

DISEÑOS ÓPTIMOS EN LA PRESENCIA DE EFECTOS DE BLOQUES ALEATORIOS

Presentado por:

EDDIE EDINSON RODRIGUEZ BOSSIO

Código: 72231783

Director:

Dr. rer. nat. JESÚS ALONSO CABRERA

Universidad del Atlántico

Facultad de Ciencias Básicas

Programa de Maestría en Ciencias Matemáticas

Barranquilla (Atlántico)

Septiembre de 2015

UNIVERSIDAD DEL ATLÁNTICO
PROGRAMA DE MATEMÁTICAS
MAESTRÍA EN CIENCIAS MATEMÁTICAS

DISEÑOS ÓPTIMOS EN LA PRESENCIA DE EFECTOS DE
BLOQUES ALEATORIOS

Trabajo Especial de Grado presentado a la Universidad del Atlántico por
EDDIE EDINSON RODRIGUEZ BOSSIO

como requisito parcial para optar al grado de
Magister en Ciencias Matemáticas

Con la dirección del profesor
Dr. rer. nat. Jesús Alonso Cabrera

Septiembre de 2015

UNIVERSIDAD DEL ATLÁNTICO
PROGRAMA DE MATEMÁTICAS
MAESTRÍA EN CIENCIAS MATEMÁTICAS

**DISEÑOS ÓPTIMOS EN LA PRESENCIA DE EFECTOS DE BLOQUES
ALEATORIOS**

Estudiante: EDDIE RODRIGUEZ BOSSIO
Código: 72231783

Este Trabajo Especial de Grado ha sido aprobado en nombre de la Universidad del Atlántico por el siguiente jurado examinador:

Dr. Julio Cesar Romero Pabón
Universidad del Atlántico - Colombia

M.Sc. Alejandro Sanchez Salazar
Universidad del Atlántico - Colombia

Dr. rer. nat. Jesús Alonso Cabrera
Tutor

Septiembre de 2015

Índice general

Aprobación del Jurado	2
Introducción	I
1. Preliminares	1
1.1. La ecuación de regresión	1
1.2. Estimación de mínimos cuadrados	2
1.3. Propiedades de los estimadores de mínimos cuadrados	12
1.4. Teorema de Gauss - Markov	14
1.5. Diseños experimentales y matrices de información	15
1.6. Criterios de optimalidad	18
1.6.1. $D - \text{criterio}$	19
1.6.2. $G - \text{criterio}$	19
1.6.3. $D_s - \text{criterio}$	19
1.6.4. Teorema de equivalencia	21
2. Diseños óptimos en la presencia de efectos de bloques aleatorios	24
2.1. Aspectos del diseño	28
3. Conclusiones y trabajos futuros	32
Bibliografía	32

Introducción

Al interior de los experimentos estadísticos la teoría de los diseños óptimos ha sido desarrollada. En general el tema de esta teoría es que para un apropiado modelo, si queremos poner énfasis sobre una cualidad particular de los parámetros a estimar, entonces la configuración experimental debería ser elegida de acuerdo a ciertos criterios con sentido estadístico. En la literatura relacionada con los diseños óptimos, un prominente autor fue Kiefer (1959), el cuál presentó los principales conceptos, tales como diseños aproximados y una variedad de criterios de optimalidad para esta rama de los diseños de experimentos; Kiefer, en particular dio el nombre D —optimalidad al criterio introducido por Wald (1943), este criterio es el más comunmente aplicado y está definido en función del determinante de la matriz de covarianza.

Más recientemente son reconocidos los libros de Atkinson y Donev (1992) y Pukelsheim (1993), donde los autores hacen una presentación estadística formal de los diseños óptimos. El presente trabajo se ha organizado en tres capítulos: el capítulo uno (Preliminares) contiene conceptos generales que sirven de apoyo y base a la teoría que se desarrolla en los siguientes dos capítulos. El capítulo dos trata sobre diseños óptimos en la presencia de efectos de bloques aleatorios, y en el capítulo tres se darán a conocer las conclusiones y una serie de problemas abiertos para futuras investigaciones relacionadas con el tema central de este trabajo de investigación.

