$ aumenta, también lo harán los valores D-óptimos de R_1 y R_2 .

Llevando a cabo un estudio de los valores óptimos obtenidos para cada uno de los z_0 's, en la Tabla 4.3 se observa que el mejor experimento a dos niveles bajo el modelo logístico, es un diseño balanceado de tres puntos con dos puntos soporte sobre la recta $ED_{.7320}$ y el otro sobre la recta $ED_{.0355}$. Si por ejemplo, $\beta_1 = 2$ y $\beta_2 = 3$, el diseño D-óptimo es tal que $R_1 = 1.0765$, $R_2 = .7177$ y el determinante de su $I(\beta)$ es $4.665 \cdot 10^{-4}n^3$. Nótese como este determinante es dos veces el determinante para el mejor diseño D-óptimo en el caso balanceado considerado en la Tabla 4.1. Además, el diseño D-óptimo soportado en los vértices de un rectángulo centrado en la dosis efectiva del 50%, es un diseño de cuatro puntos que sitúa $.328n$ ensayos en dos puntos del $ED_{.50}$, y $.172n$ en el $ED_{.0179}$ y $ED_{.9821}$. Por ejemplo, si $\beta_1 = 2$ y $\beta_2 = 3$, el experimento localmente D-óptimo es tal que $R_1 = 1.0013$, $R_2 = .6676$ y el determinante de la matriz de información es $3 \cdot 10^{-4}n^3$, que es al menos un 15% mayor que el obtenido para el mejor experimento 2^2 centrado en el $ED_{.50}$.

En el caso del link probit, el experimento D-óptimo a dos niveles está centrado y tiene dos puntos soporte en la dosis efectiva del 83.18% y otro en la del 3.79%. El

determinante de la matriz de información para este diseño es dos veces el obtenido para el mejor experimento 2^2 balanceado y cinco veces más grande que el mejor diseño a dos niveles para el modelo logístico. Por otro lado, el diseño localmente D-óptimo centrado en la dosis efectiva del 50% para el link probit, asigna $.3187n$ ensayos a dos puntos situados en la dosis del 50%, y $.182n$ ensayos en las dosis del 1.88% y del 98.11%. El determinante de su matriz de información es alrededor de un 10% mayor que el determinante de $I(\beta)$ para el mejor diseño 2^2 balanceado centrado en dicho z_0 .

En la Tabla 4.4 están recogidos los mejores diseños a dos niveles para los links utilizados en esta tesis para un modelo de datos binarios con predictor lineal de primer orden de dos factores. El mejor diseño para el link logit y para el probit ya han sido comentados en los párrafos anteriores. Para el caso del doble exponencial, el experimento D-óptimo a dos niveles es un diseño de tres puntos, que está centrado y tiene dos puntos soporte en la línea del $ED_{.50}$ y un tercero sobre el $ED_{.0097}$. El determinante de su matriz de información es un 10% mayor que el obtenido para el diseño 2^2 balanceado. Para el link complementario log-log, el mejor diseño a dos niveles tiene dos puntos soporte en el $ED_{.8914}$ y otro en el $ED_{.0369}$. El determinante de su matriz de información es dos veces mayor que el determinante del mejor diseño 2^2 balanceado. Para el link doble recíproco, no hay una solución D-óptima finita.

En la Figura 4.2 aparecen representados los diseños D-óptimos para el caso balanceado y no balanceado, cuando se considera un diseño de dos factores a dos niveles, para cada uno de los link considerados de un modelo de respuesta binaria de primer orden, salvo para el doble recíproco que no tiene solución en el caso no balanceado.

Link Logit

$F(z_0)$	$ \beta_{1,2} R_{1,2}$	$F(z_1)$	$F(z_{2,3})$	$F(z_4)$	λ_1	$\lambda_{2,3}$	λ_4	$(\beta_1\beta_2)^2 \det(I(\beta))$
.50	2.0027	.0179	.50	.9821	.172	.328	.172	.0108 n^3
.51	2.0258	.0178	.51	.9836	.264	.330	.076	.0110 n^3
.55	2.0781	.0188	.55	.9873	.333	.333	.000	.0125 n^3
.60	2.0933	.0223	.60	.9900	.333	.333	.000	.0143 n^3
.65	2.1118	.0265	.65	.9922	.333	.333	.000	.0157 n^3
.70	2.1349	.0316	.70	.9940	.333	.333	.000	.0166 n^3
.7320	2.1530	.0355	.7320	.9951	.333	.333	.000	.01689 n^3
.75	2.1646	.0380	.75	.9956	.333	.333	.000	.0167 n^3
.80	2.2046	.0464	.80	.9970	.333	.333	.000	.0159 n^3
.85	2.2619	.0579	.85	.9981	.333	.333	.000	.0138 n^3
.90	2.3537	.0751	.90	.9990	.333	.333	.000	.0102 n^3
.95	2.5382	.1060	.95	.9997	.333	.333	.000	.0053 n^3

Link Probit

$F(z_0)$	$ \beta_{1,2} R_{1,2}$	$F(z_1)$	$F(z_{2,3})$	$F(z_4)$	λ_1	$\lambda_{2,3}$	λ_4	$(\beta_1\beta_2)^2 \det(I(\beta))$
.50	1.0390	.0189	.50	.9811	.182	.318	.182	.0352 n^3
.51	1.0435	.0196	.51	.9827	.221	.319	.141	.0353 n^3
.52	1.0595	.0193	.52	.9849	.264	.323	.090	.0358 n^3
.53	1.1028	.0166	.53	.9887	.333	.333	.000	.0369 n^3
.55	1.1168	.0175	.55	.9908	.333	.333	.000	.0402 n^3
.60	1.1529	.0201	.60	.9947	.333	.333	.000	.0485 n^3
.65	1.1911	.0229	.65	.9972	.333	.333	.000	.0572 n^3
.70	1.2324	.0262	.70	.9986	.333	.333	.000	.0655 n^3
.75	1.2782	.0299	.75	.9994	.333	.333	.000	.0729 n^3
.80	1.3305	.0344	.80	.9998	.333	.333	.000	.0779 n^3
.8318	1.3689	.0379	.8318	.9999	.333	.333	.000	.07902 n^3
.85	1.3930	.0401	.85	.9999	.333	.333	.000	.0786 n^3
.90	1.4743	.0478	.90	1.000	.333	.333	.000	.0713 n^3
.95	1.5990	.0601	.95	1.000	.333	.333	.000	.0490 n^3

Tabla 4.3: Experimentos localmente D-óptimos entre todos los experimentos de dos factores soportados en los vértices de un rectángulo cuyos lados son paralelos a los ejes cartesianos, y centrado en cualquier punto de la recta $ED_{F(z_0)}$, para modelos de primer orden. Los mejores experimentos de este tipo están en negrita.

	Link			
	Logit	Probit	D.Exp.	C.log-log
$F(z_0)$	0,732	0,8318	0,50	0,8914
$F(z_0)$	0.732	0.8318	0.50	0.8914
$ \beta_i R_i$	2.153	1.3689	1.9709	2.0386
$F(z_1)$	0.0355	0.0379	0.0097	0.0369
$F(z_{2,3})$	0.7320	0.8318	0.50	0.8914
$F(z_4)$	0.9951	0.9999	0.9903	1.000
λ_1	0.333	0.333	0.1675	0.333
$\lambda_{2,3}$	0.333	0.333	0.3325	0.333
λ_4	.000	.000	0.1675	.000
$(\beta_1\beta_2)^2 \det(I(\beta))$.01689 n^3	.07902 n^3	.08807 n^3	.13631 n^3

Tabla 4.4: Experimentos D-óptimos de dos factores soportados en los vértices de un rectángulo orientado según los ejes cartesianos. Para el link doble recíproco, el determinante de $I(\beta)$ aumenta cuando aumentan R_1 y R_2 .

4.3. Diseños factoriales D-óptimos cuando hay restricciones en los factores.

Dentro de los modelos de respuesta binaria, el modelo logístico es el más utilizado. Por otro lado, los diseños más comunes en la práctica son los diseños de dos factores a dos niveles. Combinando ambos aspectos, hemos querido ampliar el estudio de los diseños D-óptimos a dos niveles, al caso donde restringimos la región de experimentación.

En modelos lineales normales cuando acotábamos la región de experimentación se definía para cada factor un intervalo donde éste tomaba sus valores. El mejor diseño era un diseño factorial de cuatro puntos situados en los vértices de la región acotada. Vamos a ver que, en general, esta propiedad no se verifica en el caso de modelos de respuesta binaria. Para comprobarlo, hemos trabajado con el diseño D-óptimo balanceado de dos factores a dos niveles para el modelo logístico que se encuentra centrado en la recta $ED_{.50}$ (ver Tabla 4.1). A partir de este diseño se nos podrían plantear algunos problemas prácticos tales como ¿qué ocurriría si no pudiésemos observar en los puntos del diseño? Un ejemplo sería el caso donde

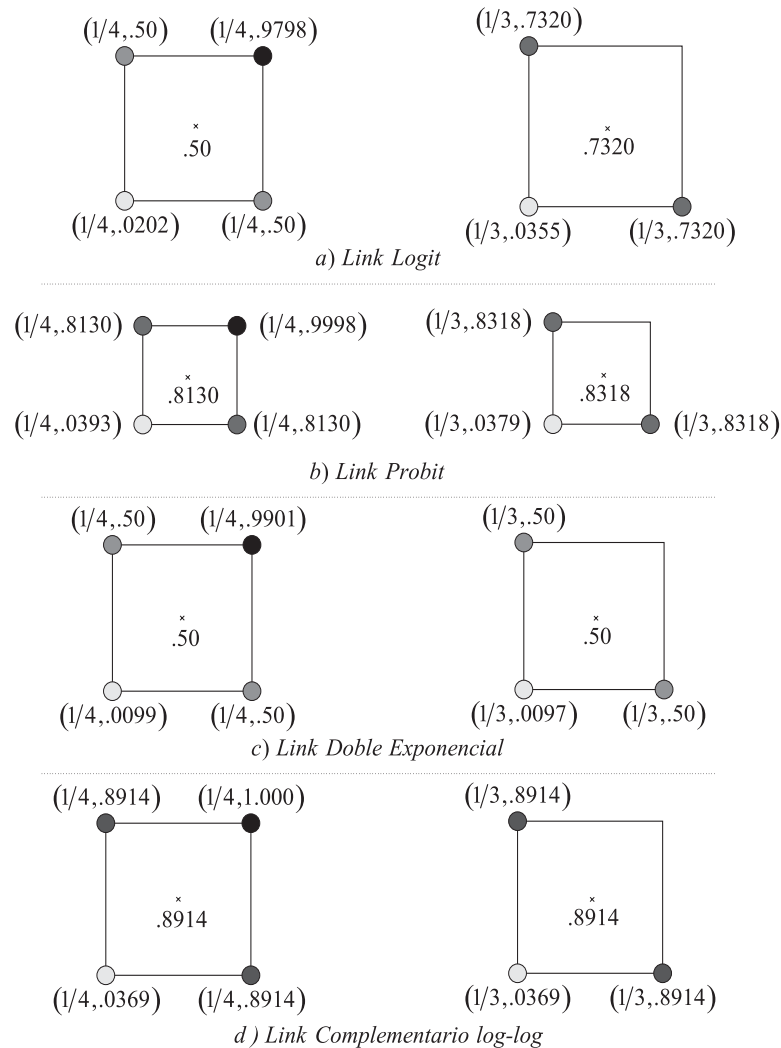


Figura 4.2: Experimentos 2^2 D-óptimos para el caso balanceado (columna de la izquierda) y no balanceado (columna de la derecha). Los valores entre paréntesis indican $(\lambda_i, F(z_i))$ asociados a cada punto.

medimos el porcentaje de fallecidos por inhalación de ciertos componentes químicos. Como podemos observar en la Tabla 4.1 el diseño D-óptimo toma el 25% de sus observaciones en las dosis que produce un 97.98% de fallecidos. Esto hace que el experimento sea totalmente inviable. Para resolver este problema, debemos definir intervalos donde se moverán los factores considerados. Para dos factores se considera una restricción rectangular cuyos lados son paralelos a los ejes cartesianos. Las restricciones que consideraron Sitter y Torsney (1995a) eran sobre el espacio de las líneas ED justificadas por cuestiones prácticas y teóricas como vimos en la Sección

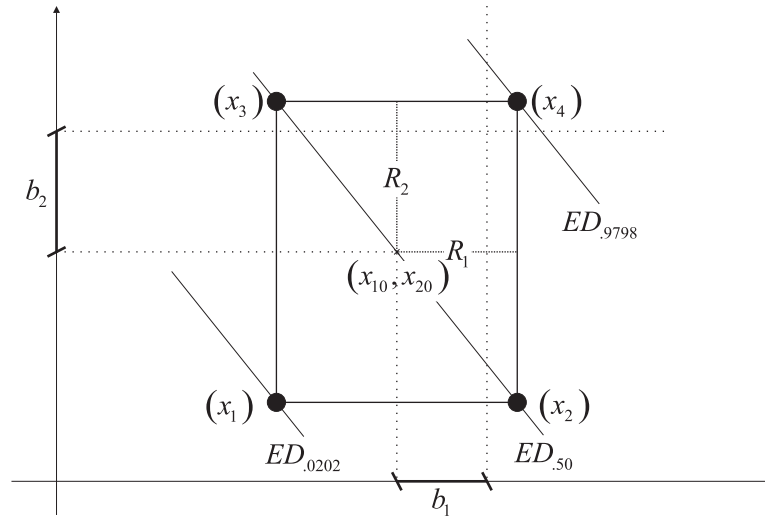


Figura 4.3: Experimento 2^2 D-óptimo balanceado para el modelo logístico. b_1 y b_2 definen las restricciones en R_1 y R_2 impuestas en la Sección 4.3.

2.5.

Considerando restricciones sobre los factores de la forma que hemos comentado en el párrafo anterior, el diseño factorial D-óptimo, que se encuentra centrado en el $ED_{.50}$, vendrá definido por los valores R_1 y R_2 óptimos, de forma que en el proceso de optimización se imponen las restricciones $R_1 \in [0, b_1]$ y $R_2 \in [0, b_2]$. En la Figura 4.3 se muestra el diseño de dos factores a dos niveles D-óptimo para el modelo logístico y las restricciones sobre R_1 y R_2 . Bajo estas condiciones se consideran los siguientes casos:

- **No existen restricciones:** En este caso, haciendo uso de la Tabla 4.1, los rangos del diseño a dos niveles son tales que

$$R_1 = \frac{1.9418}{|\beta_1|}, \text{ y } R_2 = \frac{1.9418}{|\beta_2|}$$

- **Restricción sobre una variable regresora:** Definimos la función $f(b)$, evaluada en la Tabla 4.5, como una función que nos proporciona el valor óptimo de la mitad del lado de un paralelogramo, i.e. el rango de un factor, cuando

el otro toma valores menores o iguales a b . Así obtenemos un paralelogramo de rangos b y $f(b)$ respectivamente. La función sólo está evaluada para valores menores que $b = 1.94183$ puesto que para valores superiores, el determinante comienza a disminuir, esto es, el diseño no es óptimo. Cuando la restricción supere el valor $1.94183/|\beta_i|$, siendo x_i la variable sobre la que existe la restricción, el problema se reduce al caso en el que no hay restricciones cubierto en la Sección 4.1. Por este motivo, nos limitamos a la situación donde la restricción correspondiente, sea menor que dicha cantidad.

a) Restricción sobre x_1 ; $R_1 \in [0, b_1]$.

Si $b_1 < 1.94183/|\beta_1|$ entonces $R_1 = b_1$ y $R_2 = f(|\beta_1|b_1)/|\beta_2|$

b) Restricción sobre x_2 ; $R_2 \in [0, b_2]$.

Si $b_2 < 1.94183/|\beta_2|$ entonces $R_2 = b_2$ y $R_1 = f(|\beta_2|b_2)/|\beta_1|$

- **Restricción sobre las dos variables regresoras:** Nos restringimos al caso donde las restricciones verifican que $b_1 < 1.94183/|\beta_1|$ y $b_2 < 1.94183/|\beta_2|$ respectivamente. En cualquier otra circunstancia, podemos recurrir a las situaciones anteriores, sin restricciones o con una sola restricción, asumiendo que si $b_1 \geq 1.94183/|\beta_1|$ y/o $b_2 \geq 1.94183/|\beta_2|$ no tendríamos restricción sobre la o las correspondientes componentes, respectivamente.

Los rangos del paralelogramo D-óptimo dependen, en este caso, de si $b_i < 1.2229/|\beta_i|$ o $b_i > 1.2229/|\beta_i|$, con $i = 1, 2$. Vamos a considerar el caso de b_1 . Para b_2 se haría de forma análoga simplemente sustituyendo el subíndice 1 por 2.

a) Si $b_1 < 1.2229/|\beta_1|$ entonces: $R_1 = b_1$ y $R_2 = b_2$ si $b_2 < f(|\beta_1|b_1)/|\beta_2|$
ó $R_2 = f(|\beta_1|b_1)/|\beta_2|$ si $b_2 \geq f(|\beta_1|b_1)/|\beta_2|$.

b) Si $b_1 > 1.2229/|\beta_1|$ entonces:

○ Si $b_2 < f^{-1}(|\beta_1|b_1)/|\beta_2|$ entonces $R_2 = b_2$ y $R_1 = f(|\beta_2|b_2)/|\beta_1|$

○ Si $f^{-1}(|\beta_1|b_1)/|\beta_2| \leq b_2 < f(|\beta_1|b_1)/|\beta_2|$ entonces $R_2 = b_2$ y $R_1 = b_1$.

b	$f(b)$	z_1	$F(z_1)$	$det(I(\beta))$
0.0001	1.2229	1.2230	22.74	0.000000 n^3
0.1000	1.2249	1.3249	21.00	0.020688 n^3
0.2000	1.2309	1.4309	19.30	0.081821 n^3
0.3000	1.2410	1.5410	17.64	0.180676 n^3
0.4000	1.2550	1.6550	16.04	0.312941 n^3
0.5000	1.2731	1.7731	14.52	0.473026 n^3
0.6000	1.2953	1.8953	13.06	0.654452 n^3
0.7000	1.3214	2.0214	11.70	0.850264 n^3
0.8000	1.3516	2.1516	10.42	1.053446 n^3
0.9000	1.3858	2.2858	9.23	1.257279 n^3
1.0000	1.4239	2.4239	8.14	1.455643 n^3
1.1000	1.4658	2.5658	7.14	1.643223 n^3
1.2000	1.5112	2.7112	6.23	1.815635 n^3
1.3000	1.5602	2.8602	5.42	1.969475 n^3
1.4000	1.6124	3.0124	4.69	2.102297 n^3
1.5000	1.6676	3.1676	4.04	2.212553 n^3
1.6000	1.7257	3.3257	3.47	2.299497 n^3
1.7000	1.7862	3.4862	2.97	2.363078 n^3
1.8000	1.8491	3.6491	2.54	2.403824 n^3
1.9000	1.9141	3.8141	2.16	2.422738 n^3
1.9418	1.9418	3.8836	2.02	2.424491 n^3

Tabla 4.5: Valores de $f(c)$, z_1 , $F(z_1)$ y $det(I(\beta))$ para el diseño factorial 2^2 D-óptimo centrado en el $ED_{.50}$, bajo el modelo de regresión logístico y para un valor c dado (la mitad del lado del paralelogramo en una de las dos componentes) con $f(c)$ el valor óptimo de la mitad del lado del paralelogramo en la otra componente.

○ Si $f(|\beta_1|b_1)/|\beta_2| \leq b_2$ entonces $R_1 = b_1$ y $R_2 = f(|\beta_1|b_1)/|\beta_2|$.

Para ilustrar el uso de la Tabla 4.5 para la obtención de los diseños de dos factores a dos niveles con restricciones, se ha considerado un modelo de parámetros $\beta_1 = 1$, $\beta_2 = 1$ y $\beta_0 = 0$, siendo $z = 1x_1 + 1x_2$ el predictor lineal asociado. Sin pérdida de generalidad consideramos el diseño centrado en el punto $(0, 0)$. Bajo estos supuestos se consideran las siguientes situaciones:

Sin restricciones:

1. No existen restricciones sobre x_1 ni x_2 . El paralelogramo óptimo observará en

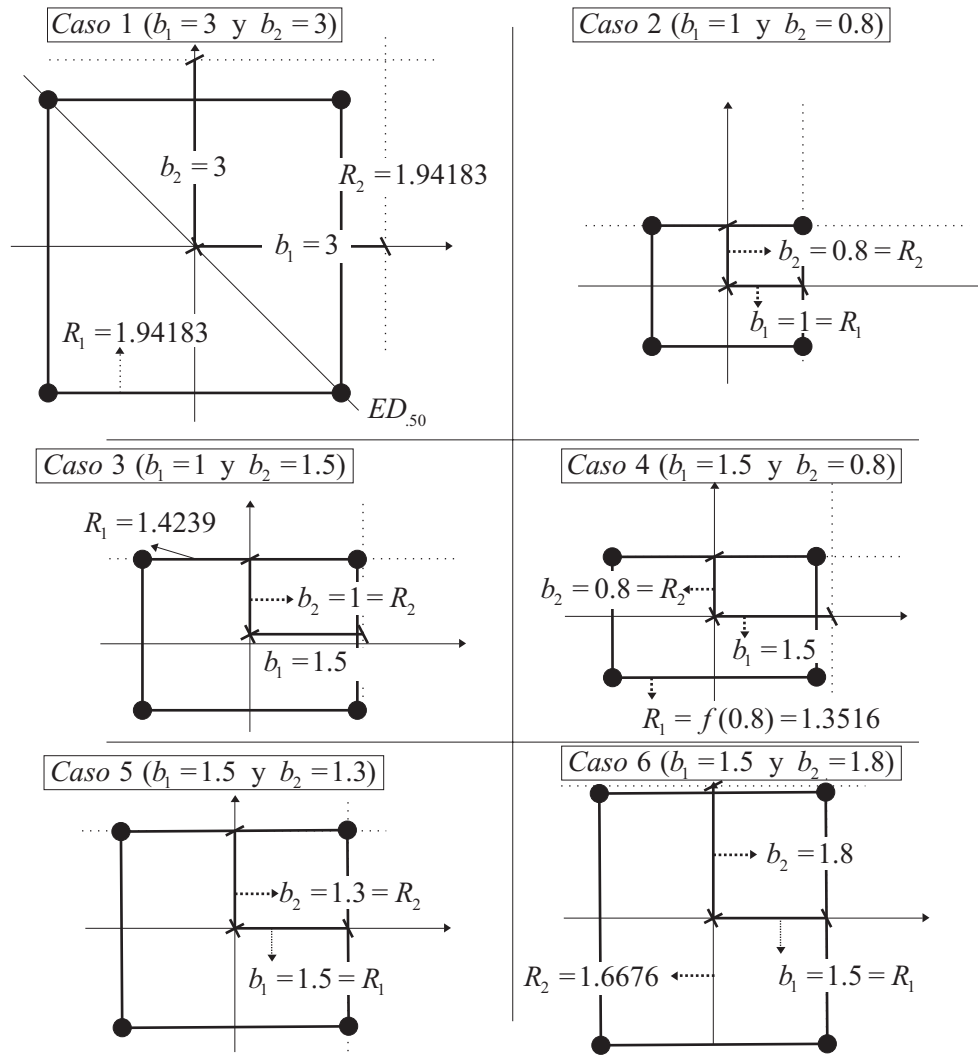


Figura 4.4: Diseños D-óptimos 2^2 para un modelo logístico con predictor lineal $z = 1x_1 + 1x_2$ y restricciones b_1 y b_2 en los factores. Los seis casos se corresponden a todas las posibilidades que se podrían dar en un problema real.

los puntos $(\pm 1.94183; \pm 1.94183)$.

Una restricción:

1. No hay restricción sobre x_2 y $b_1 = 3$, i.e. $x_1 \in [-3, 3]$. El paralelogramo óptimo observará en los puntos $(\pm 1.94183; \pm 1.94183)$. Este diseño coincide con el diseño factorial D-óptimo obtenido para el caso sin restricciones.
2. No hay restricción sobre x_2 y $b_1 = 1$, i.e. $x_1 \in [-1, 1]$. El paralelogramo óptimo observará en los puntos $(\pm 1.0; \pm f(1)) = (\pm 1.0; \pm 1.4239)$. Se trata del caso de

una restricción con $b_1 < 1.94183/|\beta_1| = 1.94183$.

Dos restricciones:

1. $b_1 = b_2 = 2.5 \Rightarrow x_1 \in [-2.5, 2.5]$ y $x_2 \in [-2.5, 2.5]$. El paralelogramo óptimo observará en los puntos $(\pm 1.94183; \pm 1.94183)$, igual que en el caso sin restricciones.
2. $b_1 = 1$ y $b_2 = 0.8 \Rightarrow x_1 \in [-1, 1]$ y $x_2 \in [-0.8, 0.8]$. El paralelogramo óptimo observará en los puntos $(\pm 1; \pm 0.8)$, puesto que $0.8 < f(1) = 1.4239$ ya que nos encontramos en el caso de $b_1 < 1.2229/|\beta_1| = 1.2229$ y $b_2 < f(|\beta_1|b_1)/|\beta_2| = 1.4239$.
3. $b_1 = 1$ y $b_2 = 1.5 \Rightarrow x_1 \in [-1, 1]$ y $x_2 \in [-1.5, 1.5]$. El paralelogramo óptimo observará en los puntos $(\pm 1; \pm 1.4239)$, puesto que $1.5 > f(1) = 1.4239$ y $b_1 < 1.2229/|\beta_1| = 1.2229$.
4. $b_1 = 1.5$ y $b_2 = 0.8 \Rightarrow x_1 \in [-1.5, 1.5]$ y $x_2 \in [-0.8, 0.8]$. El paralelogramo óptimo observará en los puntos $(\pm f(0.8); \pm 0.8)$, puesto que $0.8 < f^{-1}(1.5) = 1.17$ y $b_1 > 1.2229/|\beta_1| = 1.2229$. En este caso el valor de R_1 óptimo no es constante para todo valor de R_2 tal y como ocurría cuando $b_1 < 1.2229/|\beta_1| = 1.2229$. Esta es la causa por la cual debemos considerar las dos situaciones por separado.
5. $b_1 = 1.5$ y $b_2 = 1.3 \Rightarrow x_1 \in [-1.5, 1.5]$ y $x_2 \in [-1.3, 1.3]$. El paralelogramo óptimo observará en los puntos $(\pm 1.5, \pm 1.3)$, puesto que $f^{-1}(1.5) = 1.17 < 1.3 < f(1.5) = 1.6676$ y $b_1 > 1.2229/|\beta_1| = 1.2229$.
6. $b_1 = 1.5$ y $b_2 = 1.8 \Rightarrow x_1 \in [-1.5, 1.5]$ y $x_2 \in [-1.8, 1.8]$. El paralelogramo óptimo observará en los puntos $(\pm 1.5, \pm 1.6676)$, puesto que $f(1.5) = 1.6676 < 1.8$ y $b_1 > 1.2229/|\beta_1| = 1.2229$.

Los diseños a dos niveles de la Figura 4.4 representan a cada uno de los seis casos de dos restricciones.

4.4. Conclusiones

Se han estudiado los diseños D-óptimos para dos factores a dos niveles considerando asignaciones balanceadas y no balanceadas para un modelo de respuesta binaria sin interacciones. Si comparamos el caso balanceado y no balanceado, concluimos que los diseños D-óptimos de dos factores a dos niveles, para todas las funciones link descritas en esta memoria, son no balanceados, de tal forma que, en el vértice perteneciente al $ED_{F(z_4)}$, con $z_4 = z_0 + |\beta_1|R_1 + |\beta_2|R_2$ (el más alejado del $ED_{.50}$), no tiene asignada ninguna observación. Esto es, los diseños en los que se modifican los factores uno a uno, diseños de la columna derecha en la Figura 4.2, son más informativos que los mejores diseños factoriales a dos niveles balanceados, columna de la izquierda. De este modo, el hecho que para el modelo normal los diseños factoriales son mejores que los experimentos que varían sólo un factor a la vez, no se verifica en el modelo de respuesta binaria.

En cuanto al caso con restricciones nos hemos ceñido a los diseños D-óptimos balanceados para un modelo de respuesta binaria con dos factores y link logit. Los rangos de los factores se han obtenido de forma numérica y están recogidos en la Tabla 4.5. Mediante esta tabla, y conociendo las restricciones, se pueden determinar de forma rápida y sencilla los diseños D-óptimos de dos factores a dos niveles. La conclusión más importante que se puede extraer es que para modelos de respuesta binaria con restricciones sobre la region experimental, el diseño D-óptimo no es aquel que observa en los vértices de las restricciones. Esto contradice la idea heredada de los modelos lineales normales donde siempre se muestreaba en los vértices de las restricciones del espacio experimental.

Capítulo 5

Experimentos con 3 factores

En este capítulo se calculan los experimentos D-óptimos, dentro de la clase de experimentos soportados en los ocho vértices de un hiper-rectángulo en \mathbb{R}^3 , (x_1, \dots, x_8) , tal y como se muestra en la Figura 5.1, de forma que $x_i = (x_{10} + a_{i1}R_1, x_{20} + a_{i2}R_2, x_{30} + a_{i3}R_3)$ donde a_{i1}, a_{i2} y a_{i3} son iguales a -1 ó 1 y donde $x_0 = (x_{10}, x_{20}, x_{30})$ es cualquier punto sobre el plano $ED_{F(z_0)}$, con $z_0 = \beta_0 + \beta_1x_{10} + \beta_2x_{20} + \beta_3x_{30}$ elegido a priori.

Sean los planos $ED_{F(z_i)}$ donde se apoyan los puntos soporte del experimento, con $z_1 = z_0 - |\beta_1|R_1 - |\beta_2|R_2 - |\beta_3|R_3$, $z_2 = z_0 + |\beta_1|R_1 - |\beta_2|R_2 - |\beta_3|R_3$, $z_3 = z_0 - |\beta_1|R_1 + |\beta_2|R_2 - |\beta_3|R_3$, $z_4 = z_0 + |\beta_1|R_1 + |\beta_2|R_2 - |\beta_3|R_3$, $z_5 = z_0 - |\beta_1|R_1 - |\beta_2|R_2 + |\beta_3|R_3$, $z_6 = z_0 + |\beta_1|R_1 - |\beta_2|R_2 + |\beta_3|R_3$, $z_7 = z_0 - |\beta_1|R_1 + |\beta_2|R_2 + |\beta_3|R_3$ y $z_8 = z_0 + |\beta_1|R_1 + |\beta_2|R_2 + |\beta_3|R_3$. En la Figura 5.1 aparece representado el plano $ED_{F(z_8)}$. La expresión para el determinante de la matriz de información $I(\beta)$ se deduce de la Proposición 2.2 a través del siguiente corolario:

Corolario 5.1. *Bajo el modelo binomial con $p(x_i, \beta) = F(\beta_0 + \beta_1x_{1i} + \beta_2x_{2i} + \beta_3x_{3i})$, el determinante de $I(\beta)$ para cualquier experimento soportado en los vértices de un hiper-rectángulo en \mathbb{R}^3 , y cuya estructura ha sido explicada en el párrafo anterior, es tal que*

$$(\beta_1\beta_2\beta_3)^2 \det(I(\beta)) =$$

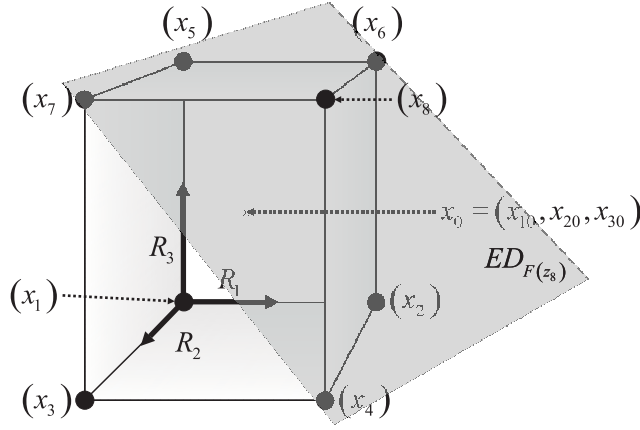


Figura 5.1: Experimento a dos niveles de tres factores. El plano denotado por $ED_{F(z_8)}$ es el plano de dosis efectiva que contiene al punto x_8 . Por cada punto (x_i) habría un plano paralelo al $ED_{F(z_8)}$ que determinaría un plano de igual probabilidad.

$$n^4(\beta_1 R_1)^2(\beta_2 R_2)^2(\beta_3 R_3)^2 \sum_{\{i_1, i_2, i_3, i_4\}} c_{i_1, i_2, i_3, i_4} \lambda_{i_1} \lambda_{i_2} \lambda_{i_3} \lambda_{i_4} h(z_{i_1}) h(z_{i_2}) h(z_{i_3}) h(z_{i_4}), \quad (5.1)$$

donde la sumatoria considera las 70 posibles combinaciones de (i_1, i_2, i_3, i_4) obtenidas como subconjuntos de $\{1, 2, 3, 4, 5, 6, 7, 8\}$ de tamaño 4, y donde $c_{1,4,6,7} = c_{2,3,5,8} = 4^4$, $c_{1,2,3,4} = c_{1,2,5,6} = c_{1,2,7,8} = c_{1,3,5,7} = c_{1,3,6,8} = c_{1,4,5,8} = c_{2,3,6,7} = c_{2,4,5,7} = c_{2,4,6,8} = c_{3,4,5,6} = c_{3,4,7,8} = c_{5,6,7,8} = 0$, y $c_{i_1, i_2, i_3, i_4} = 4^3$ para el resto de los 56 términos.

La expresión (5.1) es el determinante de la matriz de información para el experimento a dos niveles de tres factores para un modelo de respuesta binaria. Nótese que dicho determinante es, a su vez, la suma de los determinantes para cada uno de los setenta experimentos fraccionales un medio que pueden construirse a partir de un experimento factorial completo de ocho puntos. Esto es, la información contenida en este diseño medida a través del determinante de $I(\beta)$, es igual a la suma de la información contenida en los diseños fraccionales que se pueden obtener a partir del

diseño inicial.

Los subexperimentos (1, 4, 6, 7) y (2, 3, 5, 8), cuyas matrices de diseño ortogonales estandarizadas son $A_{1,4,6,7}$ y $A_{2,3,5,8}$, se corresponden con los diseños factoriales fraccionales 2^{3-1} D-óptimos para el caso del modelo lineal normal, que llamaremos a partir de este punto, diseños fraccionales estándar 2^{3-1} . El estudio de estos subexperimentos será abordado en la Sección 5.3, así como en el Apéndice B. Por otro lado, las fracciones un medio del diseño factorial de ocho puntos que tienen asociado un valor de c_{i_1, i_2, i_3, i_4} igual a 0, corresponden a las fracciones que no permiten estimar algunos de los parámetros β_i 's del modelo.

Como ya se ve el caso de tres factores es una generalización del de dos factores. Las propiedades que se daban en el Capítulo 4 también se pueden aplicar aquí con pequeñas modificaciones. Así, tratamos de encontrar los valores de (R_1, R_2, R_3) (y los λ_i 's) que maximicen el determinante de $I(\beta)$ para cualquier z_0 , distinguiendo el caso de asignaciones balanceadas y no balanceadas.

5.1. Diseños 2^3 balanceados D-óptimos

Las Tablas 5.2 y 5.3 nos proporcionan los valores de $(|\beta_1|R_1, |\beta_2|R_2, |\beta_3|R_3)$ y de $(F(z_1), \dots, F(z_8))$ para el experimento D-óptimo entre todos los diseños que asignan una octava parte de las pruebas a cada uno de los vértices del hiper-rectángulo centrado en el plano $ED_{F(z_0)}$, para varios z_0 's. La Tabla 5.2 nos proporciona los valores D-óptimos cuando se hace uso del link logit mientras que la Tabla 5.3 lo hace para el link probit. Esto nos permite obtener el diseño D-óptimo local a través de los valores de (R_1, R_2, R_3) y (z_1, z_2, \dots, z_8) para cualquier $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$.

Bajo el link logístico y dado un z_0 , la configuración D-óptima es tal que $|\beta_1|R_1 = |\beta_2|R_2 = |\beta_3|R_3$. Esto implica que $z_2 = z_3 = z_5$ y que $z_4 = z_6 = z_7$, y, por lo tanto, sus ocho puntos soporte caen exactamente en cuatro $ED_{F(z)}$'s diferentes.

Por otro lado, para el link probit la configuración D-óptima se obtiene para $|\beta_1|R_1 = |\beta_2|R_2 = |\beta_3|R_3$ únicamente en los rangos $.5 < F(z_0) < .8057$ y $.9821 <$

$F(z_0)$. Cuando el punto central es tal que $.8057 < F(z_0) < .9821$, una de las tres configuraciones D-óptimas equivalentes cumple que $|\beta_1|R_1 = |\beta_2|R_2$, con $|\beta_3|R_3$ diferente a las otras dos. De esta forma, únicamente $z_2 = z_3$ y $z_6 = z_7$. Las otras dos configuraciones D-óptimas para $.8057 < F(z_0) < .9821$ se obtienen simplemente variando los subíndices de forma que los términos de subíndices 2 y 3 sean iguales y el 1 distinto, o el 1 y 3 iguales y el 2 distinto.

Bajo el modelo logístico, el diseño 2^3 balanceado con mayor determinante de $I(\beta)$ está centrado en el plano $ED_{.9053}$ y tiene tres puntos soporte en el $ED_{.3561}$ y en el $ED_{.9940}$, y un punto soporte en el $ED_{.0018}$ y otro en el $ED_{1.000}$; si $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_3 = 1$, el experimento 2^3 D-óptimo local es tal que $R_1 = 1.4250$, $R_2 = .9500$, $R_3 = 2.8500$, y el determinante de su $I(\beta)$ es $5.918 \cdot 10^{-5}n^4$.

Para el modelo probit, el mejor diseño balanceado 2^3 está centrado en el plano $ED_{.50}$ y tiene tres puntos soporte en el $ED_{.0851}$ y en el $ED_{.9149}$, y un punto soporte en el $ED_{.0000}$ y el $ED_{1.000}$; si $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_3 = 1$, entonces $R_1 = .6858$, $R_2 = .4572$, $R_3 = 1.3716$, y el determinante de su $I(\beta)$ es $32.797 \cdot 10^{-5}$.

En la Tabla 5.1 se muestra cómo para el link doble exponencial, el experimento 2^3 balanceado D-óptimo está centrado en el $ED_{.9746}$, mientras que para el complementario log-log está centrado en el $ED_{.1015}$. Una vez más se observa que el experimento 2^3 D-óptimo para el modelo logístico es considerablemente menos informativo, que los experimentos D-óptimos para los otros link considerados.

5.2. Diseños D-óptimos a dos niveles

En las tablas 5.4 y 5.5 se muestran los experimentos D-óptimos a dos niveles entre todos los experimentos soportados en un hiper-rectángulo centrado en el $ED_{F(z_0)}$, para varios z_0 's para el link logit y probit respectivamente. Comparando estos diseños con los obtenidos en el caso balanceado, en las tablas 5.2 y 5.3 se comprueba que sus determinantes son aproximadamente tres veces más grandes.

	Link			
	Logit	Probit	D.Exp.	C.log-log
$F(z_0)$.9053	.5000	.9746	.1015
$ \beta_i R_i$	2.8500	1.3716	2.9799	2.9448
$F(z_1)$.0018	.0000	.0013	.0000
$F(z_{2,3})$.3561	.0851	.5000	.0056
$F(z_4)$.9940	.9149	.9987	.8692
$F(z_5)$.3561	.0851	.5000	.0056
$F(z_{6,7})$.9940	.9149	.9987	.8692
$F(z_8)$	1.0000	1.0000	1.0000	1.0000
$(\beta_1\beta_2\beta_3)^2 \det(I(\beta))$	$2.130 \cdot 10^{-3}n^4$	$11.807 \cdot 10^{-3}n^4$	$56.900 \cdot 10^{-3}n^4$	$42.334 \cdot 10^{-3}n^4$

Tabla 5.1: Experimentos 2^3 balanceados D-óptimos, con $\lambda_i = 1/8$, para modelos de primer orden.

Bajo el link logístico y puntos centrales que verifican $.50 \leq F(z_0) < .58197$, la solución D-óptima es un experimento de seis puntos, donde $\lambda_1 = \lambda_8 = 0$, y $|\beta_1|R_1 = |\beta_2|R_2 = |\beta_3|R_3$. Por lo tanto, $z_2 = z_3 = z_5$ y $z_4 = z_6 = z_7$. Por otro lado, para puntos centrales que verifican $.58197 < F(z_0)$, la solución D-óptima es un experimento balanceado de cuatro puntos, con $|\beta_1|R_1 = |\beta_2|R_2$, y $|\beta_3|R_3$ distinto. De esta forma, $z_2 = z_3$ y $z_6 = z_7$; esto se corresponde con una fracción un medio de un experimento completo 2^3 , aunque no coincide con la fracción por defecto 2^{3-1} para los modelos lineales normales de primer orden.

Llevando a cabo un estudio sobre los resultados obtenidos en función de z_0 , encontramos que para el modelo logístico, el mejor experimento de tres factores a dos niveles está centrado en cualquier punto del plano $ED_{.9147}$ y tiene cuatro puntos soporte, uno en el $ED_{.9853}$, otro en el $ED_{.6312}$ y dos en el plano $ED_{.1917}$. Si por ejemplo, $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_3 = 1$, el experimento localmente D-óptimo a dos niveles es tal que $R_1 = 1.4116$, $R_2 = .9411$, $R_3 = 3.8113$ y el determinante de su $I(\beta)$ es $5.1764 \cdot 10^{-4}n^4$, al menos nueve veces mayor que el obtenido para el mejor experimento 2^3 balanceado.

Link Logit

$F(z_0)$	z_0	$ \beta_1 $	R_1	$ \beta_2 $	R_2	$ \beta_3 $	R_3	$F(z_1)$	$F(z_{2,3})$	$F(z_4)$	$F(z_5)$	$F(z_{6,7})$	$F(z_8)$	$(\beta_1\beta_2\beta_3)^2 \det(I(\beta))$
.50	.000	1.8980		1.8980		1.8980		.0034	.1303	.8697	.1303	.8697	.9966	$1.5904 \cdot 10^{-3} n^4$
.55	.201	1.9090		1.9090		1.9090		.0040	.1534	.8918	.1534	.8918	.9973	$1.5989 \cdot 10^{-3} n^4$
.60	.405	1.9425		1.9425		1.9425		.0044	.1770	.9128	.1770	.9128	.9980	$1.6248 \cdot 10^{-3} n^4$
.65	.619	1.9995		1.9995		1.9995		.0046	.2009	.9320	.2009	.9320	.9987	$1.6697 \cdot 10^{-3} n^4$
.70	.847	2.0819		2.0819		2.0819		.0045	.2254	.9493	.2254	.9493	.9992	$1.7357 \cdot 10^{-3} n^4$
.75	1.099	2.1931		2.1931		2.1931		.0042	.2508	.9641	.2508	.9641	.9995	$1.8242 \cdot 10^{-3} n^4$
.80	1.386	2.3393		2.3393		2.3393		.0036	.2783	.9765	.2783	.9765	.9998	$1.9335 \cdot 10^{-3} n^4$
.85	1.735	2.5343		2.5343		2.5343		.0028	.3101	.9862	.3101	.9862	.9999	$2.0511 \cdot 10^{-3} n^4$
.90	2.197	2.8125		2.8125		2.8125		.0020	.3509	.9934	.3509	.9934	1.000	$2.1297 \cdot 10^{-3} n^4$
.9053	2.258	2.8500		2.8500		2.8500		.0018	.3561	.9940	.3561	.9940	1.000	$2.1308 \cdot 10^{-3} n^4$
.95	2.944	3.2903		3.2903		3.2903		.0010	.4144	.9980	.4144	.9980	1.000	$1.9806 \cdot 10^{-3} n^4$

Tabla 5.2: Experimento 2^3 , con $\lambda_i = 1/8$, centrado en un punto del plano de dosis efectiva $ED_{F(z_0)}$, para un modelo de respuesta binaria de primer orden cuando se usa la función link logit. El mejor diseño balanceado 2^3 aparece en negrita.

Link Probit

$F(z_0)$	z_0	$ \beta_1 R_1, \beta_2 R_2$	$ \beta_3 R_3$	$F(z_1)$	$F(z_{2,3})$	$F(z_4)$	$F(z_5)$	$F(z_{6,7})$	$F(z_8)$	$(\beta_1\beta_2\beta_3)^2 \det(I(\beta))$
.50	.000	1.3716	1.3716	.0000	.0851	.9149	.0851	.9149	1.000	11.807 $10^{-3} n^4$
.55	.126	1.3751	1.3751	.0000	.1058	.9333	.1058	.9333	1.000	11.696 $10^{-3} n^4$
.60	.253	1.3854	1.3854	.0000	.1288	.9494	.1288	.9494	1.000	11.361 $10^{-3} n^4$
.65	.385	1.4033	1.4033	.0000	.1543	.9632	.1543	.9632	1.000	10.801 $10^{-3} n^4$
.70	.524	1.4294	1.4294	.0001	.1827	.9746	.1827	.9746	1.000	10.012 $10^{-3} n^4$
.75	.674	1.4641	1.4641	.0001	.2149	.9838	.2149	.9838	1.000	8.978 $10^{-3} n^4$
.80	.842	1.5067	1.5067	.0001	.2530	.9906	.2530	.9906	1.000	7.672 $10^{-3} n^4$
.85	1.036	1.5285	2.2141	.0000	.1195	.9699	.5767	.9994	1.000	6.911 $10^{-3} n^4$
.90	1.282	1.6171	2.5188	.0000	.1080	.9771	.7144	.9999	1.000	6.655 $10^{-3} n^4$
.95	1.645	1.6919	2.8016	.0000	.1237	.9870	.8560	1.000	1.000	5.452 $10^{-3} n^4$

Tabla 5.3: Experimento 2^3 , con $\lambda_i = 1/8$, centrado en un punto del plano de dosis efectiva $ED_{F(z_0)}$, para un modelo de respuesta binaria de primer orden cuando se usa la función link probit. El mejor diseño balanceado 2^3 aparece en negrita.

En la Figura 5.2 se representa el mejor experimento de tres factores a dos niveles, el mejor experimento 2^3 balanceado y el mejor experimento fraccional “por defecto 2^{3-1} ” que será estudiado en la siguiente sección. En el conjunto de los modelos binomiales, los mejores experimentos a dos niveles no asignan ensayos a los vértices de los hiper-rectángulos más alejados del $ED_{.50}$. Cabe destacar que las configuraciones de los mejores experimentos a dos niveles, Figura 5.2.b, son poco convencionales en el sentido de que son fraccionales un medio en vez de experimentos factoriales completos. Además, a diferencia de lo que ocurre para los diseños balanceados de la Sección 5.3, los diseños factoriales fraccionales “por defecto 2^{3-1} ” no tienen matrices de diseño estandarizadas ortogonales.

Para el link probit, cuando los puntos centrales verifican que $.50 \leq F(z_0) < .60776$, la solución D-óptima es un experimento de seis puntos, con $|\beta_1|R_1 = |\beta_2|R_2 = |\beta_3|R_3$. Por otro lado, cuando $.60776 < F(z_0)$, una de las tres configuraciones D-óptimas equivalentes es un experimento de cuatro puntos balanceado, con $|\beta_1|R_1 = |\beta_2|R_2$, y con $|\beta_3|R_3$ distinto.

El mejor experimento D-óptimo de tres factores a dos niveles para el link probit, está centrado en el plano $ED_{.8928}$ y tiene cuatro puntos, uno en el $ED_{.9743}$, otro en el $ED_{.7038}$ y dos en el plano $ED_{.0982}$. Si por ejemplo, $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_3 = 1$, entonces $R_1 = .8098$, $R_2 = .5399$, $R_3 = 2.5331$, y el determinante de su $I(\beta)$ es $2.8845 \cdot 10^{-3}n^4$, más de nueve veces el valor del determinante del mejor experimento balanceado 2^3 centrado en el $ED_{.8928}$, y más de 5 veces el determinante del experimento a dos niveles para el modelo logístico.

5.3. “ 2^{3-1} por defecto” D-óptimos

Para modelos de respuesta normal, las fracciones un-medio de un factorial 2^3 con determinante máximo de $I(\beta)$, son lo que hemos denominado fracciones “por defecto 2^{3-1} ”. Hay dos fracciones en un diseño factorial completo, la formada por

Link Logit

$F(z_0)$	z_0	$ \beta_1 R_1, \beta_2 R_2$	$ \beta_3 R_3$	$F(z_{2,3})$	$F(z_4)$	$F(z_5)$	$F(z_{6,7})$	$\lambda_{2,3,5}$	λ_4	$\lambda_{6,7}$	$\lambda_{1,8}$	$(\beta_1\beta_2\beta_3)^2 \det(I(\beta))$
.50	.000	1.9804	1.9804	.1213	.8787	.1213	.8787	.1667	.1667	.1667	.000	$4.6109 \cdot 10^{-3} n^4$
.55	.201	1.9962	1.9962	.1424	.9000	.1424	.9000	.1908	.1425	.1425	.000	$4.7419 \cdot 10^{-3} n^4$
.58	.323	2.0195	2.0195	.1549	.9123	.1549	.9123	.2031	.1303	.1303	.000	$4.9396 \cdot 10^{-3} n^4$
.59	.364	1.9998	2.8676	.0756	.8170	.3169	.9620	.2500	.2500	.0000	.000	$5.1983 \cdot 10^{-3} n^4$
.60	.405	2.0081	2.8820	.0775	.8234	.3255	.9640	.2500	.2500	.0000	.000	$5.5118 \cdot 10^{-3} n^4$
.65	.619	2.0570	2.9583	.0879	.8551	.3689	.9728	.2500	.2500	.0000	.000	$7.2692 \cdot 10^{-3} n^4$
.70	.847	2.1204	3.0447	.1000	.8853	.4137	.9800	.2500	.2500	.0000	.000	$9.3451 \cdot 10^{-3} n^4$
.75	1.099	2.2035	3.1471	.1142	.9136	.4598	.9859	.2500	.2500	.0000	.000	$11.7112 \cdot 10^{-3} n^4$
.80	1.386	2.3155	3.2751	.1314	.9395	.5076	.9906	.2500	.2500	.0000	.000	$14.2685 \cdot 10^{-3} n^4$
.85	1.735	2.4742	3.4469	.1529	.9622	.5580	.9944	.2500	.2500	.0000	.000	$16.7571 \cdot 10^{-3} n^4$
.90	2.197	2.7206	3.7050	.1813	.9808	.6133	.9973	.2500	.2500	.0000	.000	$18.4954 \cdot 10^{-3} n^4$
.9147	2.372	2.8233	3.8113	.1917	.9853	.6312	.9979	.2500	.2500	.0000	.000	$18.6353 \cdot 10^{-3} n^4$
.95	2.944	3.1883	4.1882	.2238	.9941	.6805	.9992	.2500	.2500	.0000	.000	$17.3333 \cdot 10^{-3} n^4$

Tabla 5.4: Experimentos D-óptimos, entre todos los experimentos de tres factores a dos niveles soportados en los vértices de un hiper-rectángulo centrado en el plano $ED_{F(z_0)}$, bajo un modelo de respuesta binaria con predictor lineal de primer orden y función link logit. El experimento con mayor determinante aparece en negrita. Se pueden obtener dos configuraciones D-óptimas alternativas equivalentes a las que aparecen en esta tabla, simplemente intercambiando las columnas $|\beta_3|R_3$ por $|\beta_1|R_1$ ó $|\beta_2|R_2$, $F(z_4)$ por $F(z_6)$ ó $F(z_7)$ y λ_4 por λ_6 ó λ_7 .

$F(z_0)$	z_0	$ \beta_1 R_1, \beta_2 R_2$	$ \beta_3 R_3$	$F(z_{2,3})$	$F(z_4)$	$F(z_5)$	$F(z_{6,7})$	$\lambda_{2,3,5}$	λ_4	$\lambda_{6,7}$	$\lambda_{1,8}$	$(\beta_1 \beta_2 \beta_3)^2 \det(I(\beta))$
.50	.000	1.3778	1.3778	.0841	.9159	.0841	.9159	.1667	.1667	.1667	.000	37.1762 $10^{-3} n^4$
.55	.126	1.3864	1.3864	.1037	.9347	.1037	.9347	.1920	.1413	.1413	.000	37.4778 $10^{-3} n^4$
.60	.253	1.4074	1.4074	.1243	.9516	.1243	.9516	.2070	.1263	.1263	.000	38.2653 $10^{-3} n^4$
.61	.279	1.3916	1.8446	.0587	.8884	.2549	.9832	.2500	.2500	.0000	.000	38.8665 $10^{-3} n^4$
.65	.385	1.4160	1.9157	.0630	.9035	.2977	.9893	.2500	.2500	.0000	.000	47.7661 $10^{-3} n^4$
.70	.524	1.4480	2.0110	.0686	.9206	.3592	.9944	.2500	.2500	.0000	.000	60.2894 $10^{-3} n^4$
.75	.674	1.4827	2.1164	.0747	.9362	.4307	.9974	.2500	.2500	.0000	.000	73.9437 $10^{-3} n^4$
.80	.842	1.5218	2.2365	.0815	.9504	.5138	.9990	.2500	.2500	.0000	.000	87.7058 $10^{-3} n^4$
.85	1.036	1.5685	2.3796	.0896	.9636	.6099	.9997	.2500	.2500	.0000	.000	99.3256 $10^{-3} n^4$
.8928	1.241	1.6196	2.5331	.0982	.9743	.7038	.9999	.2500	.2500	.0000	.000	103.8413 $10^{-3} n^4$
.90	1.282	1.6300	2.5637	.0999	.9760	.7208	.9999	.2500	.2500	.0000	.000	103.6630 $10^{-3} n^4$
.95	1.645	1.7286	2.8430	.1154	.9881	.8486	1.000	.2500	.2500	.0000	.000	87.1781 $10^{-3} n^4$

Link Probit

Tabla 5.5: Experimentos D-óptimos, entre todos los experimentos de tres factores a dos niveles soportados en los vértices de un hiper-rectángulo centrado en el plano $ED_{F(z_0)}$, bajo un modelo de respuesta binaria con predictor lineal de primer orden y función link probit. El experimento con mayor determinante aparece en negrita. Se pueden obtener dos configuraciones D-óptimas alternativas equivalentes a las que aparecen en esta tabla, simplemente intercambiando las columnas $|\beta_3| R_3$ por $|\beta_1| R_1$ ó $|\beta_2| R_2$, $F(z_4)$ por $F(z_6)$ ó $F(z_7)$ y λ_4 por λ_6 ó λ_7 .

los vértices (x_2, x_3, x_5, x_8) y la formada por (x_1, x_4, x_6, x_7) , cuyas matrices de diseño estandarizadas son ortogonales. De hecho, para modelos normales, el determinante de $I(\beta)$ de estas dos fracciones complementarias, es igual al determinante de un experimento balanceado 2^3 con el mismo número de ensayos, n .