Agradecimientos

En primer lugar doy gracias a Dios dador de la vida y oportunidades; quien puso en mi camino a personas e instituciones que fueron piezas claves para este logro en mi vida. Ellos son mi familia, mi tutor de tesis Dr. rer. nat. Jesús Alonso Cabrera por compartir sus conocimientos y consejos, Dr. Jorge Rodriguez, Dr. Alejandro Urieles, profesores de la maestría, compañeros de estudio y las instituciones de la Universidad del Atlántico donde cursé mi pregrado y maestría, y la Universidad del Norte donde realicé mi especialización en matemáticas y se llevaron a cabo las asesorías de este trabajo de grado, ya que mi asesor es docente de planta de esta alma mater.

Preliminares

En el área de diseños de experimentos existen, principalmente, dos enfoques competitivos: uno basado en el análisis combinatorio más ajustado para modelos estadísticos de análisis de varianza y el otro enfoque basado en métodos analíticos que envuelve el análisis convexo, el cual se aplica, por ejemplo, a superficies de respuesta. Una concisa introducción a este último enfoque es dado en la breve monografía por Silvey (1980).

Mientras para modelos en la presencia de efectos fijos se han encontrado diseños óptimos para una gran variedad de casos (Schwabe, 2003) si los efectos de bloques se asumen que provienen de procesos aleatorios, aparecen dificultades adicionales. Un primer simple ejemplo de polinomios de regresión en la presencia de efectos de factores aleatorios discretos ha sido considerado en los pioneros artículos de Cheng (1995) y Atkins y Cheng (1999). Goos (2000) extendió este resultado a varias estructuras de bloque y obtuvo soluciones numéricas. A continuación presentaremos los conceptos básicos de la teoría de diseños de experimentos.

1.1. La ecuación de regresión

La siguiente ecuación es básica en la teoría de regresión:

$$Y_j = \eta(x_j, \boldsymbol{\theta}) + \epsilon_j, \quad j = 1, \dots, N, \quad (1.1)$$

donde Y_1, \dots, Y_N son resultados experimentales, $\eta(x, \boldsymbol{\theta})$ es una función dada con vector de parámetros desconocidos $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^\top$, $\epsilon_1, \dots, \epsilon_N$ corresponden al error de observación y x_1, \dots, x_N son condiciones experimentales, que pertenecen a un conjunto

compacto \mathcal{X} usualmente llamado la región de diseño.

Los casos para representar los resultados de los experimentos reales en la forma (1.1) se ha demostrado en muchos ejemplos, ver Rao (1973), Federow (1972), y Pukelsheim (1993).

Recordemos algunos supuestos básicos del modelo clásico de regresión.

- (a) Insensgamiento: $E(\epsilon_j) = 0$; ($j = 1, \dots, N$). Esto significa que $E(Y_j) = \eta(x_j, \boldsymbol{\theta})$ (es decir, el modelo está libre de un error sistemático).
- (b) Incorrelación: $E(\epsilon_i \epsilon_j) = 0$; ($i \neq j$).
- (c) Homogeneidad de varianza: $E(\epsilon_j^2) \equiv \sigma^2 > 0$; ($j = 1, \dots, N$).
- (d) La linealidad de parametrización: $\eta(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\theta}$, donde $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))^\top$, $f_i(\mathbf{x})$, $i = 1, \dots, m$, son funciones básicas y conocidas.

Como es habitual en la teoría estadística, estos supuestos proporcionan resultados observables a obtener y corresponden en cierta medida a las características de experimentos reales.

El propósito principal de un experimento es estimar un vector de parámetros desconocidos, o probar una hipótesis sobre los valores de los parámetros. Aquí, la exactitud de conclusiones estadísticas depende tanto del método de la inferencia estadística y en la elección de las condiciones experimentales.

Si (a) - (b) se asumen, entonces la técnica de mínimos cuadrados, proporciona valoración del vector $\boldsymbol{\theta}$ de parámetros bajo cualesquiera condiciones experimentales fijas.

1.2. Estimación de mínimos cuadrados

La paternidad de este método se reparte entre Legendre que lo publicó en 1805 y Gauss que lo utilizó en 1795 y lo publicó en 1809.

El método de mínimos cuadrados (MC) es utilizado para estimar los parámetros en el modelo de regresión lineal.

Por ejemplo, en el modelo de regresión lineal múltiple

$$\begin{aligned} Y_j &= \beta_0 + \beta_1 x_{j1} + \cdots + \beta_k x_{jk} + \epsilon_j \\ &= \beta_0 + \sum_{i=1}^k \beta_i x_{ji} + \epsilon_j, j = 1, \dots, N. \end{aligned}$$

Suponga que se tiene $N > k$ observaciones. Se asume que $E(\epsilon_j) = 0$ y $Var(\epsilon_j) = \sigma^2$ y que los errores son independientes. El método de mínimos cuadrados minimiza la suma de cuadrados del error dada por

$$\begin{aligned} SSE &= \sum_{j=1}^N \epsilon_j^2 \\ &= \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right)^2 \end{aligned}$$

con respecto a cada uno de los parámetros del modelo $\beta_0, \beta_1, \dots, \beta_k$.

La derivada con respecto a β_0

$$\begin{aligned} \frac{\partial SSE}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0} \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right)^2 \\ &= -2 \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) \end{aligned}$$

La derivada con respecto a β_j ($j = 1, 2, \dots, k$) es

$$\begin{aligned} \frac{\partial SSE}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right)^2 \\ &= -2 \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} \end{aligned}$$

igualando a cero las derivadas, se tiene

$$\frac{\partial SSE}{\partial \beta_0} = -2 \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) = 0$$

$$\frac{\partial SSE}{\partial \beta_j} = -2 \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} = 0$$

Simplificando para β_0 se tiene

$$\sum_{j=1}^N Y_j - \sum_{j=1}^N \beta_0 - \sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} = 0$$

$$\sum_{j=1}^N \beta_0 + \sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} = \sum_{j=1}^N Y_j$$

$$N\beta_0 + \sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} = \sum_{j=1}^N Y_j$$

$$N\beta_0 + \sum_{j=1}^N (\beta_1 x_{j1} + \beta_2 x_{j2} + \cdots + \beta_k x_{jk}) = \sum_{j=1}^N Y_j$$

$$N\beta_0 + \beta_1 \sum_{j=1}^N x_{j1} + \beta_2 \sum_{j=1}^N x_{j2} + \cdots + \beta_k \sum_{j=1}^N x_{jk} = \sum_{j=1}^N Y_j$$

Simplificando para β_j se tiene

$$\sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} = 0$$

$$\left(\sum_{j=1}^N Y_j - \sum_{j=1}^N \beta_0 - \sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} = 0$$

$$\sum_{j=1}^N Y_j x_{ji} - \beta_0 \sum_{j=1}^N x_{ji} - \left(\sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} = 0$$

$$\sum_{j=1}^N Y_j x_{ji} - \beta_0 \sum_{j=1}^N x_{ji} - \sum_{j=1}^N (\beta_1 x_{j1} + \cdots + \beta_k x_{jk}) x_{ji} = 0$$

$$\sum_{j=1}^N Y_j x_{ji} - \beta_0 \sum_{j=1}^N x_{ji} - \left(\sum_{j=1}^N \beta_1 x_{j1} + \cdots + \beta_k \sum_{j=1}^N x_{jk} \right) x_{ji} = 0$$

$$\beta_0 \sum_{j=1}^N x_{ji} + \left(\sum_{j=1}^N \beta_1 x_{j1} + \cdots + \beta_k \sum_{j=1}^N x_{jk} \right) x_{ji} = \sum_{j=1}^N Y_j x_{ji}$$