Por otro lado, en las tablas 5.4 y 5.5 se observa que para modelos de respuesta binaria, existen fracciones “especialmente no balanceadas” con matrices de diseño estandarizadas no ortogonales, y por lo tanto, que tienen un determinante mucho mas grande que aquel obtenido para el factorial completo. En esta sección, ilustramos cómo los experimentos fraccionales “ 2^{3-1} por defecto” no son los mejores en el caso de respuesta binaria. Para ello hemos calculado los experimentos fraccionales D-óptimos “ 2^{3-1} por defecto” centrados en varios x_0 's, obteniendo el determinante de su $I(\beta)$, y los hemos comparado con el determinante de $I(\beta)$ de las tablas para los diseños 2^3 D-óptimos en el caso balanceado, Tabla 5.2 y 5.3, y para el no balanceado, Tabla 5.4 y 5.5. El siguiente resultado es una consecuencia del Corolario 5.1, con $\lambda_{1,4,6,7} = 0$.

Corolario 5.2. *Bajo el modelo binomial con $p(x_i, \beta) = F(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i})$, el determinante de $I(\beta)$ para la primera fracción “ 2^{3-1} por defecto” descrito en el párrafo anterior, es tal que*

$$(\beta_1 \beta_2 \beta_3)^2 \det(I(\beta)) = 4^4 n^4 (\beta_1 R_1)^2 (\beta_2 R_2)^2 (\beta_3 R_3)^2 \lambda_2 \lambda_3 \lambda_5 \lambda_8 h(z_2) h(z_3) h(z_5) h(z_8), \quad (5.2)$$

donde $z_2 = z_0 + |\beta_1| R_1 - |\beta_2| R_2 - |\beta_3| R_3$, $z_3 = z_0 - |\beta_1| R_1 + |\beta_2| R_2 - |\beta_3| R_3$, $z_5 = z_0 - |\beta_1| R_1 - |\beta_2| R_2 + |\beta_3| R_3$ y $z_8 = z_0 + |\beta_1| R_1 + |\beta_2| R_2 + |\beta_3| R_3$.

En este caso, el número de puntos soporte es igual al número de parámetros, y de esta forma, la asignación D-óptima es balanceada, esto es, $\lambda_{2,3,5,8} = .25$. La Tabla (5.6) muestra los experimentos localmente D-óptimos cuando forzamos a que el experimento sea un diseño fraccional “ 2^{3-1} por defecto”. El estudio se ha llevado

a cabo fijando el valor central del diseño, y buscando el mejor diseño centrado en dicho punto tanto para el link logit como para el probit.

La configuración D-óptima es tal que $|\beta_1|R_1 = |\beta_2|R_2 = |\beta_3|R_3$, y por lo tanto, $z_2 = z_3 = z_5$ (ó $z_4 = z_6 = z_7$). Para el link logit, la fracción D-óptima un-medio soportada en (x_2, x_3, x_5, x_8) es mejor que la soportada en (x_1, x_4, x_6, x_7) , mientras que para el link probit, la soportada sobre (x_1, x_4, x_6, x_7) es la mejor de todas. Nótese que esto contrasta con el hecho de que para modelos lineales normales, estas dos fracciones un-medio son estadísticamente equivalentes.

El mejor diseño fraccional “ 2^{3-1} por defecto”, de entre todos los z_0 , para el link logit es aquel centrado en el plano $ED_{.6960}$, con tres puntos soporte en el $ED_{.3349}$ y uno en el $ED_{.9954}$, como se puede observar en la Figura 5.2.c. Cuando $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_3 = 1$, el experimento localmente D-óptimo “ 2^{3-1} por defecto” es tal que $R_1 = .7571$, $R_2 = .5047$, $R_3 = 1.5142$, y el determinante de su $I(\beta)$ es $1.704910^{-5}n^4$, que es menos de un tercio del determinante del mejor factorial 2^3 completo y balanceado (ver Tabla 5.1), y a su vez es una treceava parte del determinante de la mejor fracción de un experimento 2^3 para el modelo de respuesta binaria, (ver Tabla 5.2).

El mejor de los experimentos “ 2^{3-1} por defecto” centrado en el plano $ED_{.50}$, tiene tres puntos soporte en el $ED_{.2135}$ y uno en el $ED_{.9804}$; cuando $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_3 = 1$, entonces $R_1 = .6520$, $R_2 = .4343$, $R_3 = 1.3040$, y el determinante de su $I(\beta)$ es $1.243110^{-5}n^4$, que es alrededor de un tercio del determinante del mejor experimento balanceado 2^3 centrado en el $ED_{.50}$, y un décimo del determinante del mejor experimento de tres factores y dos niveles centrado en dicho plano.

Link Logit

$F(z_0)$	z_0	$ \beta_j R_j$	$F(z_{2,3,5})$	$F(z_8)$	$\lambda_{2,3,5,8}$	$(\beta_1\beta_2\beta_3)^2 \det(I(\beta))$
.50	.000	1.3040	.2135	.9804	.25	$4.475 \cdot 10^{-4} n^4$
.55	.201	1.3441	.2417	.9857	.25	$5.117 \cdot 10^{-4} n^4$
.60	.405	1.3922	.2716	.9899	.25	$5.652 \cdot 10^{-4} n^4$
.65	.619	1.4500	.3034	.9931	.25	$6.015 \cdot 10^{-4} n^4$
.6960	.828	1.5142	.3349	.9954	.25	$6.138 \cdot 10^{-4} n^4$
.70	.847	1.5204	.3378	.9955	.25	$6.137 \cdot 10^{-4} n^4$
.75	1.099	1.6078	.3754	.9973	.25	$5.935 \cdot 10^{-4} n^4$
.80	1.386	1.7203	.4173	.9986	.25	$5.329 \cdot 10^{-4} n^4$
.85	1.735	1.8730	.4655	.9994	.25	$4.252 \cdot 10^{-4} n^4$
.90	2.197	2.1014	.5239	.9998	.25	$2.714 \cdot 10^{-4} n^4$
.95	2.955	2.5235	.6037	.9999	.25	$.9594 \cdot 10^{-4} n^4$

Link Probit

$F(z_0)$	z_0	$ \beta_j R_j$	$F(z_1)$	$F(z_{4,6,7})$	$\lambda_{1,4,6,7}$	$(\beta_1\beta_2\beta_3)^2 \det(I(\beta))$
.50	.000	.7732	.0102	.7803	.25	$2.069 \cdot 10^{-3} n^4$
.5318	.080	.7758	.0123	.8039	.25	$2.091 \cdot 10^{-3} n^4$
.55	.126	.7772	.0137	.8167	.25	$2.084 \cdot 10^{-3} n^4$
.60	.253	.7807	.0184	.8494	.25	$1.986 \cdot 10^{-3} n^4$
.65	.385	.7836	.0247	.8788	.25	$1.779 \cdot 10^{-3} n^4$
.70	.524	.7858	.0334	.9049	.25	$1.481 \cdot 10^{-3} n^4$
.75	.674	.7872	.0458	.9281	.25	$1.123 \cdot 10^{-3} n^4$
.80	.842	.7874	.0642	.9483	.25	$.748 \cdot 10^{-3} n^4$
.85	1.036	.7859	.0932	.9658	.25	$.408 \cdot 10^{-3} n^4$
.90	1.282	.7813	.1440	.9804	.25	$.155 \cdot 10^{-3} n^4$
.95	1.645	.7695	.2534	.9921	.25	$.024 \cdot 10^{-3} n^4$

Tabla 5.6: Experimento D-óptimo “ 2^{3-1} por defecto”, soportado o bien en (x_2, x_3, x_5, x_8) o en (x_1, x_4, x_6, x_7) , centrado en el plano $ED_{F(z_0)}$, y todos bajo modelos de primer orden. Los mejores experimentos de este tipo son los que están en negrita.

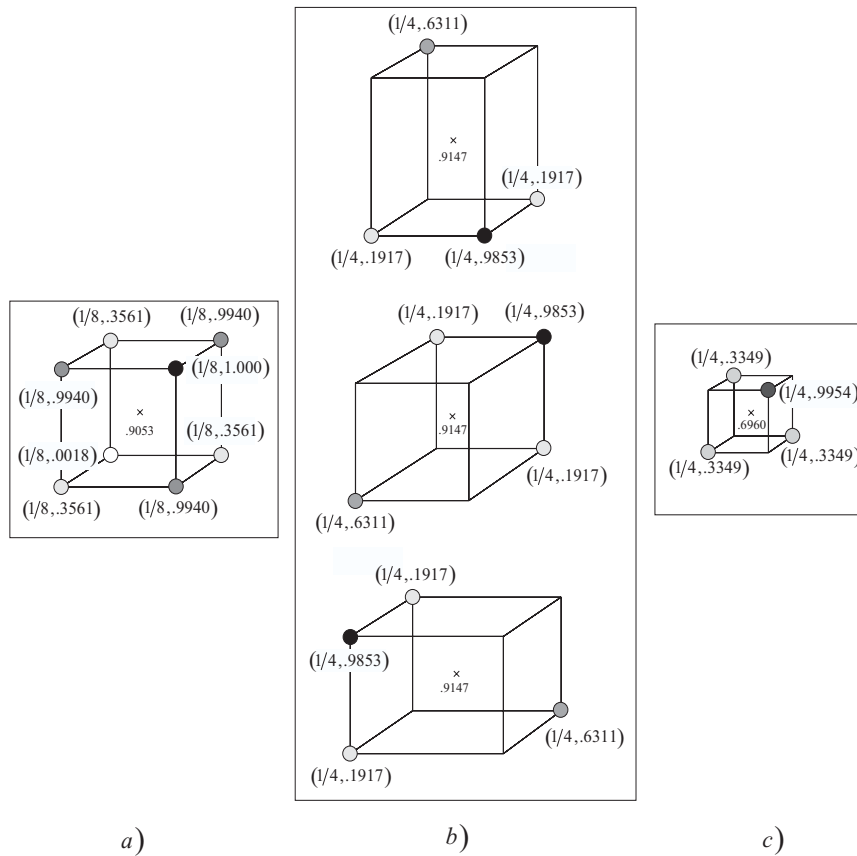


Figura 5.2: a) Experimento 2^3 D-óptimo balanceado, con $(\beta_1\beta_2\beta_3)^2\det(I(\beta)) = .00213n^4$, b) tres configuraciones D-óptimas de tres factores a dos niveles, con $(\beta_1\beta_2\beta_3)^2\det(I(\beta)) = .018635n^4$, y c) Experimento D-óptimo “ 2^{3-1} por defecto”, con $(\beta_1\beta_2\beta_3)^2\det(I(\beta)) = .00061n^4$, todos para el modelo de regresión logístico de primer orden. Las longitudes de los lados de los hiperrectángulos son proporcionales a los rangos localmente D-óptimos, cuando $\beta_1 = \beta_2 = \beta_3$. Las cantidades entre paréntesis son de la forma $(\lambda_i, F(z_i))$, y la cantidad en el centro de los hiper-rectángulos es $F(z_0)$.

Capítulo 6

Experimentos a dos niveles en presencia de interacciones

Hasta el momento sólo se han considerado modelos de respuesta binaria con predictor lineal de primer orden sin interacciones. En este capítulo se añade el término de interacción al predictor lineal, (modelo saturado), y se calculan los diseños D-óptimos a dos niveles para los link logit y probit y, en cada uno de ellos, para dos y tres factores.

En los modelos lineales normales, el diseño factorial era óptimo tanto para el modelo sin interacciones como para el modelo con interacciones. En modelos de respuesta binaria esta propiedad no se verifica. Los diseños para los modelos de respuesta binaria con interacción en el predictor lineal son distintos a aquellos en los que no se considera el término de interacción. Para modelos de respuesta binaria con interacción, los diseños a dos niveles D-óptimos dependen de dónde se encuentren centrados, tal y como ocurría cuando no se añadía el término de interacción, pero ya no se verifica que los diseños centrados en el mismo ED tengan el mismo determinante de la matriz de información de Fisher.

6.1. Caso de dos factores

Asumamos que la relación entre la respuesta, y_i , y los niveles de dosis de los dos factores, $x_i = (x_{1i}, x_{2i})$, puede ser modelada a través de

$$p(x_i; \beta) = F(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}) = F(z_i). \quad (6.1)$$

Cuando se considera este tipo de modelos, el conjunto de puntos $x_i = (x_{1i}, x_{2i})$ tal que $p(x_i; \beta) = F(z)$ dado un z , definen una hipérbola de asíntotas $x_1 = -\beta_2/\beta_{12}$ y $x_2 = -\beta_1/\beta_{12}$. En particular, cuando $z = \beta_0 - \beta_1\beta_2/\beta_{12}$, la hipérbola que se forma coincide con las asíntotas de cualquiera de las otras hipérbolas. A esta hipérbola formada por dos rectas la denominamos hipérbola degenerada.

El determinante de la matriz de información de Fisher viene definido por la expresión

$$\det(I(\beta)) = n^4 \det \left(\sum_i^q \lambda_i h(z_i) (1, x_{1i}, x_{2i}, x_{1i}x_{2i})' (1, x_{1i}, x_{2i}, x_{1i}x_{2i}) \right). \quad (6.2)$$

Aplicando un resultado análogo a la Proposición 2.2 para experimentos de dos factores a dos niveles centrados en un punto en la hipérbola $ED_{F(z_0)}$, con $z_0 = \beta_0 + \beta_1 x_{10} + \beta_2 x_{20} + \beta_{12} x_{10} x_{20}$, obtenemos que

Proposición 6.1. *Para el modelo binomial con $p(x_i, \beta) = F(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i})$, el determinante de $I(\beta)$ para un experimento soportado en los vértices de un rectángulo orientado en los ejes cartesianos, tal y como podemos ver en la Figura 6.1, y centrado en el $ED_{F(z_0)}$, con la mitad de los rangos R_1 y R_2 , es tal que*

$$\beta_{12}^4 \det(I(\beta)) = 4^4 n^4 (\beta_{12} R_1 R_2)^4 \lambda_1 \lambda_2 \lambda_3 \lambda_4 h(z_1) h(z_2) h(z_3) h(z_4), \quad (6.3)$$

donde $z_1 = z_0 - c_1 - c_2 + |\beta_{12}| R_1 R_2$, $z_2 = z_0 + c_1 - c_2 - |\beta_{12}| R_1 R_2$, $z_3 = z_0 - c_1 + c_2 - |\beta_{12}| R_1 R_2$ y $z_4 = z_0 + c_1 + c_2 + |\beta_{12}| R_1 R_2$, con $c_1 = |\beta_1 + \beta_{12} x_{20}| R_1$ y $c_2 = |\beta_2 + \beta_{12} x_{10}| R_2$.

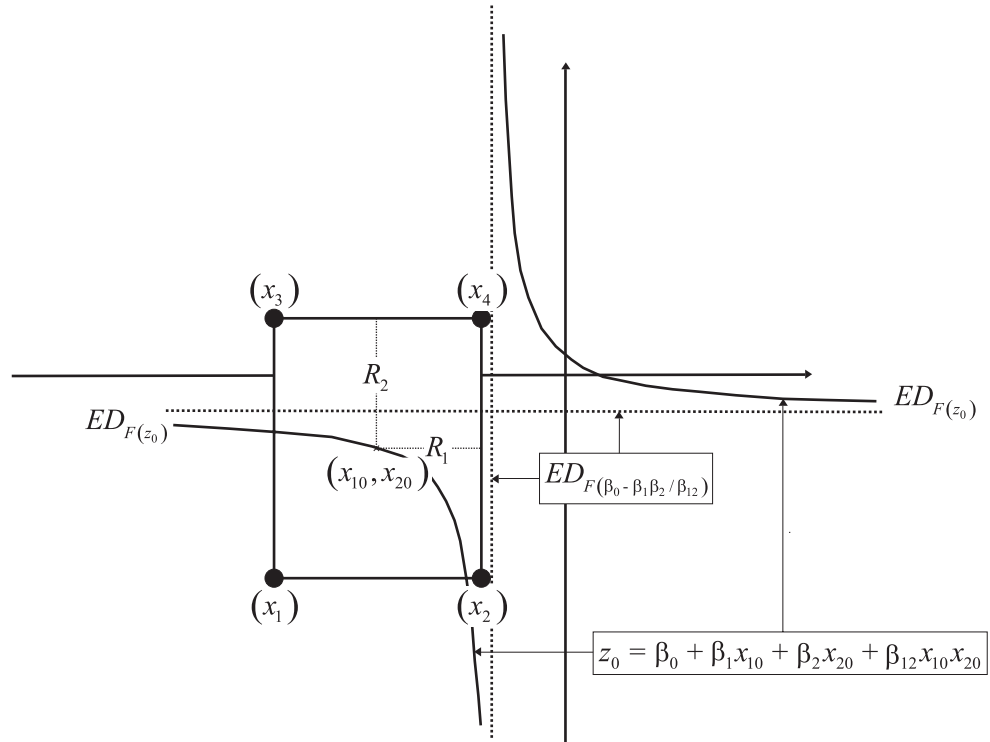


Figura 6.1: Diseño 2^2 para un modelo de respuesta binaria con predictor lineal con interacciones centrado en el $ED_{F(z_0)}$. Las curvas ED en este caso son hipérbolas.

Dado que el número de parámetros es igual al número de puntos soporte, las asignaciones D-óptimas son tales que $\lambda_{1,2,3,4} = .25$, y para éstas, $\beta_{12}^4 \det(I(\beta))$ depende únicamente de $(|\beta_{12}|R_1R_2, |\beta_1 + \beta_{12}x_{20}|R_1, |\beta_2 + \beta_{12}x_{10}|R_2)$ en vez del caso sin interacciones donde el determinante sólo dependía de $(|\beta_1|R_1, |\beta_2|R_2)$.

6.1.1. Diseño D-óptimo general

El máximo de la expresión (6.3), en función de z_0 y $(|\beta_{12}|R_1R_2, |\beta_1 + \beta_{12}x_{20}|R_1, |\beta_2 + \beta_{12}x_{10}|R_2)$, se alcanza cuando $|\beta_1 + \beta_{12}x_{20}|R_1 = 0$ y $|\beta_2 + \beta_{12}x_{10}|R_2 = 0$. Esto es, el diseño estará centrado justo en el punto de intersección de las asíntotas de la

hipérbola de forma que $x_0 = (-\beta_2/\beta_{12}, -\beta_1/\beta_{12})$ y $z_0 = \beta_0 - \beta_1\beta_2/\beta_{12}$. Para modelos saturados, (6.1), el experimento D-óptimo a dos niveles se encuentra centrado en la intersección de las asíntotas de las dosis efectivas $ED_{F(z)}$'s, y por lo tanto, x_0 y z_0 dependen de los parámetros desconocidos del modelo β 's.

Cuando centramos el experimento en $x_0 = (-\beta_2/\beta_{12}, -\beta_1/\beta_{12})$, tenemos que $z_1 = z_4$ y $z_2 = z_3$, con

$$\beta_{12}^4 \det(I(\beta)) = 4^4 n^4 (\beta_{12} R_1 R_2)^4 \lambda_1^2 \lambda_2^2 h(z_1)^2 h(z_2)^2, \quad (6.4)$$

donde $z_1 = z_0 + |\beta_{12}| R_1 R_2$ y $z_2 = z_0 - |\beta_{12}| R_1 R_2$. Debido a la similitud de la expresión (6.4) con la del determinante para el caso de un factor (3.1), los valores D-óptimos para los rangos que maximizan (6.4), pueden ser obtenidos sustituyendo $|\beta_1| R$ y $\beta_1^2 \det(I(\beta))$ por $|\beta_{12}| R_1 R_2$ y $\beta_{12}^2 (\det(I(\beta)))^{1/2}$ en (3.1). Nótese que la Tabla 6.1 es igual a la 3.1, salvo por las expresiones de la cabecera y por el significado que toma en este caso la columna z_0 . La cabecera es distinta debido a que los determinantes en ambos casos, aunque sean similares, varían sensiblemente, tal y como ya se ha comentado. Por otro lado la columna z_0 , y por consiguiente $F(z_0)$, representan el valor de la dosis efectiva sobre la que centraremos el diseño D-óptimo general, i.e. $z_0 = z_d$ y no como en el caso de un factor donde el valor de z_0 lo elegimos a priori para situar en el centro del diseño localmente D-óptimo.

El proceso para obtener el diseño D-óptimo general incluye los siguientes pasos. Primero se determinan unas estimaciones de los parámetros del modelo. Con ellas, y sabiendo que el centro del diseño D-óptimo debe estar centrado en la intersección de las asíntotas de los ED , calculamos dicho centro sin más que sustituir los parámetros en $x_0 = (-\beta_2/\beta_{12}, -\beta_1/\beta_{12})$. Una vez obtenido el centro, calculamos el ED que incluye a dicho centro y que no es más que el valor del ED de la hipérbola degenerada, esto es, $ED_{F(z_d)}$ con $z_d = z_0 = \beta_0 - \beta_1\beta_2/\beta_{12}$. En la Tabla 6.1 determinamos el valor de los rangos del diseño D-óptimo general.

A continuación presentamos un ejemplo donde se ilustra el uso de la Tabla 6.1. Asumamos que $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_{12} = 6$. El experimento localmente D-óptimo a dos niveles está centrado en $x_0 = -(1/3, 1/2)$, con $z_0 = \beta_0 - \beta_1\beta_2/\beta_{12} =$

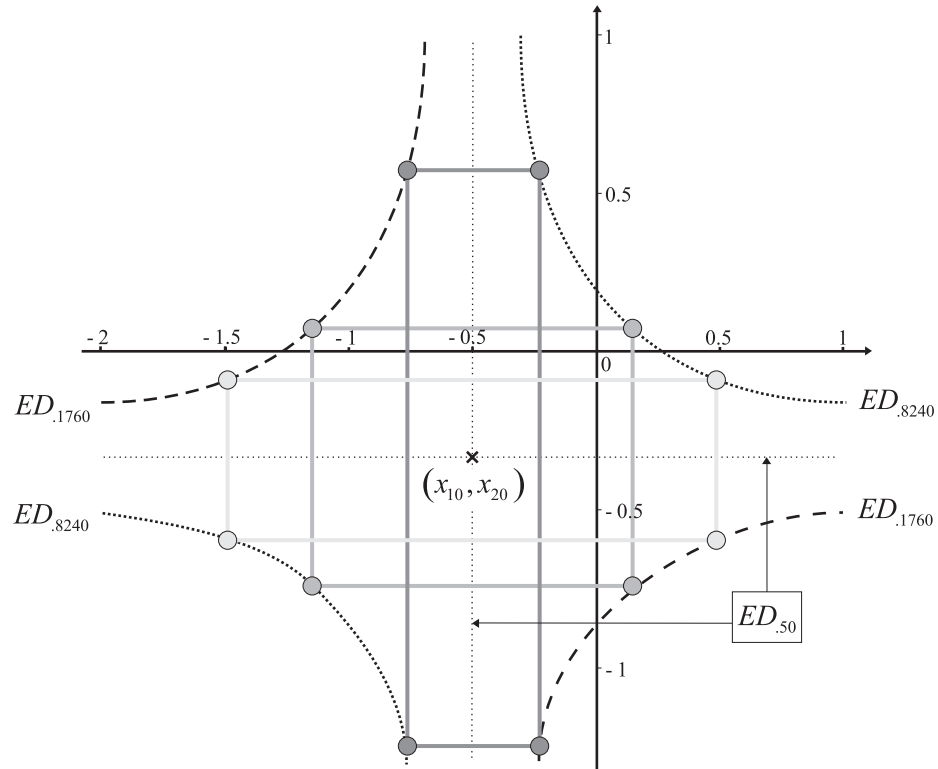


Figura 6.2: Diseños D-óptimos 2^2 para un modelo de respuesta binaria con predictor lineal saturado tal que $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$ y $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_{12} = 6$. En este caso hay infinitos diseños D-óptimos.

0. Este diseño asigna un cuarto de los ensayos a cada uno de los vértices de un rectángulo tal que $R_1 R_2 = .2572$. Dos puntos soporte se sitúan en el $ED_{.1760}$ y otros dos en el $ED_{.8240}$. En la Figura 6.2 podemos ver que, en este caso, hay infinitos diseños D-óptimos. Lo único que deben verificar es que se encuentren centrados en la intersección de las asíntotas y que sus rangos sean tales que $R_1 R_2 = .2572$.

6.1.2. Diseño D-óptimo centrado en un punto x_0

Si en vez de buscar el experimento óptimo general a dos niveles, estamos interesados en identificar el experimento D-óptimo entre todos aquellos soportados en los

Link Logit

z_0	$F(z_0)$	$ \beta_{12} R_1R_2$	$F(z_{1,4})$	$F(z_{2,3})$	$\sqrt{\lambda_i}$	$\beta_{12}^2\sqrt{(\det(I(\beta)))}$
z_0	$F(z_0)$	$ \beta_{123} R_1R_2R_3$	$F(z_{1,4,6,7})$	$F(z_{2,3,5,8})$	$\sqrt[3]{\lambda_i}$	$\beta_{123}^2\sqrt[4]{(\det(I(\beta)))}$
.000	.50	1.5434	.1760	.8240	.5	.0501 n^2
.201	.55	1.5487	.2062	.8519	.5	.0495 n^2
.405	.60	1.5652	.2387	.8777	.5	.0478 n^2
.619	.65	1.5937	.2739	.9014	.5	.0449 n^2
.847	.70	1.6371	.3122	.9230	.5	.0409 n^2
1.099	.75	1.6995	.3541	.9426	.5	.0358 n^2
1.386	.80	1.7885	.4008	.9599	.5	.0296 n^2
1.735	.85	1.9196	.4539	.9748	.5	.0225 n^2
2.197	.90	2.1286	.5171	.9870	.5	.0146 n^2
2.944	.95	2.5339	.6012	.9958	.5	.0064 n^2

Link Probit

z_0	$F(z_0)$	$ \beta_{12} R_1R_2$	$F(z_{1,4})$	$F(z_{2,3})$	$\sqrt{\lambda_i}$	$\beta_{12}^2\sqrt{(\det(I(\beta)))}$
z_0	$F(z_0)$	$ \beta_{123} R_1R_2R_3$	$F(z_{1,4,6,7})$	$F(z_{2,3,5,8})$	$\sqrt[3]{\lambda_i}$	$\beta_{123}^2\sqrt[4]{(\det(I(\beta)))}$
.000	.50	1.1382	.1275	.8725	.5	.1987 n^2
.126	.55	1.1372	.1559	.8967	.5	.1960 n^2
.253	.60	1.1345	.1891	.9174	.5	.1881 n^2
.385	.65	1.1298	.2283	.9351	.5	.1750 n^2
.524	.70	1.1230	.2747	.9503	.5	.1571 n^2
.674	.75	1.1138	.3302	.9631	.5	.1348 n^2
.842	.80	1.1018	.3974	.9740	.5	.1086 n^2
1.036	.85	1.0860	.4802	.9831	.5	.0795 n^2
1.282	.90	1.0650	.5857	.9905	.5	.0489 n^2
1.645	.95	1.0351	.7290	.9963	.5	.0196 n^2

Tabla 6.1: Experimento D-óptimo general entre todos los de dos factores, y entre todos los de tres factores a dos niveles soportados en los vértices de un hiper-rectángulo, bajo los modelos saturados (6.1) y (6.5). Se corresponden con los experimentos centrados en la intersección de las asíntotas de los ED_F 's, $x_0 = (-\beta_2/\beta_{12}, -\beta_1/\beta_{12})$ y $x_0 = (-\beta_{23}/\beta_{123}, -\beta_{13}/\beta_{123}, -\beta_{12}/\beta_{123})$, por lo que $z_0 = \beta_0 - \beta_1\beta_2/\beta_{12}$ y $z_0 = \beta_0 - \beta_{12}\beta_{13}\beta_{23}/\beta_{123}^2$ respectivamente.

vértices de un rectángulo centrado en un valor de x_0 previamente especificado, tal y como ocurría para los modelos de respuesta binaria con predictor lineal sin interacciones, debemos maximizar (6.4) con respecto a $(|\beta_{12}|R_1R_2, |\beta_1 + \beta_{12}x_{20}|R_1, |\beta_2 + \beta_{12}x_{10}|R_2)$, con β y x_0 fijos. Sin embargo, a diferencia de lo que ocurría para el modelo de primer orden sin interacciones, aquí el diseño depende tanto de $F(z_0)$ como de $F(z_d) = F(\beta_0 - \beta_1\beta_2/\beta_{12})$. En la Tabla 6.2 se caracterizan dichos experimentos para varias combinaciones de (z_0, z_d) ; para estos diseños se cumple que $|\beta_1 + \beta_{12}x_{20}|R_1 = |\beta_2 + \beta_{12}x_{10}|R_2$, salvo si $z_0 = z_d$. Si se verifica esta relación estaríamos en el caso visto en la sección anterior (6.1.1), en donde $|\beta_1 + \beta_{12}x_{20}|R_1 = 0$. En este caso tendríamos que hacer uso de la Tabla 6.1. Observe que bajo (6.1), la solución D-óptima a dos niveles centrada en un punto x_0 , no depende únicamente de z_0 sino también de x_0 .

A continuación se presenta un ejemplo donde se ilustra el uso de la Tabla 6.2. Se desea obtener un diseño óptimo a dos niveles para un modelo de respuesta binaria con función link logit y predictor lineal saturado, centrado en el $ED_{.70}$, de forma que $\beta_0 = 1$, $\beta_1 = 1$, $\beta_2 = 1$, y $\beta_{12} = 1$. El modelo de respuesta binaria es de la forma

$$p(x_i; \beta) = F(1 + 1x_{1i} + 1x_{2i} + 1x_{1i}x_{2i}) = F(z_i),$$

donde $z_d = \beta_0 - \beta_1\beta_2/\beta_{12} = 0$, que define la hipérbola degenerada, y que se corresponde con la dosis efectiva del 50%. A partir de la Tabla 6.2, podemos concluir que el diseño D-óptimo a dos niveles centrado en el $ED_{.70}$, sitúa dos de sus cuatro puntos sobre la dosis efectiva del 31.397%, y los otros dos sobre las dosis efectivas del 99.204% y 53.161% respectivamente, con un valor del determinante igual a $(6.42 \cdot 10^{-2} n^4)$. Nótese que todos los diseños centrados en el $ED_{.70}$ tienen el mismo determinante y sitúan sus puntos sobre los mismos ED . Por otro lado, los valores de R_1 y R_2 varían con el centro del diseño. Para calcularlos, haremos uso de la igualdad

$$|\beta_1 + \beta_{12}x_{20}|R_1 = |\beta_2 + \beta_{12}x_{10}|R_2 = 1.1748$$

de forma que

Link Logit

$F(z_0)$	$F(z_d)$	$ \beta_1 + \beta_{12}x_{20} R_1$	$F(z_1)$	$F(z_{2,3})$	$F(z_4)$	λ_i	$\beta_{12}^4 \det(I(\beta))$
.50	.50	.0000	.8240	.1760	.8240	.25	$25.119 \cdot 10^{-4} n^4$
.60	.50	.7959	.5929	.2392	.9723	.25	$12.813 \cdot 10^{-4} n^4$
.70	.50	1.1748	.5316	.3140	.9920	.25	$6.419 \cdot 10^{-4} n^4$
.80	.50	1.5668	.5059	.4051	.9982	.25	$2.636 \cdot 10^{-4} n^4$
.50	.60	.7827	.0441	.8192	.5136	.25	$12.038 \cdot 10^{-4} n^4$
.60	.60	.0000	.1760	.8249	.1760	.25	$22.845 \cdot 10^{-4} n^4$
.70	.60	.8547	.6881	.3087	.9854	.25	$10.529 \cdot 10^{-4} n^4$
.80	.60	1.3320	.6298	.3959	.9972	.25	$4.051 \cdot 10^{-4} n^4$
.50	.70	1.0969	.0262	.8053	.6843	.25	$5.516 \cdot 10^{-4} n^4$
.60	.70	.8086	.0635	.8682	.6325	.25	$8.675 \cdot 10^{-4} n^4$
.70	.70	.0000	.9230	.3122	.9230	.25	$16.714 \cdot 10^{-4} n^4$
.80	.70	.9898	.7728	.3938	.9944	.25	$1.155 \cdot 10^{-4} n^4$

Link Probit

$F(z_0)$	$F(z_d)$	$ \beta_1 + \beta_{12}x_{20} R_1$	$F(z_1)$	$F(z_{2,3})$	$F(z_4)$	λ_i	$\beta_{12}^4 \det(I(\beta))$
.50	.50	.0000	.8725	.1275	.8725	.25	$394.789 \cdot 10^{-4} n^4$
.60	.50	.4990	.5942	.2328	.9873	.25	$137.069 \cdot 10^{-4} n^4$
.70	.50	.6639	.5148	.3759	.9965	.25	$38.281 \cdot 10^{-4} n^4$
.80	.50	.7696	.5025	.5548	.9990	.25	$7.198 \cdot 10^{-4} n^4$
.50	.60	.5013	.0230	.8394	.5042	.25	$158.848 \cdot 10^{-4} n^4$
.60	.60	.0000	.9174	.1891	.9174	.25	$353.798 \cdot 10^{-4} n^4$
.70	.60	.5101	.6787	.3316	.9939	.25	$86.915 \cdot 10^{-4} n^4$
.80	.60	.6839	.6060	.5186	.9987	.25	$14.432 \cdot 10^{-4} n^4$
.50	.70	.6746	.0133	.8072	.6849	.25	$68.415 \cdot 10^{-4} n^4$
.60	.70	.5175	.0384	.8928	.6180	.25	$140.028 \cdot 10^{-4} n^4$
.70	.70	.0000	.2747	.9503	.2747	.25	$246.868 \cdot 10^{-4} n^4$
.80	.70	.5392	.7517	.4702	.9977	.25	$34.292 \cdot 10^{-4} n^4$

Tabla 6.2: Experimento D-óptimo, entre todos los experimentos de dos factores a dos niveles soportados en los vértices de un rectángulo centrado en la hipérbola $ED_{F(z_0)}$, bajo los modelos saturados, (6.4), con $z_d = \beta_0 - \beta_1\beta_2/\beta_{12}$. La solución D-óptima es tal que $|\beta_1 + \beta_{12}x_{20}|R_1 = |\beta_2 + \beta_{12}x_{10}|R_2$.

$$R_1 = \frac{1.1748}{|1 + x_{20}|} \text{ y } R_2 = \frac{1.1748}{|1 + x_{10}|}$$

y (x_{10}, x_{20}) tal que $(1 + 1x_{10} + 1x_{20} + 1x_{10}x_{20} = 0.847)$. En la Figura 6.3 aparece representado el diseño D-óptimo para el ejemplo expuesto. Como se puede comprobar los diseños D-óptimos centrados en el $ED_{.70}$, por su propia estructura, no conservan los valores de R_1 y R_2 . Esto se debe a que los puntos que definen los diseños óptimos, se apoyan siempre en los mismos ED y por consiguiente, al variar el centro del diseño sobre el $ED_{.70}$, variarán los rangos que definen dichos diseños.

En el caso de modelos de respuesta binaria sin interacciones los ED eran rectas (dos factores) o planos (tres factores). Debido a esto, los rangos de los diseños D-óptimos permanecían constantes, ya que tanto el centro del diseño como sus puntos soporte se movían en líneas o planos ED , que eran paralelos entre si. Ahora bien, la característica más notable de los diseños óptimos para modelos de respuesta binaria es su dependencia respecto de la hipérbola degenerada. Así, en el caso de interacciones, no sólo debemos definir sobre qué curva de dosis efectiva vamos a centrar nuestro diseño sino que además debemos conocer el valor de z_d , para poder determinar el diseño D-óptimo.

6.2. Caso de tres factores

Asumamos que la relación entre y_i y $x_i = (x_{1i}, x_{2i}, x_{3i})$, puede ser modelada a través de

$$p(x_i; \beta) = F(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_{12} x_{1i} x_{2i} + \beta_{13} x_{1i} x_{3i} + \beta_{23} x_{2i} x_{3i} + \beta_{123} x_{1i} x_{2i} x_{3i}). \quad (6.5)$$

El conjunto de puntos $x_i = (x_{1i}, x_{2i}, x_{3i})$ tal que $p(x_i; \beta) = F(z)$ dado un z , son hiperboloides cuyas asíntotas son, $x_1 = -\beta_{23}/\beta_{123}$, $x_2 = -\beta_{13}/\beta_{123}$ y $x_3 = -\beta_{12}/\beta_{123}$ (ver Figura 6.4). Aplicando un resultado análogo al la Proposición 2.2 sobre experimentos a dos niveles centrados en el $ED_{F(z_0)}$ con $z_0 = \beta_0 + \beta_1 x_{10} + \dots + \beta_{123} x_{10} x_{20} x_{30}$, obtenemos el siguiente resultado:

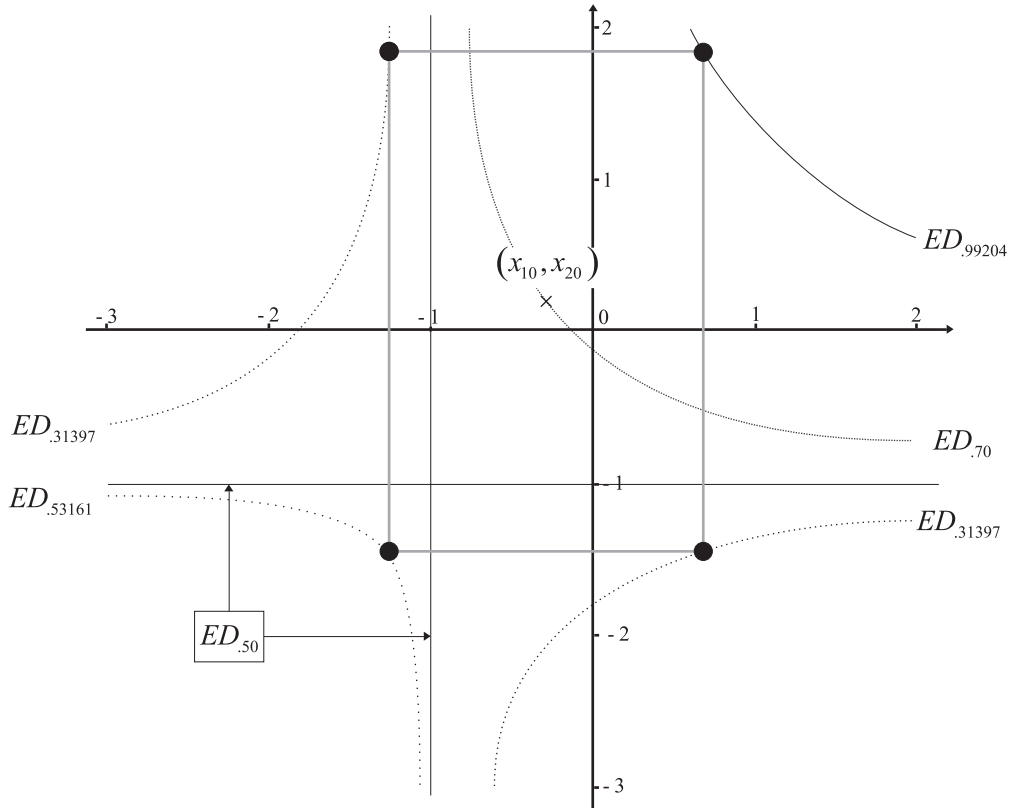


Figura 6.3: Diseño D-óptimo 2^2 para un modelo de respuesta binaria con link logit, predictor lineal saturado y centrado en el $ED_{.70}$ tal que $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$ con $\beta_0 = 1$, $\beta_1 = 1$, $\beta_2 = 1$ y $\beta_{12} = 1$.

Proposición 6.2. *Bajo modelos binomiales con parámetro (6.5), el determinante de $I(\beta)$ para un experimento soportado en los vértices de un hiper-rectángulo como el que se definió en el Capítulo 5, es tal que*

$$\beta_{123}^8 \det(I(\beta)) = 4^{12} n^8 (\beta_{123} R_1 R_2 R_3)^8 \lambda_1 h(z_1) \dots \lambda_8 h(z_8). \tag{6.6}$$

Una vez más, el número de parámetros es igual al número de puntos soporte. Por lo tanto, la asignación D-óptima es balanceada, con $\lambda_i = .125$. El determinante de $I(\beta)$ alcanza su máximo cuando $x_0 = (-\beta_{23}/\beta_{123}, -\beta_{13}/\beta_{123}, -\beta_{12}/\beta_{123})$, o lo que es lo mismo cuando $z_0 = \beta_0 - (\beta_{12}\beta_{13}\beta_{23}/\beta_{123}^2)$. Esto es, el centro del diseño coincide con la intersección de los planos asintóticos. Nótese que en el caso de dos factores,

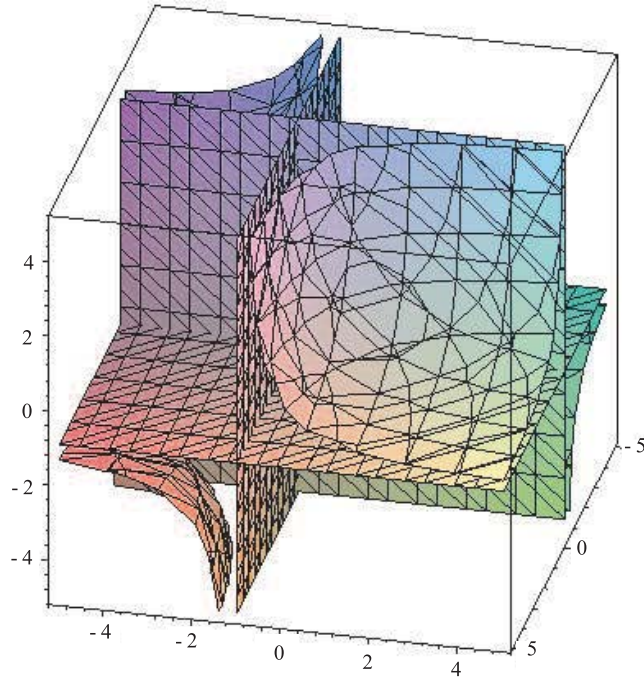


Figura 6.4: Superficies ED para un modelo de respuesta binaria de tres factores con predictor lineal saturado. Los planos definen los planos asintóticos de todos los hiperboloides que se pueden generar.

los diseños D-óptimos generales, también tienen su centro en la intersección de las asíntotas de las hipérbolas.

Así el diseño D-óptimo para tres factores a dos niveles está centrado en la intersección de los planos asintóticos de los $ED_{F(z)}$'s, y cumplen que $z_1 = z_4 = z_6 = z_7$ y $z_2 = z_3 = z_5 = z_8$, con

$$\beta_{123}^8 \det(I(\beta)) = 4^{12} n^8 (\beta_{123} R_1 R_2 R_3)^8 \lambda_1^4 \lambda_2^4 h(z_1)^4 h(z_2)^4, \quad (6.7)$$

donde $z_1 = z_0 - |\beta_{123}| R_1 R_2 R_3$ y $z_2 = z_0 + |\beta_{123}| R_1 R_2 R_3$. De nuevo, los resultados para los diseños de un factor, recogidos en la Tabla 3.1 pueden ser utilizados para calcular los valores D-óptimos de R_1 , R_2 y R_3 que maximizan (6.7). En este caso, las

expresiones $|\beta_1|R$ y $\beta_1^2 \det(I(\beta))$ se sustituyen por $|\beta_{123}|R_1R_2R_3$ y $\beta_{123}^2(\det(I(\beta)))^{1/4}$, obteniendo la Tabla 6.1.

A continuación presentamos un ejemplo donde se ilustra, como en secciones anteriores, el uso de esta tabla. Asumamos que $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$, $\beta_3 = 1$, $\beta_{12} = 3$, $\beta_{13} = 2$, $\beta_{23} = 6$ y $\beta_{123} = 6$. Los experimentos D-óptimos a dos niveles generales están centrados en el $x_0 = -(1/2, 1/3, 1)$, con $z_0 = \beta_0 - (\beta_{12}\beta_{13}\beta_{23}/\beta_{123}^2) = 0$, y asignan un octavo de los ensayos o pruebas a cada uno de los vértices de un hiper-rectángulo centrado en dicho x_0 , y donde los rangos verifican la relación $R_1R_2R_3 = .2572$; cuatro puntos soporte se encuentran situados en el $ED_{.1760}$, y los otros cuatro en el $ED_{.8240}$. En este caso la representación gráfica del problema se complica ya que se trata de un hiper-rectángulo cuyos vértices se encuentran situados en dos hiperboloides. El primero, formado por las dosis efectivas del 17.60 %, contiene al conjunto de puntos (1, 4, 6, 7) y el segundo, definido por las dosis del 82.40 %, al (2, 3, 5, 8).

Capítulo 7

Casos Particulares

En este capítulo calculamos los diseños a dos niveles cuadrados D-óptimos para modelos de respuesta binaria con dos factores. El estudio de este tipo de diseños viene justificado por el hecho de que en muchos casos los factores vienen recodificados, de forma que sus valores se mueven en el mismo rango de valores, por ejemplo si se encuentran estandarizados. Otra situación en la que pueden surgir estos diseños es cuando se aplica dos veces un mismo factor a dos niveles. Por ejemplo, en un experimento para determinar la resistencia de un determinado componente que será utilizado en un sistema de ensamblaje. Para medir dicha resistencia debemos aplicar en dos zonas específicas una determinada presión. La variable respuesta es de tipo binaria de forma que toma el valor (0) si no se produce rotura del componente y (1) en caso contrario. Nuestro objetivo es determinar un modelo que permita definir el porcentaje de veces que hay rotura en función de la presión que se va a ejercer. Para ello, se toman dos zonas sensibles del componente donde se aplicarán de forma simultánea dos presiones. Así, tenemos 4 posibles combinaciones, (zona 1-presión 1, zona 2-presión 1), (zona 1-presión 1, zona 2-presión 2), (zona 1-presión 2, zona 2-presión 1) y (zona 1-presión 2, zona 2-presión 2). Este experimento se aplica a n componentes de forma balanceada, esto es, aplicando el mismo número de veces cada uno de los tratamientos, por lo que $n_1 = n_2 = n_3 = n_4$. Bajo estos supuestos, el problema se puede modelar a través de un modelo lineal generalizado

donde la variable respuesta, y_i , se distribuye como una binomial y aunque únicamente se considera un factor, al aplicarlo dos veces se convierte en un problema de respuesta binaria con dos factores. Usando la notación de la Sección 2.4, tenemos que $k = 2$, $y_i|x_i, \beta \sim \text{Binomial}(n_i, p(x_i; \beta))$ con $x_i = (x_{1i}, x_{2i})$ donde $i = 1, \dots, 4$ y $p(x_i; \beta) = F(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) = F(z_i)$. Sabemos que aún cuando partimos de una estimación inicial de los parámetros, el diseño D-óptimo que se obtiene podríamos utilizarlo como el primero de una serie de diseños que se construyen de forma secuencial y que nos permiten definir qué niveles del factor son los adecuados para que el porcentaje de roturas sea mínimo.

En general, este problema es más sencillo que el tratado en los capítulos anteriores ya que en este caso el rango del diseño, al ser cuadrado, es tal que

$$R = R_1 = R_2 \quad (7.1)$$

Al eliminar una variable del problema se reduce su complejidad lo que permite hacer un estudio más exhaustivo de sus propiedades. Nótese que en cierta medida este capítulo es un caso particular del Capítulo 4 cuando se aplica la restricción (7.1).

En este capítulo queda demostrado que el diseño D-óptimo está centrado en el $ED_{.50}$, algo que ya intuíamos al ser un caso particular del diseño D-óptimo para dos variables de diseño. Además, se ha analizado el comportamiento del diseño D-óptimo al variar los parámetros del modelo y hemos calculado los diseños D-óptimos rotados un determinado ángulo con el fin de comparar los diseños D-óptimos a dos niveles con el diseño D-óptimo general para modelos de respuesta binaria con dos variables de diseño obtenido por Sitter y Torsney (1995a) que situaba sus puntos sobre las líneas ED .

7.1. Experimento factorial 2^2 cuadrado

Denotamos por (x_1, x_2, x_3, x_4) los vértices de un diseño cuadrado tal que $x_i = (x_{10} + a_{i1}R, x_{20} + a_{i2}R)$, donde $a_{i1} = (-1)^i$ y $a_{i2} = (-1)^{[i+1]/2}$ ($i = 1 \dots 4$). El símbolo [

] denota parte entera, (x_{10}, x_{20}) es un punto de la línea de dosis efectiva, denotada por $ED_{F(z_0)}$, donde el experimentador desea centrar el diseño, con $z_0 = \beta_0 + \beta_1 x_{10} + \beta_2 x_{20}$ y R la mitad del lado del cuadrado que lo define. Se denota por $ED_{F(z_i)}$, a la línea de dosis efectiva que contiene al punto x_i , tal que $z_i = z_0 + R(a_{i1}|\beta_1| + a_{i2}|\beta_2|)$. En la Figura 7.1 se muestra la estructura de estos diseños. Las líneas z_1 y z_4 son las dosis efectivas de menor y mayor probabilidad respectivamente.

Bajo estos supuestos, a partir de la Proposición (2.2) obtenemos el siguiente corolario.

Corolario 7.1. *Bajo el modelo binomial con $p(x_i, \beta) = F(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})$, el determinante de $I(\beta)$ para cualquier experimento soportado en los vértices de un cuadrado tal y como quedó definido en el párrafo anterior, centrado en la línea $ED_{F(z_0)}$, es de la forma*

$$(\beta_1 \beta_2)^2 \det(I(\beta)) = 4^2 n^3 (\beta_1 R)^2 (\beta_2 R)^2 \sum_{i < j < r} \lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r), \quad (7.2)$$

donde $z_1 = z_0 - |\beta_1|R - |\beta_2|R$, $z_2 = z_0 + |\beta_1|R - |\beta_2|R$, $z_3 = z_0 - |\beta_1|R + |\beta_2|R$ y $z_4 = z_0 + |\beta_1|R + |\beta_2|R$, y donde $\lambda_i = n_i/n$.

Una vez más comprobamos que la matriz de información depende de los parámetros desconocidos del modelo. Si consideramos que éstos son conocidos a priori, la expresión (7.2) depende únicamente de z_0 y R para una asignación balanceada, esto es $\lambda_i = \lambda = 1/4$, $\forall i = 1 \dots 4$. Derivando respecto a z_0 e igualando a cero obtenemos la expresión

$$\sum_{i < j < r} \left(\lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r) \left(\sum_{l=i,j,k} \frac{1 - e^{z_l}}{1 + e^{z_l}} \right) \right) = 0 \quad (7.3)$$

Para el link logit, el valor $z_0 = 0$ verifica la ecuación anterior por lo que se convierte en un punto crítico para el determinante de la matriz de información. Además, numéricamente se demuestra que este valor es un máximo. Por lo tanto, el diseño

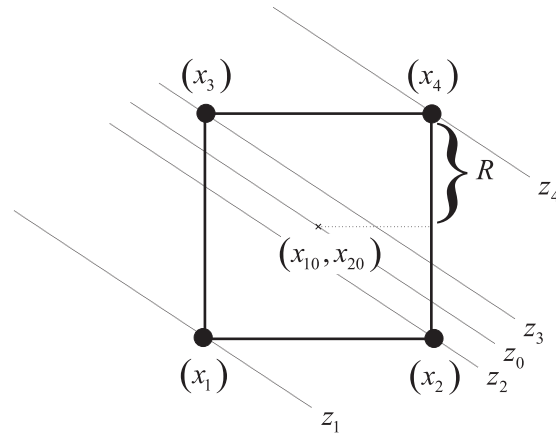


Figura 7.1: Experimento factorial 2^2 cuadrado.

factorial 2^2 D-óptimo cuadrado para un modelo de regresión logística está centrado en el $ED_{.50}$ independientemente de los parámetros del modelo. Esto es, dados β_1 y β_2 , existe un valor óptimo de R de forma que el diseño centrado en el $ED_{.50}$ maximiza el determinante de $I(\beta)$.