Luego las ecuaciones normales son:

$$\begin{aligned}
 N\beta_0 + \beta_1 \sum_{j=1}^N x_{j1} + \cdots + \beta_k \sum_{j=1}^N x_{jk} &= \sum_{j=1}^N Y_j \\
 \beta_0 \sum_{j=1}^N x_{j1} + \beta_1 \sum_{j=1}^N x_{j1}^2 + \cdots + \beta_k \sum_{j=1}^N x_{jk} x_{j1} &= \sum_{j=1}^N Y_j x_{j1} \\
 \beta_0 \sum_{j=1}^N x_{j2} + \beta_1 \sum_{j=1}^N x_{j1} x_{j2} + \cdots + \beta_k \sum_{j=1}^N x_{jk} x_{j2} &= \sum_{j=1}^N Y_j x_{j2} \\
 &\vdots \\
 \beta_0 \sum_{j=1}^N x_{jk} + \beta_1 \sum_{j=1}^N x_{j1} x_{jk} + \cdots + \beta_k \sum_{j=1}^N x_{jk}^2 &= \sum_{j=1}^N Y_j x_{jk}
 \end{aligned}$$

Observe que hay $p = k + 1$ ecuaciones. Para obtener la solución es conveniente utilizar notación matricial. En esta notación el modelo se expresa como

$$Y = X\beta + \epsilon$$

donde

Y es el vector de observaciones

X es una matriz $n \times p$ de niveles de la variable

β es un vector $p \times 1$ de coeficientes de regresión

ϵ es el vector aleatorio error de orden $p \times 1$

La suma de cuadrados del error es dada por

$$SSE = \sum_{j=1}^N \epsilon_j^2 = \epsilon' \epsilon = (Y - X\beta)'(Y - X\beta)$$

Luego se obtiene que las ecuaciones normales son

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$$

Para solucionar las ecuaciones normales se requiere que exista la inversa de la matriz $\mathbf{X}'\mathbf{X}$. Esta existe siempre que las variables regresoras sean linealmente independientes. Así, la solución de mínimos cuadrados de vector paramétrico $\boldsymbol{\beta}$ es

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

Ejemplo 1

Myers y Montgomery (1985) describen un experimento que trata con motores de turbina de gas. El voltaje de salida de los motores se midió en diversas combinaciones de velocidad de la cuchilla y extensión del sensor de medición de tensión. Los datos del experimento se dan en el Cuadro 1.1.

Corrida	Voltaje	Velocidad de la cuchilla (pulg/seg)	Extension (pulg)
1	1.23	5300	0.000
2	3.13	8300	0.000
3	1.22	5300	0.012
4	1.92	8300	0.012
5	2.02	6800	0.000
6	1.51	6800	0.012
7	1.32	5300	0.006
8	2.62	8300	0.006
9	1.65	6800	0.006
10	1.62	6800	0.006
11	1.59	6800	0.006

Cuadro 1.1: Experimento turbina de gas descrito por Myers y Montgomery (1995).

El propósito del experimento fue estimar un modelo cuadrático completo con el voltaje de salida como la variable dependiente y la velocidad de la cuchilla y extensión del sensor de medición de tensión como las variables explicativas.

Variables codificadas

Es conveniente para la mayoría de aplicaciones describir el experimento en términos de variables codificadas, porque esto facilita la comparación de los diseños de diferentes experimentos. Por lo tanto, se reajustarán las variables cuantitativas. Es característico de una variable u cuantitativa o continua que varía entre un valor mínimo y máximo, u_{min} y u_{max} . Por lo general, los niveles de los factores se reajustarán a estar entre -1 y $+1$. Los valores codificados pueden ser calculados por

$$z = \frac{u - u_0}{\Delta}$$

donde u_0 es el punto medio del intervalo $[u_{min}, u_{max}]$ y Δ es la mitad de la diferencia entre u_{max} y u_{min} . Para la interpretación de los resultados experimentales, sin embargo, es deseable para volver a los niveles de los factores originales. Para el experimento de turbina de gas, los niveles codificados se pueden obtener de la siguiente manera:

$$x_1 = \frac{Velocidad\ del\ cuchilla - 6800}{1500}$$

y

$$x_2 = \frac{Extension - 0,006}{0,006}$$

donde x_1 y x_2 representan los niveles codificados de los factores de velocidad de la cuchilla y extensión del sensor de medición de tensión, respectivamente. Los niveles codificados se muestran en el Cuadro 1.2. Vamos a utilizar esta forma de analizar los datos. Del Cuadro 1.2, es fácil ver que las corridas 9, 10 y 11 se llevan a cabo en el nivel medio de los factores experimentales.

Análisis

Corrida	Voltaje	Velocidad de la cuchilla (pulg/seg)	Extension (pulg)
1	1.23	-1	-1
2	3.13	+1	-1
3	1.22	-1	+1
4	1.92	+1	+1
5	2.02	0	-1
6	1.51	0	+1
7	1.32	-1	0
8	2.62	+1	0
9	1.65	0	0
10	1.62	0	0
11	1.59	0	0

Cuadro 1.2: Forma codificada del experimento de turbinas de gas.

El propósito del experimento fue estimar un modelo cuadrático completo en las dos variables. Como resultado, la expansión polinómica

$$\mathbf{f}^\top(\mathbf{x}) = \begin{bmatrix} 1 & x_1 & x_2 & x_1x_2 & x_1^2 & x_2^2 \end{bmatrix}$$

y

$$\boldsymbol{\beta}^\top = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_{12} & \beta_{11} & \beta_{22} \end{bmatrix},$$

de manera que $p = 6$ y el modelo estadístico se pueden escribir como

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \epsilon.$$

La matriz de diseño de todo el experimento viene dada por

$$X = \begin{pmatrix} 1 & -1 & -1 & +1 & +1 & +1 \\ 1 & +1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 & +1 & +1 \\ 1 & 0 & -1 & 0 & 0 & +1 \\ 1 & 0 & +1 & 0 & 0 & +1 \\ 1 & -1 & 0 & 0 & +1 & 0 \\ 1 & +1 & 0 & 0 & +1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

la matriz $X^T X$ es

$$\begin{pmatrix} 11 & 0 & 0 & 0 & 6 & 6 \\ 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 6 & 0 & 0 & 0 & 6 & 4 \\ 6 & 0 & 0 & 0 & 4 & 6 \end{pmatrix}$$

y entonces $(X^T X)^{-1}$ es

$$\begin{pmatrix} 0,2632 & 0 & 0 & 0 & -0,1579 & -0,1579 \\ 0 & 0,1667 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,1667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,25 & 0 & 0 \\ -0,1579 & 0 & 0 & 0 & 0,3947 & -0,1053 \\ -0,1579 & 0 & 0 & 0 & -0,1053 & 0,3947 \end{pmatrix}$$

El vector Y es

$$Y = \begin{pmatrix} 1,23 \\ 3,13 \\ 1,22 \\ 1,92 \\ 2,02 \\ 1,51 \\ 1,32 \\ 2,62 \\ 1,65 \\ 1,62 \\ 1,59 \end{pmatrix}$$

y el vector $X^T Y$ es

$$\begin{pmatrix} 19,83 \\ 3,9 \\ -1,73 \\ -1,2 \\ 11,44 \\ 11,03 \end{pmatrix}$$

el estimador de mínimos cuadrados de β es

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

o

$$\begin{aligned}
\hat{\beta} &= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_{12} \\ \hat{\beta}_{11} \\ \hat{\beta}_{22} \end{pmatrix} \\
&= \begin{pmatrix} 0,2632 & 0 & 0 & 0 & -0,1579 & -0,1579 \\ 0 & 0,1667 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,1667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,25 & 0 & 0 \\ -0,1579 & 0 & 0 & 0 & 0,3947 & -0,1053 \\ -0,1579 & 0 & 0 & 0 & -0,1053 & 0,3947 \end{pmatrix} \begin{pmatrix} 19,83 \\ 3,9 \\ -1,73 \\ -1,2 \\ 11,44 \\ 11,03 \end{pmatrix} \\
&= \begin{pmatrix} 1,6705 \\ 0,6500 \\ -0,2883 \\ -0,3000 \\ 0,2237 \\ 0,0187 \end{pmatrix}
\end{aligned}$$