Una vez centrado el diseño en la dosis efectiva del 50% y suponiendo que conocemos una estimación inicial de los parámetros, el siguiente paso es determinar el valor óptimo para R . Derivando (7.2) respecto a R e igualando a cero, obtenemos una ecuación no lineal de la cual no se puede extraer la solución analítica. Esto nos lleva, una vez más, a aplicar métodos numéricos para su resolución. En la Figura 7.2 se presenta el gráfico de contorno del valor de R obtenido cuando maximizamos la ecuación (7.2) en función de β_1 y β_2 . Este gráfico no nos aporta una idea clara de lo que sucede con el valor óptimo de R . Únicamente podemos concluir que conforme los parámetros del modelo se aproximan a 0, el valor óptimo de R aumenta pero no hay una forma analítica de calcularlo a priori en función de β_1 y β_2 . Por lo tanto, si queremos determinar el diseño D-óptimo debemos llevar a cabo la maximización de (7.2) a través de métodos computacionales. Para solventar este problema, vamos a considerar el diseño desde otro punto de vista. Teniendo en cuenta que las rectas

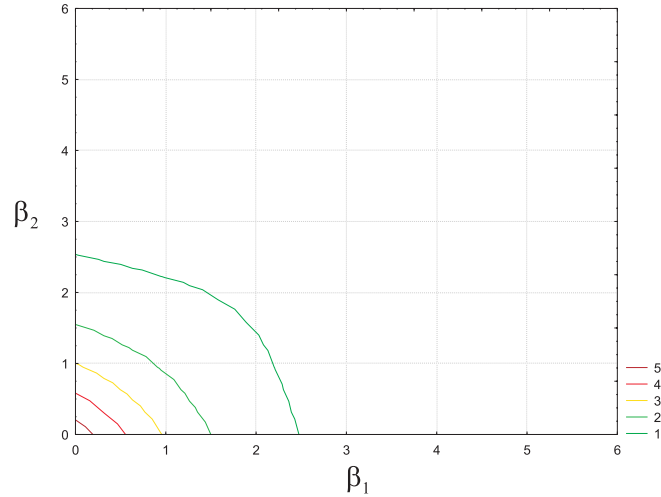


Figura 7.2: Gráfico de contorno del valor óptimo de R en función de β_1 y β_2 .

de probabilidad $z_1 = z_0 - R(|\beta_1| + |\beta_2|)$ y $z_4 = z_0 + R(|\beta_1| + |\beta_2|)$ definen las rectas sobre las que se apoyan los puntos de máxima y mínima probabilidad, si calculamos el valor óptimo para la expresión,

$$R' = R(|\beta_1| + |\beta_2|) \quad (7.4)$$

obtenemos dónde se deben situar los puntos del diseño. La Figura 7.3 presenta el gráfico de contorno para el valor óptimo de R' . Se puede observar que dichos valores son aproximadamente constantes cuando nos movemos sobre líneas de pendiente β_2/β_1 . Este cociente, que se denota por m , es clave para calcular los diseños D-óptimos cuadrados, pues mediante él se simplifican enormemente los cálculos, pudiendo obtener de forma rápida y sencilla dichos diseños. El estudio se ha limitado a la situación donde $|\beta_1| \geq |\beta_2|$. En otro caso basta con alternar el nombre de los parámetros y los resultados se siguen verificando. Analíticamente se puede demostrar que el valor óptimo de (7.4) sólo depende del parámetro m . Dicha demostración queda recogida en el siguiente corolario:

Corolario 7.2. *Dado un diseño de la forma que aparece en la Figura 7.1 centrado*

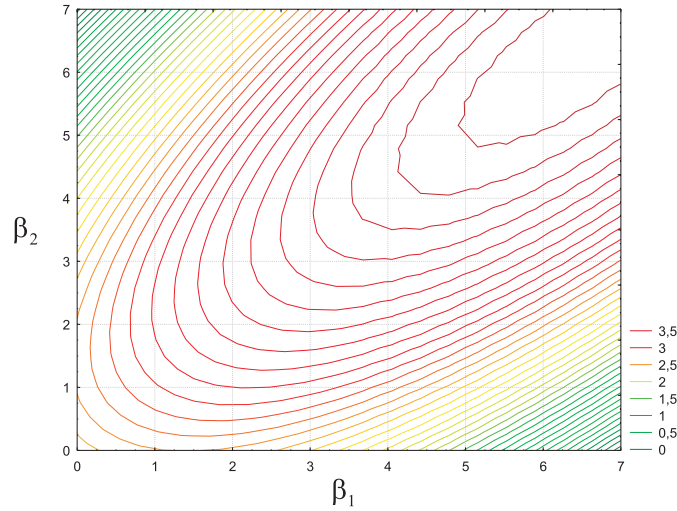


Figura 7.3: Gráfico de contorno del valor óptimo de la expresión $R(|\beta_1| + |\beta_2|)$ en función de β_1 y β_2 .

en el $ED_{.50}$ y asumiendo que $|\beta_1| \geq |\beta_2|$ y $m = \beta_2/\beta_1$ con $m \in [0, 1]$, se cumple que:

$$1. |z_i| = R \left| \left((-1)^i \beta_1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor} \beta_2 \right) \right| = \frac{R'}{(1 + |m|)} \left(1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor + i} |m| \right) \text{ con } i = 1, \dots, 4,$$

2. el valor de $R' = R(|\beta_1| + |\beta_2|)$ que maximiza el determinante de la matriz de información sólo depende de m .

3. Sea $R_{(\beta_1, \beta_2)}$ el valor de R para el modelo con parámetros β_1 y β_2 . Entonces

$$R_{(\beta_1, \beta_2 = m\beta_1)} = \frac{R'}{(|\beta_1| + |\beta_2|)} = \frac{R'}{|\beta_1|(1 + |m|)} = \frac{R_{(1, m)}}{|\beta_1|}$$

$$4. \text{Det}(I(\beta_1, \beta_2 = m\beta_1)) = \frac{\text{Det}(I(1, m))}{\beta_1^4}$$

La demostración de este corolario la podemos encontrar en el Apéndice 7.2.

Por lo tanto, a la hora de calcular un diseño factorial 2^2 D-óptimo cuadrado para un modelo de regresión logístico de parámetros (β_1, β_2) , lo único que necesitamos conocer es el cociente de los parámetros del modelo, $m = \beta_2/\beta_1$. Denotamos por $R_{(1,m)}$ a la mitad del rango del diseño factorial cuadrado para un modelo de regresión logístico de parámetros $(1, m)$. Teniendo en cuenta el apartado (3.) del Corolario 7.2, se verifica que $R_{(\beta_1, \beta_2=m\beta_1)} = R_{(1,m)}/|\beta_1|$, por lo que si conocemos el óptimo de $R_{(1,m)}$ conoceremos el óptimo de $R_{(\beta_1, \beta_2=m\beta_1)}$. Nótese que en el caso de $R_{(1,m)}$ sólo son necesarios los óptimos para los valores de m comprendidos entre 0 y 1, por lo que hemos conseguido simplificar de forma notable nuestro problema.

Para el cálculo del determinante haremos uso del apartado (4.) del Corolario 7.2 ya que si conocemos el valor óptimo de $R_{(1,m)}$, podemos determinar el valor óptimo del determinante del modelo con parámetros $(\beta_1, \beta_2 = m\beta_1)$ sin más que aplicar la fórmula que aparece en dicho apartado. Así, mediante el uso de la Tabla 7.1, podemos obtener los diseños D-óptimos para cualquier combinación de parámetros (β_1, β_2) . En dicha tabla podemos encontrar los valores óptimos de $R_{(1,m)}$, del $Det(I(1, m))$, de R' y la mínima dosis efectiva en tanto por cien, denotada por $\%ED_{min}$, del diseño que tratábamos de encontrar.

Del Corolario 7.2 podemos deducir que dado un modelo de regresión logística de parámetros (β_1, β_2) , los valores óptimos de $R_{(\beta_1, \beta_2)}$ y del $Det(I(\beta_1, \beta_2))$ dependen de los parámetros del modelo, mientras que el óptimo de R' depende únicamente de m . Supongamos que se desea obtener el diseño D-óptimo para un modelo de regresión logístico donde se conocen a priori buenas estimaciones de los parámetros del modelo, $\beta_1 = 5$ y $\beta_2 = 4$. Para calcular los valores óptimos $R_{(\beta_1, \beta_2)}$ y $Det(I(\beta_1, \beta_2))$, haremos uso de la Tabla 7.1. Primero calculamos $m = 4/5 = 0.8$ y, a partir de este valor, buscamos en la Tabla 7.1 los valores óptimos de $R_{(1,m)}$ y $Det(I(1, m))$. Por último, usando los apartados (3.) y (4.) del Corolario 7.2, se tiene que:

$$R_{(5,4)} = \frac{R_{(1,0.8)}}{|5|} = \frac{2.057876}{5} = 0.411575$$

$$Det(I(5, 4)) = \frac{Det(I(1, 0.8))}{5^4} = \frac{0.84896n^3}{5^4} = 1.3583n^3$$

Link Logit

m	$R_{(1,m)}$	$Det(I(1, m))$	R'	$\%ED_{min}$
0.00	1.8380	1.21355 n^3	1.8380	13.7288
0.01	1.8381	1.21351 n^3	1.8564	13.5118
0.05	1.8395	1.21260 n^3	1.9314	12.6591
0.10	1.8439	1.20974 n^3	2.0283	11.6268
0.15	1.8512	1.20491 n^3	2.1289	10.6321
0.20	1.8614	1.19799 n^3	2.2337	9.6762
0.25	1.8745	1.18883 n^3	2.3432	8.7609
0.30	1.8904	1.17722 n^3	2.4575	7.8889
0.35	1.9089	1.16290 n^3	2.5770	7.0636
0.45	1.9521	1.12471 n^3	2.8305	5.5698
0.50	1.9756	1.10000 n^3	2.9634	4.9109
0.55	1.9991	1.07091 n^3	3.0985	4.3168
0.60	2.0211	1.03698 n^3	3.2338	3.7915
0.65	2.0400	0.99785 n^3	3.3660	3.3375
0.70	2.0538	0.95333 n^3	3.4914	2.9558
0.75	2.0604	0.90354 n^3	3.6056	2.6452
0.80	2.0579	0.84896 n^3	3.7042	2.4029
0.85	2.0449	0.79053 n^3	3.7832	2.2245
0.90	2.0209	0.72954 n^3	3.8398	2.1046
0.95	1.9862	0.66754 n^3	3.8730	2.0372
1.00	1.9418	0.60612 n^3	3.8837	2.0161

Tabla 7.1: Valores óptimos asociados a un diseño factorial cuadrado centrado en el $ED_{.50}$, para un modelo de regresión logístico de parámetros $\beta_1 = 1$ y $\beta_2 = m$, donde, $R_{(1,m)}$ es la mitad del rango del diseño, $Det(I(1, m))$, es el determinante, $R' = R_{(1,m)}(1 + |m|)$ y $\%ED_{min}$ es la mínima dosis efectiva, en tanto por cien, del diseño.

Nótese que en ambos casos es necesario el uso de β_1 para calcular el valor final de $R_{(5,4)}$ y del $Det(I(5, 4))$. Por el contrario, el valor R' sólo depende del valor de m y por lo tanto, para este ejemplo el valor óptimo de R' es 3.7042, que es el mismo que para cualquier modelo de regresión logística con $m = 0.8$.

En la Figura 7.4 se muestran, de forma gráfica, los valores óptimos para $R_{(1,m)}$, $Det(I(1, m))$, R' y $\%ED_{min}$. Como ya se ha comentado, únicamente es necesario considerar $|m| \in [0, 1]$. En caso contrario, $|m| > 1$, se deben tomar los valores óptimos para $1/|m|$ y proceder como se ha indicado anteriormente.

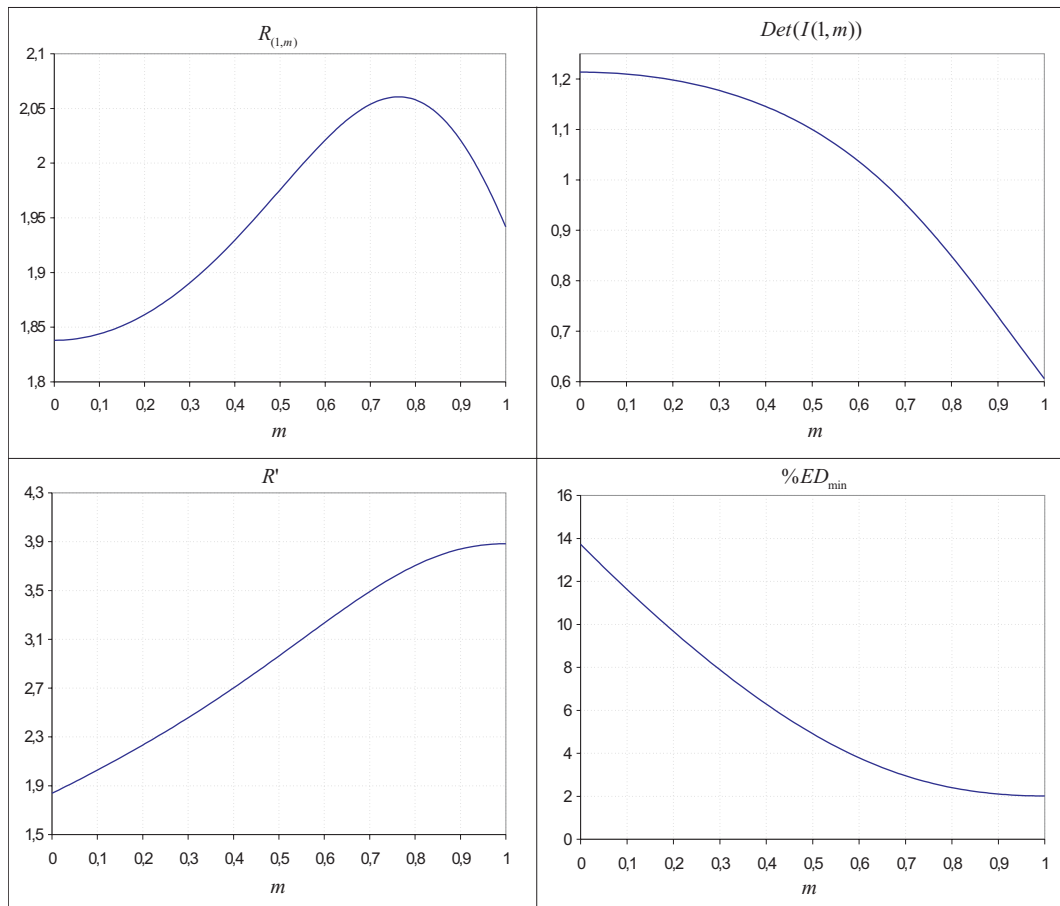


Figura 7.4: Valores óptimos asociados a un diseño factorial cuadrado para un modelo de regresión logístico de parámetros $\beta_1 = 1$ y $\beta_2 = m$, donde, $R_{(1,m)}$ es la mitad del rango del diseño, $Det(I(1,m))$, es el determinante, $R' = R_{(1,m)}(1 + |m|)$ y $\%ED_{min}$ es la mínima dosis efectiva, en tanto por cien, del diseño.

Con todo esto podemos decir que las características de estos diseños no se corresponden con las de los diseños factoriales en modelos lineales generales. Además, en ningún momento hemos tenido que acotar la región del diseño, como ya ocurría en capítulos anteriores. Esto se debe a que, para diseños de este tipo, el determinante de $I(\beta)$ está acotado, algo que no ocurre para modelos lineales normales. Así, un experimentador que desee llevar a cabo un experimento no debe definir acotaciones a priori, ya que el diseño D-óptimo en estos casos se puede calcular sin necesidad de ellas.

7.2. Experimento cuadrado con ángulo de giro θ

Sitter y Torsney (1995 a y b), obtienen los diseños D-óptimos globales para modelos de respuesta binaria con dos factores. Los puntos de estos diseños se sitúan sobre líneas de igual probabilidad como hemos indicado en las secciones anteriores. Los diseños de dos factores a dos niveles que hemos obtenido hasta el momento, tanto en este capítulo como en los anteriores, tienen sus lados paralelos a los ejes cartesianos. En esta sección vamos a llevar a cabo la rotación del diseño cuadrado de dos factores a dos niveles con el fin de poder comparar los diseños obtenidos por nosotros y los obtenidos por Sitter y Torsney. Nuestro objetivo es rotar los diseños hasta situarlos en la misma dirección que el diseño D-óptimo global y estudiar sus características. A continuación veremos que los diseños rotados se pueden introducir como casos particulares de los diseños sin rotar.

Vamos a considerar diseños a dos niveles centrados en la línea de dosis efectiva del 50%. Además, sin pérdida de generalidad, vamos a estudiar el comportamiento del diseño factorial rotado centrado en el punto $(-\beta_0/(\beta_1 + \beta_2), -\beta_0/(\beta_1 + \beta_2))$. Definimos como *eje del factorial* a la recta que une los puntos medios del cuadrado formado con los puntos del factorial, con la condición de que si se utiliza un diseño factorial sin rotar, este eje sea perpendicular al eje x ; y definimos como *eje del modelo*, a la recta correspondiente al $ED_{.50}$. Consideramos además, tal como se muestra en la Figura 7.5, en el supuesto de que $\beta_1\beta_2 < 0$, los ángulos siguientes:

α = ángulo que forman eje del modelo y la recta $x_2 = -\beta_0/(\beta_1 + \beta_2)$.

γ = ángulo que forman el eje del factorial y la recta $x_1 = -\beta_0/(\beta_1 + \beta_2)$.

θ = ángulo que forman el eje del modelo y el eje del factorial, esto es, $\theta = \alpha - \gamma$.

Sin pérdida de generalidad, vamos a tomar valores de β_1 y β_2 tales que $\beta_1\beta_2 > 0$. En el supuesto de que $\beta_1\beta_2 < 0$, el problema se puede resolver para el caso de $\beta_1\beta_2 > 0$ sin más que cambiar los subíndices a los parámetros, resolverlo en este nuevo caso y posteriormente trasladar los resultados al caso inicial. Bajo estos supuestos, y considerando ángulos medidos en sentido contrario a las agujas del reloj, se tiene que:

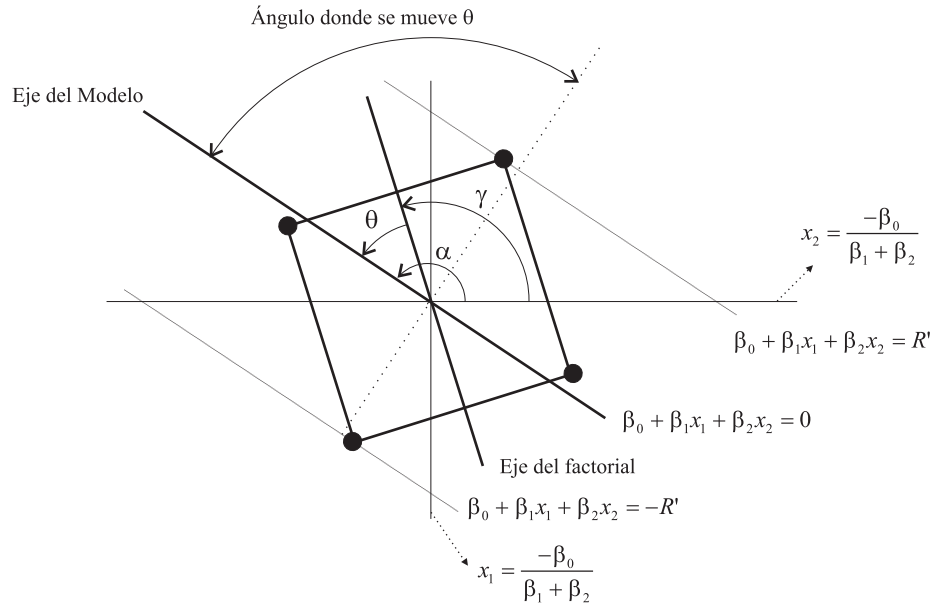


Figura 7.5: Diseño factorial rotado de dos factores a dos niveles centrado en el $ED_{.50}$, con puntos de mayor y menor probabilidad situados en el $ED_{1/(1+e^{\pm R'})}$.

1. $90^\circ \leq \alpha \leq 180^\circ$, con $\left(\sin \alpha = |\beta_1|/\sqrt{\beta_1^2 + \beta_2^2}\right)$ y $\left(\cos \alpha = -|\beta_2|/\sqrt{\beta_1^2 + \beta_2^2}\right)$.
2. $0^\circ \leq \theta \leq 90^\circ$ (o de forma equivalente, $\alpha - 90^\circ \leq \gamma \leq \alpha$), con $\gamma = 90^\circ$ ($\theta = \alpha - 90^\circ$) cuando se tiene el diseño factorial sin rotar y $\theta = 90^\circ$ ó $\theta = 0^\circ$ ($\gamma = \alpha - 90^\circ$ ó $\gamma = \alpha$ respectivamente) para el factorial orientado sobre los ejes de igual probabilidad.
3. $\sin \gamma = \sin(\alpha - \theta) = (|\beta_1| \cos \theta + |\beta_2| \sin \theta)/\sqrt{\beta_1^2 + \beta_2^2}$ y
 $\cos \gamma = \cos(\alpha - \theta) = (|\beta_1| \sin \theta - |\beta_2| \cos \theta)/\sqrt{\beta_1^2 + \beta_2^2}$.

En base a esta notación, los puntos del factorial en función del ángulo γ quedan

como sigue:

$$x'_i = (x_{1i}, x_{2i}) = \left(\frac{-\beta_0}{\beta_1 + \beta_2} + R' \frac{(-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} \cos \gamma + (-1)^i \sen \gamma}{(|\beta_1| + |\beta_2|) \sen \gamma + (|\beta_1| - |\beta_2|) \cos \gamma}, \right. \\ \left. \frac{-\beta_0}{\beta_1 + \beta_2} + R' \frac{(-1)^{i+1} \cos \gamma + (-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} \sen \gamma}{(|\beta_1| + |\beta_2|) \sen \gamma + (|\beta_1| - |\beta_2|) \cos \gamma} \right)$$

con

$$R' = R ((|\beta_1| + |\beta_2|) \sen \gamma + (|\beta_1| - |\beta_2|) \cos \gamma)$$

y en función de θ como:

$$x'_i = \left(\frac{-\beta_0}{\beta_1 + \beta_2} + R' \frac{|\beta_1| \left((-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} \sen \theta + (-1)^i \cos \theta \right) + |\beta_2| \left((-1)^i \sen \theta + (-1)^{\lfloor \frac{i-1}{2} \rfloor} \cos \theta \right)}{(\beta_1^2 + \beta_2^2)(\sen \theta + \cos \theta)}, \right. \\ \left. \frac{-\beta_0}{\beta_1 + \beta_2} + R' \frac{|\beta_1| \left((-1)^i \sen \theta + (-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} \cos \theta \right) + |\beta_2| \left((-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} \sen \theta + (-1)^i \cos \theta \right)}{(\beta_1^2 + \beta_2^2)(\sen \theta + \cos \theta)} \right) \quad (7.5)$$

Además, las rectas de máxima y mínima probabilidad son tales que $(\beta_0 + \beta_1 x_1 + \beta_2 x_2 = R')$ y $(\beta_0 + \beta_1 x_1 + \beta_2 x_2 = -R')$ respectivamente, donde

$$R' = R \left((\beta_1^2 + \beta_2^2)(\sen \theta + \cos \theta) \right).$$

Nótese que en el caso del diseño sin rotar, el problema se simplificaba de manera considerable cuando se tomaba el cociente entre los parámetros del modelo. En un principio, al rotar el diseño, esta característica no se verifica. Si comparamos ambos casos, y más concretamente las expresiones para R' , podemos comprobar que en el caso de la rotación no hay ninguna característica que a simple vista nos permita reducir la complejidad del problema. Por otro lado, llevando a cabo una transformación en el espacio del diseño rotado, obtenemos un diseño sin rotar sin mas que transformar el sistema de referencia (x_1, x_2) en uno nuevo denotado por (u_1, u_2) donde, el centro del diseño, $(-\beta_0/(\beta_1 + \beta_2), -\beta_0/(\beta_1 + \beta_2))$, pasa a ser el origen de coordenadas, y la perpendicular al eje factorial y el eje factorial se convierten en los nuevos ejes de referencia. Analíticamente, dado un punto (x_1, x_2)

en el espacio de origen, las coordenadas de dicho punto en el espacio transformado son

$$(u_1, u_2) = \left(\left(x_1 - \frac{-\beta_0}{\beta_1 + \beta_2} \right) \text{sen } \gamma - \left(x_2 - \frac{-\beta_0}{\beta_1 + \beta_2} \right) \text{cos } \gamma, \right. \\ \left. \left(x_1 - \frac{\beta_0}{\beta_1 + \beta_2} \right) \text{cos } \gamma + \left(x_2 - \frac{\beta_0}{\beta_1 + \beta_2} \right) \text{sen } \gamma \right)$$

siendo la transformación inversa,

$$(x_1, x_2) = \left(\frac{-\beta_0}{\beta_1 + \beta_2} + u_1 \text{sen } \gamma + u_2 \text{cos } \gamma, \frac{-\beta_0}{\beta_1 + \beta_2} - u_1 \text{cos } \gamma + u_2 \text{sen } \gamma \right)$$

con Jacobiano de la transformación igual a 1. En resumen, se ha llevado a cabo una traslación y una rotación, cuyo resultado es un diseño cuadrado centrado en el punto $(0, 0)$, y con rango R . El predictor lineal del modelo de respuesta binaria en el nuevo sistema de referencia es de la forma,

$$z = \sqrt{\beta_1^2 + \beta_2^2} (u_1 \text{cos } \theta + u_2 \text{sen } \theta)$$

y los puntos del factorial en función de γ vienen dados por

$$u'_i = (u_{1i}, u_{2i}) = \left(R' \frac{(-1)^i}{(|\beta_1| + |\beta_2|) \text{sen } \gamma + (|\beta_1| - |\beta_2|) \text{cos } \gamma}, \right. \\ \left. R' \frac{(-1)^{\lfloor \frac{i-1}{2} \rfloor + 1}}{(|\beta_1| + |\beta_2|) \text{sen } \gamma + (|\beta_1| - |\beta_2|) \text{cos } \gamma} \right)$$

Denotamos por $\beta_{1u} = (\sqrt{\beta_1^2 + \beta_2^2} \text{cos } \theta)$ y $\beta_{2u} = (\sqrt{\beta_1^2 + \beta_2^2} \text{sen } \theta)$ a los nuevos parámetros del modelo y por $R_{(\beta_{1u}, \beta_{2u})}$ al valor de la mitad del lado del diseño rotado un ángulo θ , de forma que

$$R_{(\beta_{1u}, \beta_{2u})} = \frac{R'}{|\sqrt{\beta_1^2 + \beta_2^2} \text{sen } \theta| + |\sqrt{\beta_1^2 + \beta_2^2} \text{cos } \theta|} = \frac{R'}{|\beta_{1u}| + |\beta_{2u}|}. \quad (7.6)$$

Por lo tanto, los puntos del diseño en función de R' y los parámetros del modelo son de la forma

$$u'_i = (u_{1i}, u_{2i}) = \left(R' \frac{(-1)^i}{|\beta_{1u}| + |\beta_{2u}|}, R' \frac{(-1)^{\lfloor \frac{i-1}{2} \rfloor + 1}}{|\beta_{1u}| + |\beta_{2u}|} \right) \quad (7.7)$$

Sustituyendo (7.6) en la expresión del determinante para un diseño factorial cuadrado, Corolario 7.1, se obtiene de forma directa el siguiente corolario.

Corolario 7.3. *El determinante de la matriz de información de Fisher para el diseño factorial con ángulo de rotación θ que observa en los puntos (7.5), para un modelo de regresión logística con dos variables de diseño es:*

$$(\beta_{1u}\beta_{2u})^2 \det(I(\beta)) = 4^2 n^3 (\beta_{1u} R_{(\beta_{1u}, \beta_{2u})})^2 (\beta_{2u} R_{(\beta_{1u}, \beta_{2u})})^2 \sum_{i < j < r} \lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r), \quad (7.8)$$

donde

$$h(z_i) = \frac{e^{z_i}}{(1 + e^{z_i})^2} \quad y \quad z_i = R' \frac{(-1)^i |\beta_{1u}| + (-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} |\beta_{2u}|}{(|\beta_{1u}| + |\beta_{2u}|)}$$

De este corolario podemos deducir que trabajar con diseños factoriales rotados un ángulo θ , es equivalente a trabajar con diseños factoriales sin rotar suponiendo que los parámetros del diseño son β_{1u} y β_{2u} . Así, el cálculo de los valores óptimos de R' y $R_{(\beta_{1u}, \beta_{2u})}$ depende únicamente de m_θ . En la Figura 7.6 se muestra cómo queda el diseño una vez hecho el cambio de base. El diseño factorial aparece en este caso como un diseño a dos niveles sin rotar. El eje del modelo en este caso pasa por el origen de coordenadas y tiene de parámetros β_{1u} y β_{2u} . De esta forma, reducimos el problema al caso recogido en la Sección 7.1. Esto permite la utilización de las gráficas que aparecen en la Figura 7.4 y de la Tabla 7.1 para calcular los valores óptimos de $R_{(\beta_{1u}, \beta_{2u})}$ y R' que definen el diseño a dos niveles D-óptimo. Denotamos por m_c a la relación de los parámetros para el factorial sin rotar, esto es $m_c = \beta_2/\beta_1$ con $m_c \in [0, 1]$, y por m_θ el valor de esta relación en el nuevo sistema de referencia,

$$m_\theta = \frac{\beta_{2u}}{\beta_{1u}} = \frac{\sen \theta}{\cos \theta} = \frac{\cos \gamma + |m_c| \sen \gamma}{\sen \gamma - |m_c| \cos \gamma}.$$

Sin pérdida de generalidad para poder hacer uso de la Tabla 7.1, debemos considerar las siguientes propiedades:

1. Si $m_\theta < 0 \implies$ tomar los valores correspondientes a $|m_\theta|$ en la Tabla 7.1.

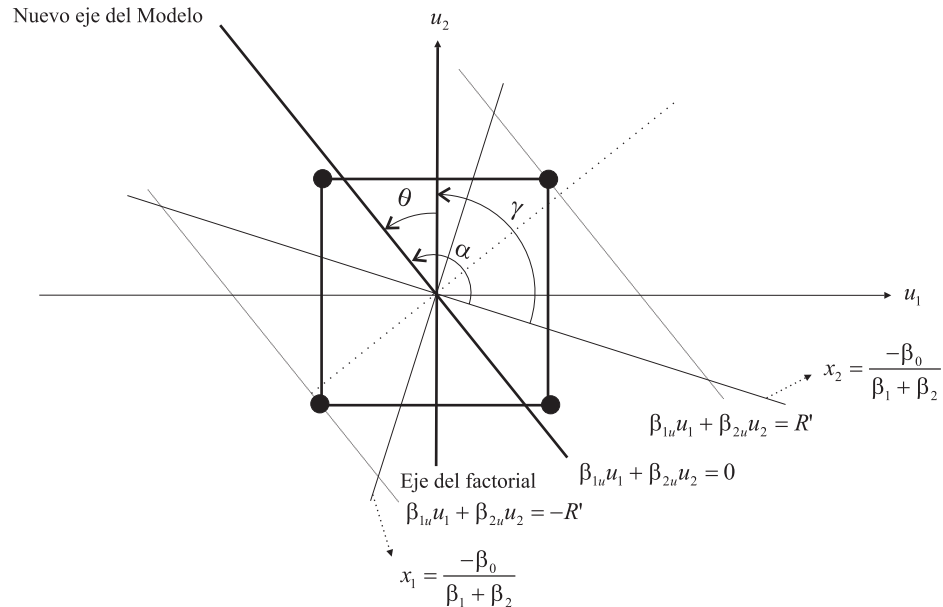


Figura 7.6: Diseño factorial rotado en el nuevo sistema de referencia (u_1, u_2) .

2. Si $|m_\theta| > 1 \implies$ tomar los valores correspondientes a $1/|m_\theta|$ en la Tabla 7.1.

Así, por ejemplo, cuando $\theta = 0^\circ$ ó $\theta = 90^\circ$, esto es, el eje del diseño coincide con el eje del modelo, proporcionando así el diseño orientado sobre rectas de igual probabilidad, se obtiene que $m_\theta = 0$ y, por lo tanto, equivaldría a un factorial sin rotar cuando uno de los parámetros del modelo es igual a cero.

Si $\theta = 0^\circ$, el diseño factorial tiene sus puntos dos a dos sobre las líneas de probabilidad constante. En este caso $m_\theta = 0$ y, usando la Tabla 7.1, el valor óptimo de R' es 1.838. Así, las líneas de probabilidad constante que soportan los puntos del diseño factorial cuadrado son, la $ED_{0.1373}$ y la $ED_{0.8627}$. Este diseño es muy interesante ya que, tal y como hemos introducido, los diseños factoriales D-óptimos obtenidos por Sitter y Torsney (1995 a y b) situaban los puntos dos a dos sobre rectas de igual probabilidad, aunque en ese caso el espacio debía estar acotado.

Una vez más hemos demostrado que la acotación no es necesaria obteniendo

de forma directa nuestro diseño. Esto es, rotando los diseños factoriales cuadrados centrados en el $ED_{.50}$, se sigue conservando la propiedad de no acotación, incluso cuando situamos los diseños en la dirección de los diseños D-óptimos obtenidos por Sitter y Torsney.

En general, el diseño D-óptimo a dos niveles con dos factores rotado un ángulo θ , se obtiene simplemente calculando el valor de m_θ que siempre se puede reducir al intervalo $[0, 1]$. Además se observa que los diseños se repiten de forma cíclica en el sentido de que hay diseños que son equivalentes en cuanto a su tamaño. Así, por ejemplo, para $\theta = \alpha - 90^\circ$ (diseño sin rotar), se tiene que $m_\theta = m_c$ y el valor óptimo de R es 1.838. En el caso de que $m_\theta = -m_c$ obtendríamos el mismo diseño. Así, despejando θ de la expresión

$$m_\theta = \frac{\text{sen } \theta}{\text{cos } \theta} = -m_c$$

obtenemos que $\theta = \text{arc tg}(-m_c)$ y por lo tanto, el diseño D-óptimo para este ángulo y para $\theta = 0^\circ$ son diseños equivalentes, entendiendo por equivalentes, a aquellos que tienen el mismo valor óptimo de R .

Con todo esto, hemos caracterizado los diseños D-óptimos a dos niveles rotados un ángulo θ y centrado en el $ED_{.50}$ para el modelo logístico perteneciente a los modelos de respuesta binaria. Nótese que mediante la Tabla 7.1 quedan definidos los diseños, siendo relativamente fácil poderlos determinar a partir del valor m_θ .

7.3. Comparación entre diseños D-óptimos

Como se mencionó en el Capítulo 4, el diseño D-óptimo para el caso de modelos con dos variables de diseño puede hacerse tan grande como se quiera. Esto hace que en estas situaciones se consideren espacios acotados, eligiendo así el diseño optimal para estas restricciones.

Sitter y Tornsey (1995) proponen trabajar con espacios de diseño en forma de paralelogramo, construidos a partir de las restricciones $b_1 \leq \beta_1(x_1 - \mu) + \beta_2(x_2 - \mu) \leq$

b_2 , es decir, rectas de igual probabilidad con b_1 y b_2 elegidas por el usuario, y de las restricciones $-1 \leq a_1\beta_1(x_1 - \mu) + a_2\beta_2(x_2 - \mu) \leq 1$, con a_1 y a_2 también elegidas por el usuario. El diseño D-óptimo en este caso no es directo y depende de las restricciones, de forma que:

1. Si $[-1.22, 1.22] \in [b_1, b_2]$ entonces los puntos dos a dos se sitúan sobre los $ED_{0.227}$ y $ED_{0.773}$.
2. Si $[b_1, b_2] \in [-1.22, 1.22]$ entonces los puntos dos a dos se sitúan sobre los $ED_{1/(1+e^{-b_1})}$ y $ED_{1/(1+e^{-b_2})}$.

Jia y Myers (2001) y Myers, Montgomery y Vining (2002) muestran además, que el criterio de la D-optimalidad es invariante a la localización y al cambio de ángulo en dicho paralelogramo.

7.3.1. Diseño D-óptimo de referencia

Con la intención de comparar los diseños factoriales óptimos, y teniendo en cuenta la necesidad de definir el diseño D-óptimo general con referencia al espacio de diseño que se considere, se propone la siguiente definición:

Definición 7.1. El diseño D-óptimo de referencia asociado a un diseño factorial D-óptimo es aquel diseño D-óptimo que se obtiene al considerar como espacio de diseño el menor de los paralelogramos en la orientación de los contornos ED que contiene a dicho factorial.

Así, dado un diseño factorial D-óptimo cuadrado para un modelo de regresión logístico, las restricciones para construir el diseño D-óptimo de referencia son de dos tipos. Las primeras son las líneas que soportan los puntos situados en la mayor y menor dosis efectiva. En la Figura 7.7 son las líneas correspondientes a las expresiones $\beta_0 + \beta_1x_1 + \beta_2x_2 = R'$ y $\beta_0 + \beta_1x_1 + \beta_2x_2 = -R'$. El segundo grupo de

restricciones vienen definidas a partir de las líneas perpendiculares a las anteriores y que pasan por los otros dos puntos, esto es,

$$\beta_0 \left(\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \right) - \beta_2 x_1 + \beta_1 x_2 = R'$$

$$\beta_0 \left(\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \right) - \beta_2 x_1 + \beta_1 x_2 = -R'$$

Estas cuatro líneas definen el espacio del diseño donde se va a calcular el diseño D-óptimo global, siguiendo la estructura presentada por Sitter y Torsney en sus artículos. Nótese que aún cuando el predictor lineal usado por estos autores es distinto al nuestro, podemos seguir usando las conclusiones a las que llegaron ya que, por un lado, el criterio de la D-optimalidad es invariante respecto de transformaciones lineales, y por otro, a que si consideramos que $\mu = -\beta_0/(\beta_1 + \beta_2)$ las restricciones que obtenemos usando el predictor lineal $z = \beta_1(x_1 - \mu) + \beta_2(x_2 - \mu)$ coinciden con las que hemos obtenido usando el predictor lineal $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. En la Figura 7.7 se representa un diseño factorial D-óptimo cuadrado, las restricciones que se imponen para buscar el diseño D-óptimo global y por último, el Diseño D-óptimo global asociado al diseño D-óptimo cuadrado. Nótese que en este caso el diseño D-óptimo global observa en puntos que se encuentran dos a dos en las dosis efectivas del 22.7% y del 77.3%. Ahora bien, hay que recordar que existen dos posibilidades: la primera si $[-1.22, 1.22] \in [b_1, b_2]$ y la segunda si $[b_1, b_2] \in [-1.22, 1.22]$. Los b_i , en el caso de diseños Factoriales D-óptimos cuadrados vienen dados por el valor óptimo de R' . Así, $b_1 = -R'$ y $b_2 = R'$. Además, en la Sección 7.1, se demostró que los valores óptimos de R' se mueven únicamente en el intervalo $[1.838, 3.883]$. Por lo tanto, en este caso siempre se cumple la segunda de las condiciones y de esta manera el diseño D-óptimo global siempre observa en puntos que se sitúan dos a dos sobre las dosis efectivas del 22.7% y del 77.3%.

Haciendo uso de este diseño de referencia utilizaremos la D-eficiencia de un diseño factorial D-óptimo tal como se describe en la siguiente definición para la comparación de su efectividad según el criterio de la D-optimalidad.

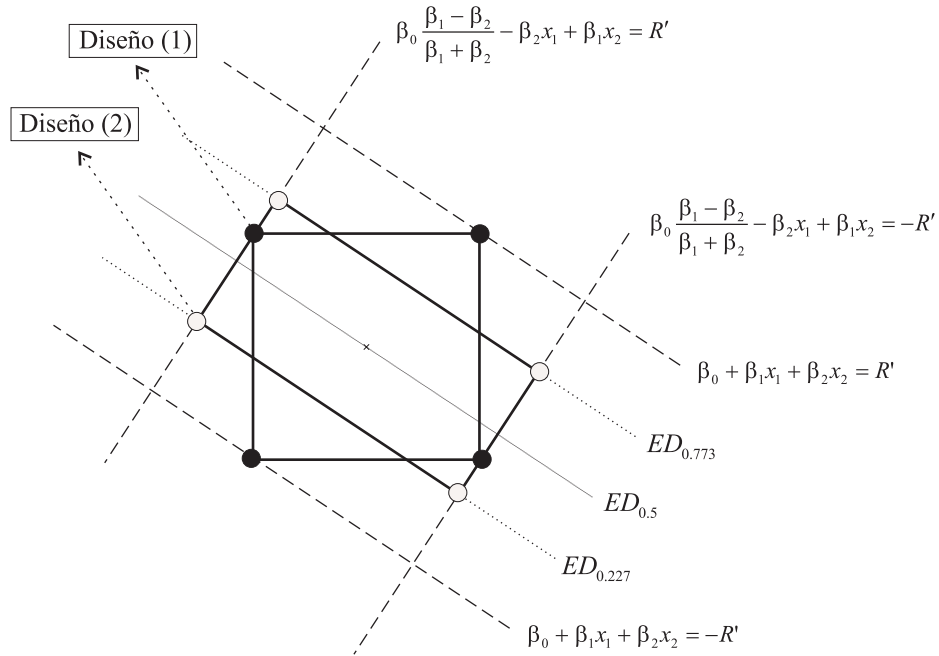


Figura 7.7: Diseño (1): Diseño Factorial D-óptimo. Diseño (2): Diseño D-óptimo de referencia del Diseño (1) según la Definición (7.1).

Definición 7.2. La D-eficiencia del diseño factorial D-óptimo, D_{efic} , se define como el cociente entre el valor del determinante de la matriz de información de Fisher del diseño factorial D-óptimo y el determinante del correspondiente diseño D-óptimo de referencia asociado.

En el siguiente apartado vamos a llevar a cabo el estudio de la D_{efic} de los diseños factoriales cuadrados sin rotar y rotados que hemos obtenido en la Secciones 7.1 y 7.2 respectivamente.

7.3.2. D-eficiencia de los diseños D-óptimos cuadrados

Teniendo en cuenta que el diseño factorial D-óptimo está centrado en el $ED_{.50}$ y que los puntos con mayor y menor probabilidad se encuentran en el $ED_{1/(1+e^{\pm R'})}$ con $R' \in [1.838, 3.883]$, tal como se muestra en la Figura 7.7, es fácil comprobar que el menor paralelogramo orientado sobre rectas de igual probabilidad que contiene a dichos factoriales se construye a partir de las siguientes restricciones:

$$\begin{aligned} -R' &\leq \beta_0 + \beta_1 x_1 + \beta_2 x_2 \leq R' \\ -R' &\leq \beta_0 \left(\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \right) - \beta_2 x_1 + \beta_1 x_2 \leq R' \end{aligned}$$

A partir de estas restricciones, el diseño D-óptimo de referencia observará en puntos de la forma:

$$\begin{aligned} x'_i = (x_{1i}, x_{2i}) = & \left(\frac{-\beta_0}{\beta_1 + \beta_2} + \frac{1.22(-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} \beta_1 + R'(-1)^i \beta_2}{\beta_1^2 + \beta_2^2}, \right. \\ & \left. \frac{-\beta_0}{\beta_1 + \beta_2} + \frac{1.22(-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} \beta_1 + R'(-1)^i \beta_2}{\beta_1^2 + \beta_2^2} \right) \end{aligned} \quad (7.9)$$

El valor de la D-eficiencia del diseño factorial D-óptimo con $m = \beta_2/\beta_1$ y R' que observa el mismo número de veces sobre cada uno de sus puntos con respecto al diseño D-óptimo de referencia cuyos puntos son de la forma (7.9), es:

$$\begin{aligned} D_{efic} = & \frac{(1 + m^2)^2}{(1 + |m|)^4} \left(\frac{R'}{1.22} \right)^2 (1 + e^{1.22})(1 + e^{-1.22}) \\ & \frac{e^{R'}}{(1 + e^{R'})^2} \frac{e^{\frac{(1-|m|)R'}{(1+|m|)}}}{\left(1 + e^{\frac{(1-|m|)R'}{(1+|m|)}}\right)^2} \left(\frac{e^{R'}}{(1 + e^{R'})^2} \frac{e^{\frac{(1-|m|)R'}{(1+|m|)}}}{\left(1 + e^{\frac{(1-|m|)R'}{(1+|m|)}}\right)^2} \right) \end{aligned} \quad (7.10)$$

Nótese que la ecuación (7.10) únicamente depende del valor de m , además de ser simétrica respecto de $m = 0$. En la Figura 7.8 se muestra el porcentaje de la D-eficiencia, $\%D_{efic}$, en función de los valores de m , de forma que, el valor de la

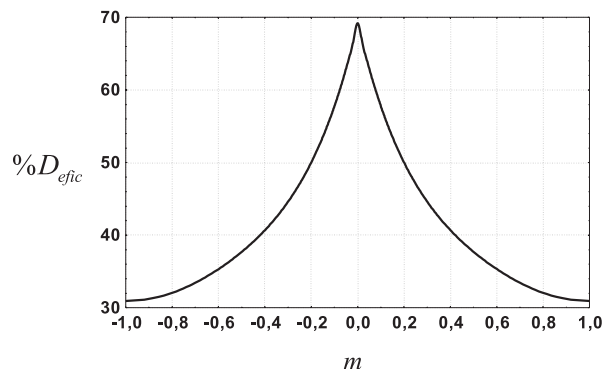


Figura 7.8: Evolución del porcentaje de la D-eficiencia para el diseño de dos factores a dos niveles 2^2 en función del parámetro m .

D-eficiencia aumenta al acercarnos a $m = 0$ (equivalente a un factorial orientado sobre las rectas de igual probabilidad) pasando de un 30.9626 % cuando $m = 1$ hasta alcanzar un máximo de 69.1934 % cuando $m = 0$. Hay que recordar que el diseño D-óptimo a dos niveles general (obtenido por Sitter y Torsney) situaba sus puntos sobre líneas de igual probabilidad. Esto hace que la D-eficiencia cuando $m = 0$ sea máxima, pues el diseño a dos niveles óptimo obtenido se encuentra en la misma dirección que el diseño D-óptimo global.

7.3.3. D-eficiencia de un diseño a dos niveles cuadrado rotado un ángulo θ

Sea un diseño a dos niveles rotado un ángulo θ que observa en puntos de la forma

$$x'_i = (x_{1i}, x_{2i}) = \left(\frac{-\beta_0}{\beta_1 + \beta_2} + R' \frac{(-1)^i \cos \theta + (-1)^{\lfloor \frac{i-1}{2} \rfloor} \sin \theta}{|\beta_1 \cos \theta + \beta_2 \sin \theta| + |-\beta_1 \sin \theta + \beta_2 \cos \theta|}, \right. \\ \left. \frac{-\beta_0}{\beta_1 + \beta_2} + R' \frac{(-1)^{\lfloor \frac{i-1}{2} \rfloor + 1} \cos \theta + (-1)^i \sin \theta}{|\beta_1 \cos \theta + \beta_2 \sin \theta| + |-\beta_1 \sin \theta + \beta_2 \cos \theta|} \right)$$

Trabajando con la transformación (u_1, u_2) se puede demostrar de forma directa

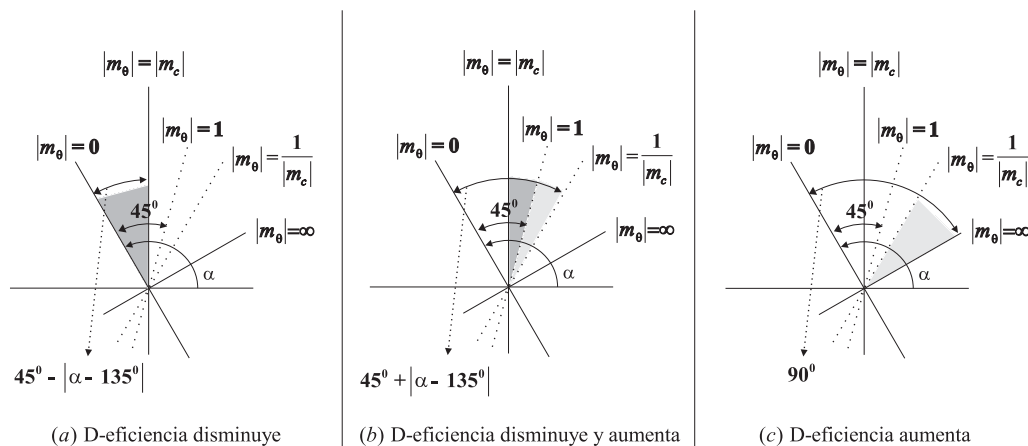


Figura 7.9: Evolución de la D-eficiencia del diseño factorial rotado. El arco denotado por θ indica dónde se mueve el eje del diseño, y el sentido de la rotación en cada caso.

que la D-eficiencia para este caso tiene la misma expresión que para el caso sin rotar. Esto es, la expresión (7.10) y la Figura 7.8 también son válidas en este caso. Únicamente hay que notar que en la expresión (7.10), m se corresponde con m_θ y R' con el valor que se puede deducir a partir de la expresión (7.6). Además, la D-eficiencia varía en función del ángulo de giro (θ) y del ángulo del eje del modelo (α). Para éste último debemos considerar dos casos, si $\alpha > 135^\circ$, donde $|\beta_2| > |\beta_1|$ y $|m_c| > 1$, o si $\alpha < 135^\circ$, en cuyo caso $|\beta_2| < |\beta_1|$ y $|m_c| < 1$. Esto hace que cuando partimos del caso sin rotar, esto es, $\theta = 45^\circ + |\alpha - 135^\circ|$, el valor de $|m_\theta| = |m_c|$ para ambos casos la única diferencia es que en el primer caso este valor es mayor de 1 y en el segundo es menor.

En la Figura 7.10 se muestra el comportamiento de la D-eficiencia en función del ángulo de giro θ cuando $\alpha < 135^\circ$. Si partimos del diseño cuyos lados son paralelos a las líneas ED , ($|m_\theta| = 0 \Rightarrow$ punto (1) en la Figura 7.10), y aumentamos el ángulo θ de 0° a ($\theta = 45^\circ$), la D-eficiencia disminuye tal y como se aprecia en la Figura 7.9-(a) y Figura 7.10-tramo (A). El diseño sin rotar se obtiene cuando $\theta = (45^\circ - |\alpha - 135^\circ|)$.

Conforme θ aumenta desde $\theta = 45^\circ$, la D-eficiencia también aumenta y los diseños

Ángulo de giro θ	D-eficiencia	m_θ	
		$\alpha < 135^\circ$ ($ m_\theta < 1$)	$\alpha > 135^\circ$ ($ m_\theta > 1$)
$[0^\circ, 45^\circ - \alpha - 135^\circ]$	Se empeora	$[0, m_c]$	$[\infty, m_c]$
$[45^\circ - \alpha - 135^\circ , 45^\circ]$	Se empeora	$[m_c , 1]$	$[m_c , 1]$
$[45^\circ, 45^\circ + \alpha - 135^\circ]$	Se mejora	$[1, 1/ m_c]$	$[1, 1/ m_c]$
$[45^\circ + \alpha - 135^\circ , 90^\circ]$	Se mejora	$[1/ m_c , \infty]$	$[1/ m_c , 0]$

Tabla 7.2: Comportamiento de la D-eficiencia y del parámetro m_θ respecto del parámetro m_c en función del ángulo θ .

obtenidos son equivalentes a los calculados en el Tramo (A) de la Figura 7.10. Esto se debe a que, como ya hemos comentado, cuando $|m_\theta| > 1$, para hacer uso de la Tabla 7.1, debemos tomar el valor correspondiente a $1/|m_\theta|$, ya que los diseños en ambos casos son equivalentes.

Como podemos ver la D-eficiencia del diseño sin rotar no siempre es mínima. Dicha D-eficiencia está en función del ángulo α y por consiguiente de los valores de β . En la Figura 7.10 este comportamiento queda reflejado en el valor de la D-eficiencia para el valor de $|m_c|$ y $1/|m_c|$.

En la Tabla 7.2 se muestra el comportamiento de la D-eficiencia en función del ángulo de giro y de la posición del eje del modelo (definida a partir del ángulo α). Las dos últimas columnas representan el valor de m_θ cuando $\alpha < 135^\circ$ y $\alpha > 135^\circ$ respectivamente en función del ángulo θ (primera columna). Se verifica que en ambos casos el comportamiento del valor m_θ es el mismo para las dos posibles alternativas de elección del valor α . La única diferencia consiste en la forma en que se debe interpretar la Figura 7.10. Así, en el caso de que $\alpha < 135^\circ$, comenzamos de izquierda a derecha, mientras que si $\alpha > 135^\circ$ la lectura se hará de derecha a izquierda. Nótese que en los dos casos el comportamiento coincide; en un primer instante la D-eficiencia es máxima, desciende hasta el mínimo, que se alcanza para $\theta = 45^\circ$, y vuelve a aumentar hasta el máximo.

En resumen, tanto los diseños del Tramo (A) como los del Tramo (B) de la Figura 7.10 son equivalentes y por lo tanto, podríamos reducir el estudio al caso en el que

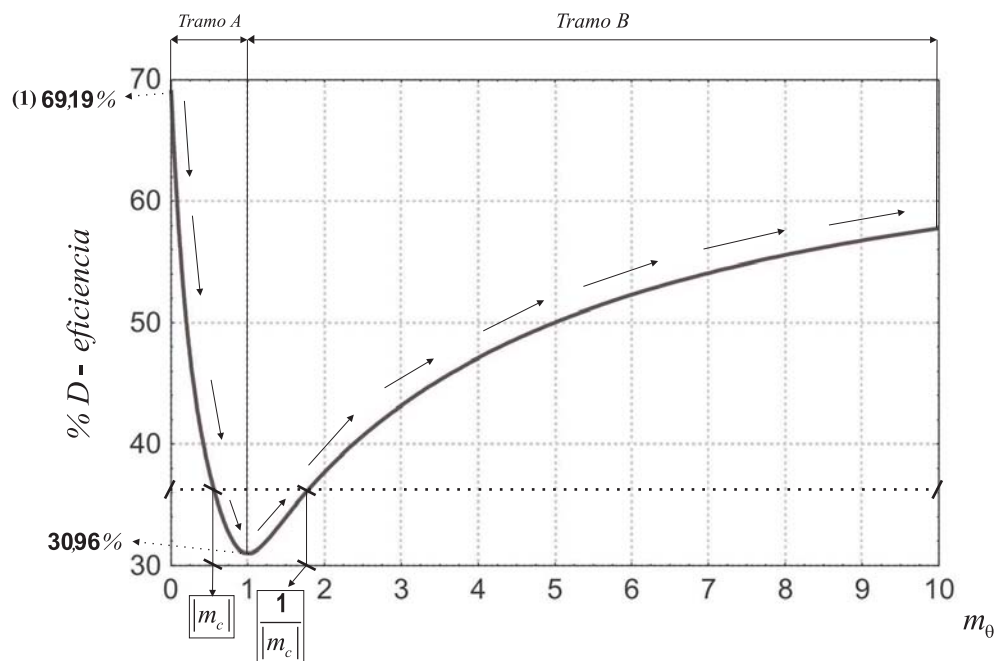


Figura 7.10: Porcentaje de la D-eficiencia en función del valor m_θ .

$|m_\theta| < 1$ lo cual nos permite hacer uso de la Tabla 7.1.

Capítulo 8

Conclusiones y líneas futuras

En esta tesis hemos tratado el problema del cálculo de los diseños D-óptimos a dos niveles para modelos de respuesta binaria en el caso de una, dos y tres variables de diseño. Uno de las conclusiones más importantes que se pueden extraer de los resultados obtenidos es que los experimentos factoriales a dos niveles y los factoriales fraccionales no son los candidatos más apropiados cuando se desea estimar un modelo de respuesta binaria. Esto se debe a que, en general, no se pueden aplicar como bloques para la exploración de superficies respuesta, ya que sus dimensiones se salen de las regiones de interés, ni como experimentos de escrutinio, debido a que observan en puntos donde $p(x_i; \beta)$ está muy próximo a 0 ó 1. Por otro lado, resultan de gran interés como punto de partida para evaluar la eficiencia de experimentos más realistas, y como una forma de ver que las propiedades de los factoriales a dos niveles enumeradas en el capítulo de introducción, no se extienden fuera del conjunto de modelos lineales normales.

En particular, hemos ilustrado cómo en el conjunto de modelos de respuesta binaria de primer orden, el determinante de la matriz de información de los experimentos a dos niveles depende de forma compleja, de la localización de sus puntos centrales, de sus dimensiones, de su orientación relativa a las curvas de nivel de la superficie respuesta, de la cantidad de réplicas y de la presencia o ausencia de interacciones.

En cuanto a la información que aportan los experimentos en modelos de respuesta binaria podemos destacar los siguientes puntos:

- Los diseños con asignaciones no balanceadas tienen un determinante mayor que en el caso de asignaciones balanceadas.
- Los experimentos factoriales fraccionales y los experimentos que varían un factor cada vez pueden tener mayor determinante y, por lo tanto, aportar mayor información para construir los modelos de respuesta binaria que los diseños factoriales completos con el mismo número de ensayos.
- Los experimentos factoriales y factoriales fraccionales con matrices de diseño no ortogonales, pueden tener mayor determinante que los diseños factoriales y factoriales fraccionales con matrices de diseño ortogonales.
- Las fracciones de un factorial en un modelo de respuesta binaria dejan de ser estadísticamente equivalentes.

También se llevó a cabo un estudio de los diseños D-óptimos cuando se incluyen términos de interacción en el predictor lineal de los modelos de respuesta binaria. Se comprobó que las propiedades que habíamos obtenido al trabajar con predictores lineales sin interacción variaban considerablemente. Esto sorprende por el hecho de que en el caso de modelos lineales normales, las propiedades de los diseños óptimos eran las mismas tanto para el caso del modelos con interacciones como sin ellas. Una vez más demostramos que en modelos de respuesta binaria hay que tener cuidado al elegir los diseños con los que trabajamos puesto que muchas veces estamos considerando diseños que se comportan de forma fiable en modelos normales.

En cuanto a la comparativa entre diseños D-óptimos a dos niveles, nos podríamos preguntar si debíamos realizar o no un estudio de eficiencia respecto a los diseños D-óptimos generales. Nótese que, para llevar a cabo una comparación clara, deberíamos calcular tanto los experimentos D-óptimos a dos niveles como los D-óptimos generales para regiones elegidas cuidadosamente, como las que definieron Sitter y Torsney (1995a), y posteriormente comparar los determinantes de las correspondientes matrices de información. Dado que los resultados dependen en gran medida de las regiones elegidas, decidimos no llevar a cabo dicha comparativa en el caso

general. Únicamente, en el Capítulo 7, se introducen los conceptos de la D-eficiencia y se lleva a cabo una comparativa entre ambos diseños para un caso muy particular que por sus propiedades permitía rotar el diseño en la dirección del que obtuvieron Sitter y Torsney (1995a).

A continuación se enumeran brevemente algunas líneas de investigación futuras en relación al trabajo realizado en esta tesis:

1. El cálculo de los diseños D-óptimos a dos niveles y el estudio de los factoriales se puede extender a otros modelos de la familia exponencial. El más importante es el modelo de Poisson. Algunos autores ya han estudiado este problema de forma general. Ver, por ejemplo, Chiacherrini (1996); van Mullekom y Myers (2001); Wang, Myers, Smith y Ye (2004).
 2. Otra variación importante está basada en asignar una distribución a priori a los parámetros del modelo. Con ello obtendríamos diseños D-óptimos Bayesianos. Ver, por ejemplo, Chaloner y Larntz (1989); Minkin (1993); Atkinson, Chaloner, Herzberg y Juritz (1993); Dumouchel y Jones (1994); Letsinger (1995); Chaloner y Verdinelli (1995).
 3. Una última extensión consiste en tomar una estimación puntual de los parámetros, o considerar el caso bayesiano, y obtener un diseño D-óptimo secuencialmente. Este tipo de diseños en modelos de respuesta binaria han sido estudiados por los siguientes autores: Box y Lucas (1959); Abdelbasit y Plackett (1983); Wu (1985); Minkin (1987); Myers, Myers, Carter, y White (1996); Letsinger (1995); Sitter y Wu (1999); Müller y Ponce (1996); McLeish (1990); Lin, Myers y Ye (2000); Schwartz et al. (2001).
-

Apéndices

Apéndice A

Demostración de la Proposición 2.1

Lema A.1. Sean C, D y E matrices de dimensión $n \times n$, y denotamos por c_i, d_i y e_i a las i -ésimas columnas de C, D y E respectivamente. Si, para algún k , $c_k = d_k + e_k$ y $c_i = d_i = e_i$ para $i = 1, \dots, k-1, k+1, \dots, n$, entonces

$$\det(C) = \det(D) + \det(E). \quad (\text{A.1})$$

Si $c_i = tc_j$ para $i \neq j$, donde t es un escalar conocido, entonces

$$\det(C) = 0. \quad (\text{A.2})$$

La primera parte de este lema está enunciado y probado como Teorema 13.7.2 en Harville (1997), mientras que la segunda parte es su lema 13.2.8. Al repetir de forma reiterada (A.1) en:

$$\det(I(\beta)) = \det \begin{pmatrix} \sum_i^q n_i h(z_i) & \sum_i^q n_i h(z_i) x_{1i} & \dots & \sum_i^q n_i h(z_i) x_{ki} \\ \sum_i^q n_i h(z_i) x_{1i} & \sum_i^q n_i h(z_i) x_{1i}^2 & \dots & \sum_i^q n_i h(z_i) x_{1i} x_{ki} \\ \dots & \dots & \dots & \dots \\ \sum_i^q n_i h(z_i) x_{ki} & \sum_i^q n_i h(z_i) x_{1i} x_{ki} & \dots & \sum_i^q n_i h(z_i) x_{ki}^2 \end{pmatrix}, \quad (\text{A.3})$$

se prueba que

$$\det(I(\beta)) =$$

$$\sum_{i_1=1}^q \sum_{i_2=1}^q \cdots \sum_{i_{k+1}=1}^q \det \begin{pmatrix} n_{i_1} h(z_{i_1}) & n_{i_2} h(z_{i_2}) x_{1i_2} & \cdots & n_{i_{k+1}} h(z_{i_{k+1}}) x_{ki_{k+1}} \\ n_{i_1} h(z_{i_1}) x_{1i_1} & n_{i_2} h(z_{i_2}) x_{1i_2}^2 & \cdots & n_{i_{k+1}} h(z_{i_{k+1}}) x_{1i_{k+1}} x_{ki_{k+1}} \\ \cdots & \cdots & \cdots & \cdots \\ n_{i_1} h(z_{i_1}) x_{ki_1} & n_{i_2} h(z_{i_2}) x_{1i_2} x_{ki_2} & \cdots & n_{i_{k+1}} h(z_{i_{k+1}}) x_{ki_{k+1}}^2 \end{pmatrix}, \quad (\text{A.4})$$

y debido a (A.2), únicamente los términos donde los $(k+1)$ valores de i_j sean distintos, tienen un determinante distinto a 0, y por lo tanto

$$\det(I(\beta)) = \sum_{(i_1, \dots, i_{k+1})} \det \begin{pmatrix} n_{i_1} h(z_{i_1}) & n_{i_2} h(z_{i_2}) x_{1i_2} & \cdots & n_{i_{k+1}} h(z_{i_{k+1}}) x_{ki_{k+1}} \\ n_{i_1} h(z_{i_1}) x_{1i_1} & n_{i_2} h(z_{i_2}) x_{1i_2}^2 & \cdots & n_{i_{k+1}} h(z_{i_{k+1}}) x_{1i_{k+1}} x_{ki_{k+1}} \\ \cdots & \cdots & \cdots & \cdots \\ n_{i_1} h(z_{i_1}) x_{ki_1} & n_{i_2} h(z_{i_2}) x_{1i_2} x_{ki_2} & \cdots & n_{i_{k+1}} h(z_{i_{k+1}}) x_{ki_{k+1}}^2 \end{pmatrix}, \quad (\text{A.5})$$

donde la sumatoria es sobre todas las $q(q-1)\dots(q-k+1)$ posibles variaciones de tamaño $k+1$ de los primeros q enteros, $\{1, 2, \dots, q\}$. Esto puede ser re-escrito como

$$\det(I(\beta)) = \sum_{\{i_1, \dots, i_{k+1}\}} \det(X'_{i_1, \dots, i_{k+1}} H_{i_1, \dots, i_{k+1}} X_{i_1, \dots, i_{k+1}}), \quad (\text{A.6})$$

donde ahora la sumatoria es sobre todos los $\binom{q}{k+1}$ posibles subconjuntos de tamaño $k+1$ de los primeros q enteros, y donde $H_{i_1, \dots, i_{k+1}}$ denota la matriz diagonal de dimensión $(k+1) \times (k+1)$, y cuyos términos de la diagonal son $(n_{i_1} h(z_{i_1}), \dots, n_{i_{k+1}} h(z_{i_{k+1}}))$. El hecho de que el determinante del producto de matrices es el producto de sus determinantes y que el determinante de una matriz cuadrada y de su traspuesta sean iguales, prueba (2.7).

Apéndice B

Fracciones mitad de diseños 2^3 en orden estándar

En este apéndice tratamos de relacionar los diseños en orden estándar de Box, Hunter & Hunter (1978) con la notación utilizada a lo largo de toda la tesis para el cálculo de los diseños D-óptimos en modelos de respuesta binaria. Nótese que el orden estándar es el más utilizado en problemas de diseño experimental. En nuestro caso, el uso de dicho orden hizo que los cálculos computacionales se complicaran, lo que nos obligó a tomar otra alternativa. Vamos a exponer cómo resolvimos los problemas de notación y a qué fueron debidos. Para ilustrarlo vamos a llevar a cabo un estudio de las fracciones un medio de los diseños factoriales 2^3 . Estos diseños están formados por cuatro puntos de los ocho que lo forman. La razón por la que hemos elegido este tipo de diseños, es su importancia en modelos lineales normales y a que los diseños D-óptimos para modelos de respuesta binaria con tres factores están formados por 4 puntos. Además, como dato relevante de la tesis, hemos demostrado que las fracciones un medio estándar para los modelos lineales normales, dejan de ser óptimas en modelos de respuesta binaria.

Un diseño factorial está en orden estándar, cuando la primera columna de la

matriz del diseño consta de todos los valores iguales a 1; la segunda, de (-1) y (+1) alternados; la tercera, de parejas de (-1) y (+1) alternadas; la cuarta, de cuatro (-1) seguidos de cuatro (+1) y así sucesivamente. Siguiendo estas reglas, la matriz del diseño estandarizada para un factorial 2^3 en orden estándar viene dada por

$$A = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & -1 \\ 1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 \\ 1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 \end{pmatrix} \quad (\text{B.1})$$

En la Figura (B.1) se muestra un diseño factorial 2^3 completo en orden estándar. A cada uno de los vértices se le ha asignado un número que coincide con el orden de los puntos en la matriz (B.1). Por ejemplo, el punto (3) se corresponde con la fila 3 de la Matriz A .

En modelos de respuesta binaria, cada uno de los puntos del diseño pertenece a una dosis efectiva o ED , que en el caso de tres factores se corresponde con un plano. A la hora de calcular los diseños D-óptimos a dos niveles para estos modelos, el determinante de la matriz de información depende de los β_i a través de los valores del predictor lineal z . Para cada uno de los 8 puntos del diseño 2^3 completo tenemos un z_i de la forma:

$$z_i = z_0 + (-1)^i |\beta_1| R_1 + (-1)^{[(i-1)/2]+1} |\beta_2| R_2 + (-1)^{[(i-1)/4]+1} |\beta_3| R_3$$

El determinante de la matriz de información para un modelo logístico de tres factores viene dado por la expresión (5.1), que depende de los z_i y de los $|\beta_i| R_i$. El proceso de optimización para calcular el diseño D-óptimo, se llevó a cabo sobre los valores de $|\beta_i| R_i$. Obtuvimos que el mejor diseño D-óptimo estaba formado por los puntos (2345). Este resultado se obtiene cuando todos los parámetros del modelo son positivos, puesto que así lo hemos impuesto con el valor absoluto. Pero ¿qué ocurriría

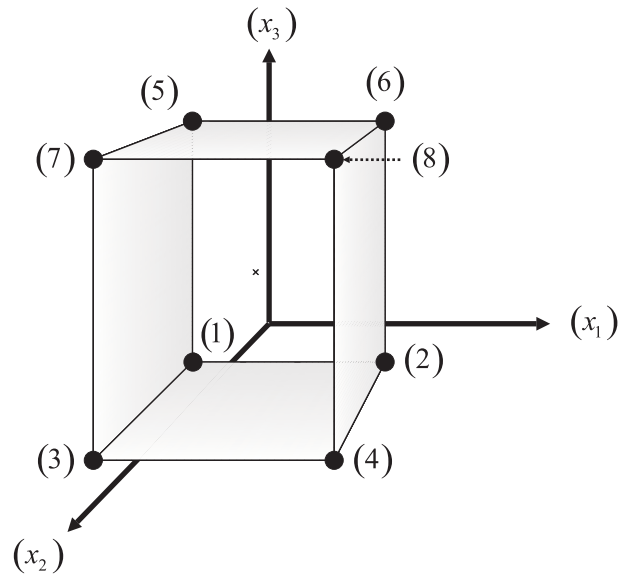


Figura B.1: Diseño 2^3 en orden estándar.

si algún β_i cambiara de signo? Está claro que el diseño seguiría siendo el mismo ya que todos los β_i s están en valor absoluto. Así, dado un conjunto de parámetros, el diseño D-óptimo no varía cuando lo hacen los signos de dichos parámetros. ¿Cuál es la razón para que esto ocurra? Si indagamos un poco más en el problema, nos damos cuenta de que al variar el signo de un parámetro, los planos ED sobre los que se apoyan los puntos del diseño siguen asociados a la misma probabilidad que tenían cuando todos los parámetros eran positivos. Lo que está sucediendo es que los puntos del diseño y los planos ED cambian de posición, pero no así las probabilidades asociadas a los planos. Veamos un ejemplo. Supongamos que $\beta_i > 0$ con $i = 1, \dots, 4$. Como ya hemos comentado, el diseño D-óptimo para el modelo logístico es aquel que observa en los puntos (2345) (ver Tabla 5.4). La matriz estandarizada

de este diseño viene dada por

$$A = \begin{pmatrix} 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & -1 \\ 1 & -1 & -1 & +1 \end{pmatrix}, \quad (\text{B.2})$$

y su representación gráfica se muestra en la Figura B.2 caso $\beta_1 > 0$. Supongamos ahora que β_1 cambia de signo. En este caso, al calcular el diseño D-óptimo mediante la optimización de la expresión (5.1) obtenemos el diseño (2345) que coincide con el anterior. Esto es lógico puesto que en dicha expresión el cambio de signo lo ha anulado el valor absoluto. Ahora bien, al representar el diseño en el espacio, debemos utilizar la expresión del predictor lineal sin valor absoluto, esto es

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Por supuesto, los planos ED , definidos a partir de los z_i , cambian de dirección al variar el signo del β_1 . Si calculamos el diseño haciendo uso de la Tabla 5.4 y lo representamos en el espacio tridimensional, obtenemos el diseño de la Figura B.2 para $\beta_1 < 0$. Nótese que el diseño para $\beta_1 > 0$ es distinto al de $\beta_1 < 0$, puesto que sus puntos se distribuyen de forma diferente. Ahora bien, ambos diseños tienen el mismo determinante y sitúan sus puntos en los mismos ED 's. Esta propiedad es la que nos permitió trabajar con valores absolutos en los parámetros del modelo.

En resumen, dada una estimación de los parámetros del modelo, los diseños D-óptimos para modelos de respuesta binaria son invariantes respecto al signo de dichos parámetros. Esto es, son invariantes en el sentido de que tienen el mismo determinante y que los puntos se sitúan sobre las mismas dosis efectivas o planos ED .

Veamos ahora lo que ocurre cuando hacemos uso de la notación estándar. En este caso, si todos los parámetros son positivos, el diseño D-óptimo coincide con el (2345). Por otro lado, si $\beta_1 < 0$ el diseño D-óptimo en notación estándar pasa a ser

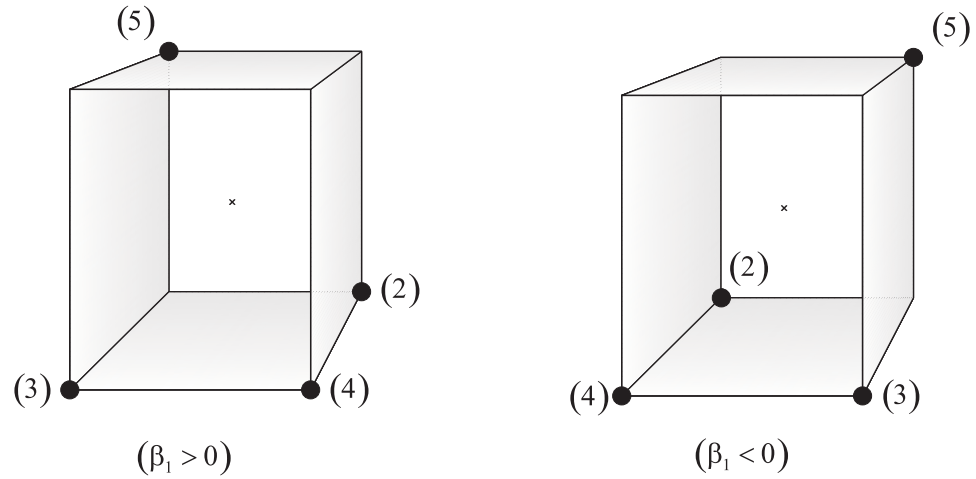


Figura B.2: Diseños D-óptimos para el modelo logístico con tres factores cuando el signo de β_1 cambia.

el (1346). La matriz de información para este diseño viene dada por

$$A = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & -1 \\ 1 & +1 & -1 & +1 \end{pmatrix} \quad (\text{B.3})$$

Es fácil comprobar que (B.3) se obtiene al multiplicar la segunda columna de (B.2) por (-1) ya que el parámetro que cambió de signo fue β_1 . Si el cambio se hubiese producido en β_2 hubiésemos multiplicado la tercera columna de (B.2) por (-1) y así con el resto. De esta forma, podemos decir que tanto el diseño (2345) como el (1346) son equivalentes ya que sitúan sus puntos en los mismos ED 's y tienen el mismo determinante.

En este punto, nos preguntamos si hay alguna forma de obtener todos los diseños equivalentes al (2345). Parece que sí por las reglas de multiplicación que vimos anteriormente. Pero además nos surge la duda de si todos los diseños de cuatro puntos son equivalentes al diseño (2345). En este sentido, ya conocemos la respuesta

puesto que hemos demostrado que los fraccionales un medio para los modelos lineales normales tienen menor determinante que el diseño D-óptimo (2345). A partir de aquí planteamos un estudio computacional exhaustivo de todas las posibilidades de elegir cuatro puntos de los ocho de que consta un diseño 2^3 . Como resultado obtuvimos que los diseños de cuatro puntos se pueden dividir en cinco clases, diferenciadas por la forma geométrica que toman los cuatro puntos del diseño. A continuación se enumeran estas cinco clases junto con una tabla que nos permite calcular los diseños equivalentes al diseño de referencia. Estamos considerando diseño de referencia a aquel que muestra el óptimo cuando se consideran todos los parámetros positivos. Además el estudio se lleva a cabo considerando todas las posibilidades, tanto de los signos de los parámetros como sobre qué ED se sitúa el centro del diseño.

Clase (a)

Puntos que se sitúan en los vértices de las diagonales de caras opuestas del diseño 2^3 , tal como se muestra en la Figura B.3. Estos puntos se corresponden con la elección de puntos del 2^{3-1} por defecto en modelos lineales normales.

Hay dos diseños de referencia, el (1467) y (2358). Además esta clase está formada únicamente por estos dos diseños.

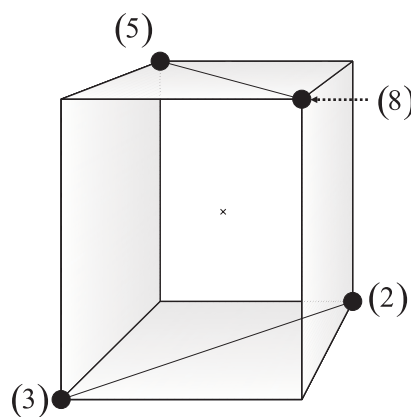


Figura B.3: Diseño de referencia (2358) de la Clase (a).

$\beta_3 > 0$				
	$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
> 50 %	(2358)	(1467)	(1467)	(2358)
< 50 %	(1467)	(2358)	(2358)	(1467)

$\beta_3 < 0$				
	$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
> 50 %	(1467)	(2358)	(2358)	(1467)
< 50 %	(2358)	(1467)	(1467)	(2358)

Tabla B.1: Diseños D-óptimos de la Clase (a) a utilizar según si están centrados por encima o por debajo del 50 % y según el signo de los parámetros.

La Tabla B.1 muestra los mejores diseños de la clase (a) a utilizar según si están centrados por encima o por debajo del $ED_{.50}$ y según el signo de los parámetros. Los diseños correspondientes a una misma fila tendrán sus puntos situados en los mismos ED .

Nótese que si usamos el diseño (2358) pero consideramos que uno de los parámetros es negativo entonces el diseño se transformaría en el (1467). Si dos de los parámetros son negativos seguirá siendo el (2358), y si los tres parámetros son negativos se transformará en el (1467).

Clase (b)

En este caso, la disposición de los puntos es la de tres puntos en una de las caras y el otro punto en la cara opuesta, justo en el vértice opuesto al que no se tiene en la cara de los tres puntos, tal como se muestra en la Figura B.4

Hay seis diseños de referencia. Por un lado los diseños (2345), (2357) y (2356) centrados por encima del $ED_{.50}$ y por otro, los diseños (4567), (2467) y (3467) centrados por debajo del $ED_{.50}$.

La Tabla B.2 muestra los mejores diseños de la clase (b) a utilizar según si están centrados por encima o por debajo del $ED_{.50}$ y según el signo de los parámetros. Los diseños correspondientes a una misma fila tendrán sus puntos situados en los mismos $ED's$.

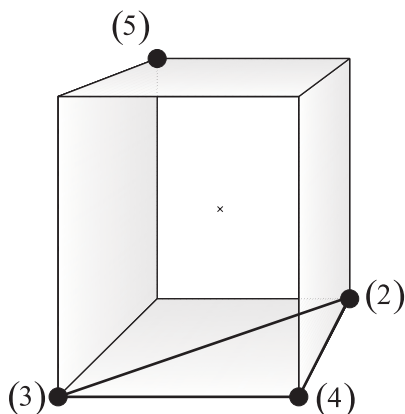


Figura B.4: Diseño de referencia (2345) de la Clase (b)

Veamos un ejemplo que ilustra el uso de la Tabla B.2. Supongamos que queremos elegir el mejor diseño centrado en el 60%. Los diseños que podríamos obtener en función del signo de los parámetros serían:

1. Si todos los parámetros son positivos, entonces deberíamos elegir el (2345).
2. Si $\beta_1 < 0$ y el resto positivos, entonces deberíamos elegir el (1346).
3. Si $\beta_2 < 0$ y el resto positivos, entonces deberíamos elegir el (1247).
4. Si $\beta_1 < 0$, $\beta_2 < 0$ y $\beta_3 > 0$, entonces deberíamos elegir el (1238).
5. Si $\beta_3 < 0$ y el resto positivos, entonces deberíamos elegir el (1678).
6. Si $\beta_3 < 0$, $\beta_1 < 0$ y $\beta_2 > 0$, entonces deberíamos elegir el (2578).
7. Si $\beta_3 < 0$, $\beta_2 < 0$ y $\beta_1 > 0$, entonces deberíamos elegir el (3568).
8. Si todos son negativos, entonces deberíamos elegir el (4567).

Clase (c)

		$\beta_3 > 0$			
		$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
> 50 %		(2345)	(1346)	(1247)	(1238)
		(2357)	(1468)	(1457)	(2368)
		(2356)	(1456)	(1478)	(2378)
< 50 %		(4567)	(3568)	(2578)	(1678)
		(2467)	(1358)	(2458)	(1367)
		(3467)	(3458)	(1258)	(1267)

		$\beta_3 < 0$			
		$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
> 50 %		(1678)	(2578)	(3568)	(4567)
		(1367)	(2458)	(1358)	(2467)
		(1267)	(1258)	(3458)	(3467)
< 50 %		(1238)	(1247)	(1346)	(2345)
		(2368)	(1457)	(1468)	(2357)
		(2378)	(1478)	(1456)	(2356)

Tabla B.2: Diseños D-óptimos de la Clase (b) a utilizar según si están centrados por encima o por debajo del 50 % y según el signo de los parámetros.

La clase (c) está formada por todos los diseños que presentan tres puntos en una de las caras y el otro en la cara opuesta, justo en el vértice opuesto al vértice central de los que está en la cara de los tres puntos, ver Figura B.5 adjunta.

Hay dos diseños de referencia. Por un lado el (1235) centrado por encima del $ED_{.50}$ y por otro el (4678) centrado por debajo del $ED_{.50}$.

Una vez más la Tabla B.3 muestra los mejores diseños de la clase (c) a utilizar según si están centrados por encima o por debajo del $ED_{.50}$ y según el signo de los parámetros. Los diseños correspondientes a una misma fila tendrán sus puntos situados en los mismos ED .

Clase (d)

La clase (d) está formada por los diseños que presentan tres puntos en una de las caras y el otro punto en la cara opuesta, justo en uno de los vértices opuesto a los vértices esquina de los que están en la cara de los tres puntos, ver Figura B.6

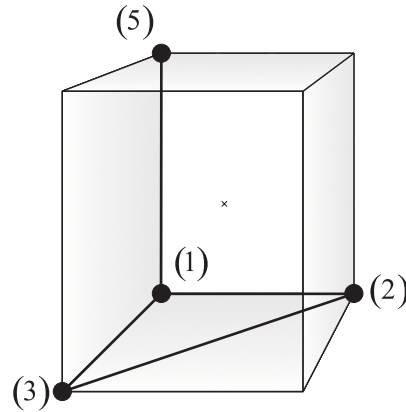


Figura B.5: Diseño de referencia (1235) de la Clase (c)

adjunta. Este grupo tiene la propiedad de que el valor del determinante es el mismo para diseños centrados por encima o por debajo del 50 %.

Hay seis diseños de referencia, el (2346), (2347), (3567), (2456), (2567) y (3457).

La Tabla B.4 muestra los mejores diseños de la clase (d) a utilizar según el signo de los parámetros. Los diseños correspondientes a una misma fila tendrán sus puntos situados en los mismos ED .

Clase (e)

La última clase estaría formada por aquellos diseños con determinante igual a cero. En este caso los cuatro puntos caen en un mismo plano. Tendremos dos subgrupos: los que caen en las caras del paralelogramo y los que se sitúan en los planos diagonales del paralelepípedo, ver Figura B.7 adjunta.

Hay seis diseños de referencia, el (1234), (1357), (1256), (1278), (1368), y (1458).

La Tablas B.5 y B.6 muestran los diseños de la Clase (e) que localizan sus cuatro puntos sobre caras del paralelepípedo y los diseños de la forma (e) que localizan sus cuatro puntos sobre planos conteniendo diagonales del paralelepípedo respectivamente.

Reglas de cambio general

$\beta_3 > 0$				
	$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
> 50 %	(1235)	(1246)	(1347)	(2348)
< 50 %	(4678)	(3578)	(2568)	(1567)

$\beta_3 < 0$				
	$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
> 50 %	(1567)	(2568)	(3578)	(4678)
< 50 %	(2348)	(1347)	(1246)	(1235)

Tabla B.3: Diseños D-óptimos de Clase (c) a utilizar según si están centrados por encima o por debajo del 50 % y según el signo de los parámetros.

Veamos ahora cómo podemos resumir todas las reglas de cambio haciendo uso únicamente de dos tablas.

Dado un diseño cualquiera formado por cuatro puntos de los ocho de un diseño factorial 2^3 , definimos diseños equivalentes a aquellos que tienen igual determinante además de observar en los mismos ED 's. Los diseños pertenecientes a una clase de equivalencia se podrán construir partiendo de los diseños generados según las reglas de cambio de la Tabla B.7 y luego aplicándoles a ellos las reglas de cambio dadas en la Tabla B.8.

La Tabla B.7 nos permite determinar los diseños de referencia asociados a uno dado, mientras que la B.8 permite calcular los diseños equivalentes a uno de referencia cuando cambiamos el signo de los parámetros.

Así por ejemplo, dado el diseño (1234), estará en una clase de equivalencia formada por: Aplicando las reglas de cambio de la Tabla B.7 por los diseños {(1234), (1357), (1256)}, y a estos los de la Tabla B.8 por {(1234),(5678)} {(1357)(2468)} {(1256),(3478)}. En este caso son todos los diseños que tienen determinante igual a cero y que sitúan los cuatro puntos sobre una cara del paralelogramo.

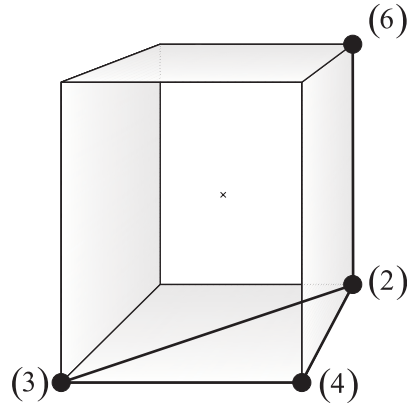


Figura B.6: Diseño de referencia (2346) de la Clase (d)

$\beta_3 > 0$			
$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
(2346)	(1345)	(1248)	(1237)
(2347)	(1348)	(1245)	(1236)
(3567)	(4568)	(1578)	(2678)
(2456)	(1356)	(2478)	(1378)
(2567)	(1568)	(4578)	(3678)
(3457)	(3468)	(1257)	(1268)

$\beta_3 < 0$			
$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
(2678)	(1578)	(4568)	(3567)
(3678)	(4578)	(1568)	(2567)
(1237)	(1248)	(1345)	(2346)
(1268)	(1257)	(3468)	(3457)
(1236)	(1245)	(1348)	(2347)
(1378)	(2478)	(1356)	(2456)

Tabla B.4: Muestra los mejores diseños de la Clase (d) a utilizar según el signo de los parámetros.

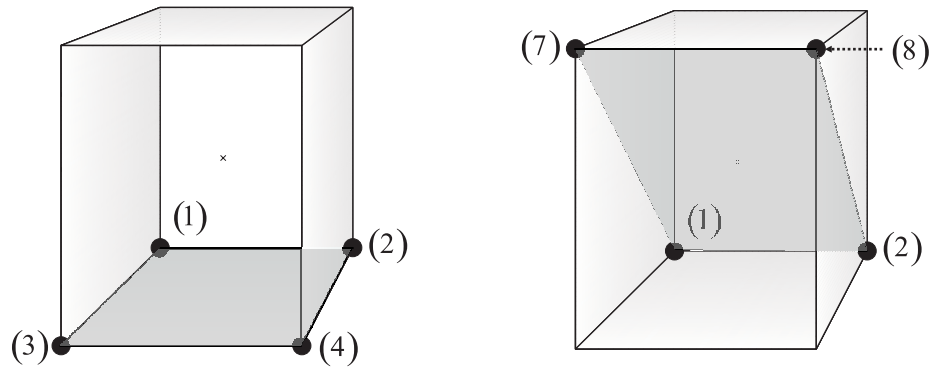


Figura B.7: Diseños de referencia (1234) y (1278) de la Clase (e).

$\beta_3 > 0$			
$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
(1234)	(1234)	(1234)	(1234)
(1357)	(2468)	(1357)	(2468)
(1256)	(1256)	(3478)	(3478)

$\beta_3 < 0$			
$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
(5678)	(5678)	(5678)	(5678)
(1357)	(2468)	(1357)	(2468)
(1256)	(1256)	(3478)	(3478)

Tabla B.5: Muestra los diseños de la Clase e) que localizan sus cuatro puntos sobre caras del paralelepípedo.

$\beta_3 > 0$			
$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
(1278)	(1278)	(3456)	(3456)
(1368)	(2457)	(1368)	(2457)
(1458)	(2367)	(2367)	(1458)

$\beta_3 < 0$			
$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
(3456)	(3456)	(1278)	(1278)
(2457)	(1368)	(2457)	(1368)
(1458)	(2367)	(2367)	(1458)

Tabla B.6: Muestra los diseños de la Clase (e) que localizan sus cuatro puntos sobre planos conteniendo diagonales del paralelepípedo.

Columnas que se intercambian				
2 y 3	2 y 4	3 y 4	3 por 2, 4 por 3 y 2 por 4	4 por 2, 2 por 3 y 3 por 4
			2 \rightarrow 5	2 \rightarrow 3
			3 \rightarrow 2	3 \rightarrow 5
2 \leftrightarrow 3	2 \leftrightarrow 5	3 \leftrightarrow 5	4 \rightarrow 6	4 \rightarrow 7
6 \leftrightarrow 7	4 \leftrightarrow 7	4 \leftrightarrow 6	5 \rightarrow 3	5 \rightarrow 2
			6 \rightarrow 7	6 \rightarrow 4
			7 \rightarrow 4	7 \rightarrow 6

Tabla B.7: Muestra las reglas de cambio de los puntos de un diseño dado al intercambiar dos de las cuatro columnas. Los puntos que no aparecen permanecen inalterados.

$\beta_3 > 0$			
$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
Diseño de referencia	1 ↔ 2	1 ↔ 3	1 ↔ 4
	3 ↔ 4	2 ↔ 4	2 ↔ 3
	5 ↔ 6	5 ↔ 7	5 ↔ 8
	7 ↔ 8	6 ↔ 8	6 ↔ 7

$\beta_3 < 0$			
$\beta_1 > 0, \beta_2 > 0$	$\beta_1 < 0, \beta_2 > 0$	$\beta_1 > 0, \beta_2 < 0$	$\beta_1 < 0, \beta_2 < 0$
1 ↔ 5	1 ↔ 6	1 ↔ 7	1 ↔ 8
2 ↔ 6	2 ↔ 5	2 ↔ 8	2 ↔ 7
3 ↔ 7	3 ↔ 8	3 ↔ 5	3 ↔ 6
4 ↔ 8	4 ↔ 7	4 ↔ 6	4 ↔ 5

Tabla B.8: Muestra las reglas de cambio de los puntos de un diseño dado según los signos de los parámetros. La \leftrightarrow significa que es recíproca la relación.

Apéndice C

Demostración del Corolario 7.2

1. Sea $|z_i| = R|((-1)^i\beta_1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor}\beta_2)|$, entonces, sacando factor común $(-1)^i$ y multiplicando y dividiendo por $(|\beta_1| + |\beta_2|)$ obtenemos que

$$\begin{aligned} |z_i| &= \frac{R(|\beta_1| + |\beta_2|)}{(|\beta_1| + |\beta_2|)} |(-1)^i| \left| \left(\beta_1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor + i} \beta_2 \right) \right| \\ &= \frac{R'}{(|\beta_1| + |\beta_2|)} |\beta_1| \left| \left(1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor + i} \frac{\beta_2}{\beta_1} \right) \right| = \frac{R'}{(1 + |m|)} \left| \left(1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor + i} \frac{\beta_2}{\beta_1} \right) \right| \end{aligned}$$

Sin pérdida de generalidad se verifica la siguiente igualdad

$$\left| \left(1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor + i} \frac{\beta_2}{\beta_1} \right) \right| = \left| \left(1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor + i} \left| \frac{\beta_2}{\beta_1} \right| \right) \right|$$

ya que, por hipótesis $m \in [0, 1]$ y, además, los índices que definen a los distintos puntos del diseño factorial en estudio pueden alternarse de manera que se verifique dicha expresión. Como consecuencia, se obtiene lo que queríamos ver, esto es

$$|z_i| = \frac{R'}{(1 + |m|)} \left| \left(1 + (-1)^{\lfloor \frac{i-1}{2} \rfloor + i} |m| \right) \right| \quad (\text{C.1})$$

2. Para demostrar el segundo apartado, debemos tener en cuenta que para el link logit las funciones $h(\cdot)$ son pares respecto de z_i , esto es

$$h(z_i) = \frac{e^{z_i}}{(1 + e^{z_i})^2} = \frac{e^{-z_i}}{(1 + e^{-z_i})^2}$$

y por lo tanto,

$$h(z_i) = \frac{e^{|z_i|}}{(1 + e^{|z_i|})^2} \quad (\text{C.2})$$

Sustituyendo (C.1) en (C.2), concluimos que las funciones $h(\cdot)$ sólo dependen de R' y m . Así, la expresión del determinante de la matriz de información queda de la forma:

$$\text{Det}(I(\beta_1, \beta_2 = m\beta_1)) = 4^2 n^3 (R')^4 \frac{1}{\beta_1^4 (1 + |m|)^4} \sum_{i < j < r} \lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r)$$

y por lo tanto, el valor de R' que lo maximiza, sólo depende de m .

3. Su demostración es trivial.

4. Tenemos que

$$\text{Det}(I(\beta_1, \beta_2 = m\beta_1)) = 4^2 n^3 (R_{(\beta_1, \beta_2 = m\beta_1)})^4 \sum_{i < j < r} \lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r)$$

Por el apartado (3.) tenemos que $R_{(\beta_1, \beta_2 = m\beta_1)} = \frac{R_{(1, m)}}{|\beta_1|}$, así

$$\begin{aligned} \text{Det}(I(\beta_1, \beta_2 = m\beta_1)) &= 4^2 n^3 \left(\frac{R_{(1, m)}}{|\beta_1|} \right)^4 \sum_{i < j < r} \lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r) \\ &= \frac{1}{(\beta_1)^4} \left(4^2 n^3 (R_{(1, m)})^4 \sum_{i < j < r} \lambda_i \lambda_j \lambda_r h(z_i) h(z_j) h(z_r) \right) = \frac{\text{Det}(I(1, m))}{(\beta_1)^4} \end{aligned}$$

como queríamos ver.

Bibliografía

- Abdelbasit, K.M., y Plackett, R.L. (1983), “Experimental Design for Binary Data”, *Journal of the American Statistical Association*, 78, 90-98.
- Aitkin, M., Anderson, D., Francis, B. y Hinde, J. (1989), *Statistical Modelling in GLIM*, Oxford: Oxford Science Publications.
- Atkinson, A.C., Chaloner, K., Herzberg, A.M. y Juritz, J. (1993), “Optimum Experimental Designs for Properties of a Compartmental Model”, *Biometrics*, 49, 325-337.
- Atkinson, A.C., y Donev, A.N. (1992), *Optimum Experimental Designs*, Oxford: Clarendon Press.
- Berkson, J. (1944), “Application of the logistic function to bio-assay”, *Journal of the American Statistical Association*, 39, 357-365.
- Berkson, J. (1951), “Why I prefer logits to probits”, *Biometrics*, 7, 327-339.
- Bisgaard, S., y Fuller, H., T. (1994), “Analysis of factorial experiments with defects or defectives as the response”, *Quality Engineering*, 7, 429-443.
- Bisgaard, S., y Fuller, H., T. (1995), “Sample Size Estimates for 2^{k-p} Designs with Binary Responses”, *Journal of Quality Technology*, 27, 344-354.
- Bliss, C.I. (1935), “The calculation of the dosage-mortality curve”, *Annals of Applied Biology*, 22, 134-167.

-
- Box, G. E. P. y Behnken, D. W. (1960), “Some new level designs for the study of quantitative variables”, *Technometrics*, 2, 455-476.
 - Box, M.J., y Draper, N. (1971), “Factorial Designs, the $|X'X|$ Criterion and some Related Matters”, *Technometrics*, 13, 731-742.
 - Box, G.E.P., y Draper, N. (1987), *Empirical Model-Building and Response Surfaces*, New York: Wiley.
 - Box, G.E.P., y Hunter, J.S. (1961a), “The 2^{k-p} fractional factorial designs, Part I”, *Technometrics*, 2, 311-352.
 - Box, G.E.P., y Hunter, J.S. (1961b), “The 2^{k-p} fractional factorial designs, Part II”, *Technometrics*, 2, 449-458.
 - Box, G.E.P., Hunter, W.G., y Hunter, J.S. (1978), *Statistics for Experimenters*, New York: Wiley.
 - Box, G.E.P. y Lucas, H.L. (1959), “Design of Experiments in Nonlinear Situations”, *Biometrics*, 46, 77-90.
 - Chaloner, K. y Verdinelli, I. (1995), “Bayesian Experimental Design: A Review”, *Statistical Science*, 10, 273-304.
 - Chaloner, K., y Larntz, K. (1989), “Optimal Bayesian Design Applied to Logistic Regression Experiments”, *Journal of Statistical Planning and Inference*, 21, 198-208.
 - Chernoff H. (1953). “Locally optimal designs for estimating parameters”, *Annals Mathematics Statistics*, 24, 586-602.
 - Chiachierini, L.M. (1996), *Experimental Design Issues in Impaired Reproduction Studies*, PhD Thesis, Virginia Tech.
 - Collet, D. (2003), *Modelling Binary Data 2nd edn.*, Florida: Chapman & Hall/CRC.
-

-
- Dobson, A.J. (1990), *An introduction to generalized linear models*, London: Chapman Hall.
 - Dorta-Guerra, R. y Gonzalez-Davila, E. (2004), “D-optimal 2^2 factorial designs for binary response data”, *Proceedings of COMPSTAT 2004 (ed. Physica Verlag)*, 951-958.
 - Dumouchel, W. y Jones, B. (1994), “A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model”, *Technometrics*, 36, 37-47.
 - Dyke, G.V. y Patterson, H.D. (1952), “Analysis of factorial arrangements when the data are proportions”, *Biometrics*, 8, 1-12.
 - Fedorov, V.V. (1972), *Theory of Optimal Experiments*, New York: Academic Press.
 - Finney (1971), *Probit Analysis*, Cambridge: Cambridge University Press.
 - Ford, I., Torsney, B., y Wu, C.F.J. (1992), “The Use of Canonical form in the Construction of Locally Optimal Designs for Non-linear Problems”, *Journal of the Royal Statistical Society, Ser. B*, 54, 569-583.
 - Francis, B., Green, M. y Payne, C. (1993), *The GLIM System. Release 4 Manual*, Oxford: Clarendon Press.
 - Goel, P.K., y Ginebra, J. (2003), “When is One Experiment Always Better than Another?”, *Journal of the Royal Statistical Society, Ser. D*, 52, 515-537.
 - Hamada, M., y Nelder, J.A. (1997), “Generalized Linear Models for Quality-improvement Experiments”, *Journal of Quality Technology*, 29, 292-304.
 - Harville, D.A. (1997), *Matrix Algebra from a Statistician’s Perspective*, New York: Springer.
-

-
- Heise, M.A. y Myers, R.H. (1996), “Optimal designs for bivariate logistic regression”, *Biometrika*, 52, 613-624.

 - Jia, Y., y Myers, R.H. (2001), “Some Optimal and Robust Designs for the Two Variable Logistic Model”, Tech. Report 01-3, Department of Statistics, VPI&SU, Blacksburg, Virginia.

 - Kalish, L.R. (1990), “Efficient Design for estimation of median lethal dose and quantal dose-response curves”, *Biometrics*, 46, 737-748.

 - Kalish, L.R. y Rosenberger, J.L. (1978), “Optimal Designs for the estimation of the logistic function”, Tech. Report 33, Pennsylvania State University.

 - Khan, M.K., y Yazdi, A.A. (1988), “On D-optimal Designs for Binary Data”, *Journal of Statistical Planning and Inference*, 18, 83-91.

 - Kiefer, J. (1959), “Optimum Experimental Designs”, *Journal of the Royal Statistical Society*, Ser. B, 21, 272-319.

 - Kiefer, J. y Wolfowitz, J. (1959), “Optimum designs in regression problems”, *Annals of Mathematical Statistics*, 30, 271-294.

 - Kleinbaum, D.G. (1994), *Logistic Regression: A self-learning text*, New York: Springer-Verlag.

 - Lee, Y. y Nelder, J.A. (1998b), “Generalized linear models for the analysis of quality-improvement experiments”, *Canadian Journal of Statistics*, 26, 95-105.

 - Lehmann, E.L. (1983), *Theory of point Estimation*, New York: John Wiley & Sons.

 - Lestingier, W.C. (1995), *Optimal One and Two-Stage Designs for the Logistic Regression Model*, PhD Thesis, Virginia Tech.
-

-
- Lewis, S.L., Montgomery, D.C. y Myers, R.H. (2001a), “Examples of designed experiments with nonnormal responses”, *Journal of Quality technology*, 33, 265-278.
 - Lin, H., Myers, R.H. y Ye, K. (2000), “Bayesian Two-Stage Optimal Design for Mixture Models”, *Journal of Statistical Computation and Simulation*, 66, 209-231.
 - Mathew, T., y Sinha, B.K. (2001), “Optimal Designs for Binary Data Under Logistic Regression”, *Journal of Statistical Planning and Inference*, 93, 295-307.
 - Maxim, L.D., Hendrickson, A.D. y Cullen, D.E. (1977), “Experimental Design for sensitivity testing: The Weibull model”, *Technometrics*, 19, 405-413.
 - McCullagh, P., y Nelder, J.A. (1989), *Generalized Linear Models*, 2nd ed., London: Chapman Hall.
 - McLeish, D.L. y Tosh, D. (1990), “Sequential Designs in Bioassay”, *Biometrics*, 46, 103-116.
 - Meeker, W.Q. y Hahn, G.J. (1977), “Asymptotically optimum over-stress test to estimate the survival probability at a condition with a low expected failure probability”, *Technometrics*, 19, 381-404.
 - Minkin, S. (1987), “Optimal Designs for Binary Data”, *Journal of the American Statistical Association*, 82, 1098-1103.
 - Minkin, S. (1993), “Experimental design for clonogenic assays in chemotherapy”, *Journal of the American Statistical Association*, 88, 410-420.
 - Minkin, S. y Kundhal, K (1999), “Likelihood-based experimental designs for estimation of ED_{50} ”, *Biometrics*, 55, 1030-1037.
 - Müller, W.G. y Ponce de Leon, A.C. (1996), “Discrimination between Two Binary Data Models: Sequentially Designed Experiments”, *Journal of Statistical Planning and Inference*, 55, 87-100.
-

-
- Myers, W.R., Myers, R.H., y Carter, W.H. (1994), “Some Alphabetic Optimal Designs for the Logistic Regression Model”, *Journal of Statistical Planning and Inference*, 42, 57-77.
 - Myers, W.R., Myers, R.H., Carter, W.H. Jr. y White, K.L. (1996), “Two-stage Designs for the Logistic Regression Model in a Single Agent Bioassay”, *Journal of Biopharmaceutical Statistics*, 6, 283-301.
 - Myers, R. H., y Montgomery, D. C. (1997), “A Tutorial on Generalized Linear Models”, *Journal of Quality Technology*, 29, 274-291.
 - Myers, R. H., y Montgomery, D. C. (2002), *Response Surface Methodology, Process and Product Optimization Using Designed Experiments (2nd ed.)*, New York: Wiley.
 - Myers, R.H., Montgomery, D.C., y Vinning, G.G. (2002), *Generalized Linear Models*, New York: Wiley.
 - Nelder, J.A. y Wedderburn, R.W.M. (1972), “Generalized Linear Models”, *Journal of the Royal Statistical Society A*, 132, 107-120.
 - Plackett, R.L y Burman, J.P. (1946), “The design of optimal multifactorial experiments”, *Biometrika*, 33, 305-325.
 - Pukelsheim, F. (1993), *Optimal Design of Experiments*, New York: Wiley.
 - Schwartz, P.F., Gennings, C., Teuschler, L.K. y Fariss, M.W. (2001), “Optimizing the Precision of Toxicity Threshold Estimation Using a Two-Stage Experimental Design”, *Journal of Agricultural, Biological, and Environmental Statistics*, 6, 409-428.
 - Sebastiani, P. y Settimi, R. (1997), “A Note on D-optimal Designs for a Logistic Regression Model”, *Journal of Statistical Planning and Inference*, 59, 359-368.
 - Silvey, D. (1980), *Optimal Design*, London: Chapman and Hall.
-

-
- Sitter, R.R. (1992), “Robust Designs for Binary Data”, *Biometrics*, 48, 1145-1156.
 - Sitter, R.R., y Fainaru, I. (1997), “Optimal Design for the Logit and Probit Models for Binary Data”, *Canadian Journal of Statistics*, 25, 175-190.
 - Sitter, R.R., y Forbes, B.E. (1997), “Optimal Two-stage Designs for Binary Response Experiments”, *Statistica Sinica*, 7, 941-956.
 - Sitter, R.R., y Torsney, B. (1995a), “Optimal Designs for Binary Response Experiments with Two Design Variables”, *Statistica Sinica*, 5, 405-419.
 - Sitter, R.R., y Torsney, B. (1995b), “D-Optimal Designs for Generalized Linear Models”, in *Proceedings of MODA 4*, eds. C.P. Kitsos and W.G. Muller, W.G., Heidelberg: Physica Verlag, Heidelberg, pp. 87-102.
 - Sitter, R.R., y Wu, C.F.J. (1993), “Optimal Designs for Binary Response Experiments: Fieller, D, and A Criteria”, *Scandinavian Journal of Statistics*, 20, 329-342.
 - Sitter, R.R., y Wu, C.F.J. (1999), “Two-stage Design of Quantal Response Studies”, *Biometrics*, 55, 396-402.
 - Torsney, B. y Gunduz, N. (2000), “On optimal Designs for High Dimensional Binary Regression Models”, *Technical Report, Department of Statistics, University of Glasgow*, vol. 00-8.
 - Torsney, B. y Musrati, A. (1993), “On the construction of optimal designs with applications to binary response and to weighted regression models”, in *Proceedings of MODA 4*, eds. W.G. Muller, H.P. Wynn y A. Zhigljavsky, Heidelberg: Physica Verlag, pp. 37-52.
 - Van Mullekom J. y Myers R.H. (2001), “Optimal experimental designs for Poisson impaired reproduction”, *Technical Report, Department of Statistics, Virginia Tech.*, vol. 01-1.
-

- White, L. V. (1975), *The optimal design of experiments for estimation in non-linear models*, PhD Thesis, University of London.
 - Williams, D.A. (1986), “Interval estimation of the median lethal dose”, *Biometrics*, 42, 641-645.
 - Wu, C.F.J. (1985), “Efficient Sequential Designs with Binary Data”, *Journal of the American Statistical Association*, 80, 974-984.
 - Wu, C.F.J. (1988), “Optimal Design for Percentile Estimation of a Quantal Response Curve”, in *Optimal Design and Analysis of Experiments*, eds. Y. Dodge, V. Fedorov, and H.P. Wynn, Elsevier: Amsterdam, pp. 213-223.
 - Wu, C.F.J., y Hamada, M. (2000), *Experiments. Planning, Analysis, and Parameter Design Optimization*, New York: Wiley.
 - Wynn, H. (1972a), *Results in the theory and construction of D-optimal experimental designs*, New York: Wiley.
 - Wynn, H. (1972b), “Results in the theory and construction of D-optimal experimental designs”, *Journal of the Royal Statistical Society B*, 34, 133-147.
-