luego el modelo ajustado por mínimos cuadrados es

$$\hat{Y} = 1,6706 + 0,6500x_1 - 0,2883x_2 - 0,3000x_1x_2 + 0,2237x_1^2 + 0,0187x_2^2$$

1.3. Propiedades de los estimadores de mínimos cuadrados

(1) $\hat{\beta}$ es un estimador insesgado de β . Ésto es, $E(\hat{\beta}) = \beta$.

Demostración:

$$\begin{aligned} E(\hat{\beta}) &= E[(X^T X)^{-1} X^T Y] \\ &= (X^T X)^{-1} X^T E(Y) \end{aligned}$$

y como $E(Y) = X\beta$, entonces

$$\begin{aligned} E(\hat{\beta}) &= (X^T X)^{-1} X^T X\beta \\ &= \beta \end{aligned}$$

- (2) La matriz de varianzas y covarianzas del vector $\hat{\beta}$ es $Cov(\hat{\beta}) = Var(\hat{\beta}) = \sigma^2(X^T X)^{-1}$

Demostración:

$$Var(\hat{\beta}) = Var((X^T X)^{-1} X^T Y)$$

Sea $A = (X^T X)^{-1} X^T$ y como A es una matriz y Y un vector columna y por la propiedad de varianzas $Var(AY) = AVar(Y)A^T$, se tiene que

$$\begin{aligned} Var(\hat{\beta}) &= (X^T X)^{-1} X^T Var(Y) ((X^T X)^{-1} X^T)^T \\ &= (X^T X)^{-1} X^T \sigma^2 ((X^T X)^{-1} X^T)^T \\ &= (X^T X)^{-1} X^T \sigma^2 (X^T)^T ((X^T X)^{-1})^T \\ &= (X^T X)^{-1} X^T \sigma^2 X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} (X^T X) (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} I \\ &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

Ejemplo 2

Para los datos del ejemplo 1, se tiene que la estimación de la matriz de varianzas-covarianzas del vector $\hat{\beta}$ es

$$Var(\hat{\beta}) = \sigma^2(X^T X)^{-1}$$

$$= 0,0083 \begin{pmatrix} 0,2632 & 0 & 0 & 0 & -0,1579 & -0,1579 \\ 0 & 0,1667 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,1667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,25 & 0 & 0 \\ -0,1579 & 0 & 0 & 0 & 0,3947 & -0,1053 \\ -0,1579 & 0 & 0 & 0 & -0,1053 & 0,3947 \end{pmatrix}$$

$$= \begin{pmatrix} 0,00218456 & 0 & 0 & 0 & -0,00131057 & -0,00131057 \\ 0 & 0,00138361 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,00138361 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,002075 & 0 & 0 \\ -0,00131057 & 0 & 0 & 0 & 0,00327601 & -0,00087399 \\ -0,00131057 & 0 & 0 & 0 & -0,00087399 & 0,00327601 \end{pmatrix}$$

Los errores estándar de cada parámetro es dado en la tabla:

Parámetro	Error estándar
β_0	$\sqrt{0,00218456} = 0,04673928$
β_1 , β_2	$\sqrt{0,00138361} = 0,03719691$
β_{12}	$\sqrt{0,002075} = 0,04555217$
β_{11} , β_{22}	$\sqrt{0,00327601} = 0,05723644$

1.4. Teorema de Gauss - Markov

Si no se asume normalidad el estimador mínimo cuadrático $\hat{\beta}$ es el mejor estimador dentro de los estimadores lineales insesgados de β , en el sentido que es el de la varianza más pequeña.

Demostración:

Conocemos que $\hat{\beta} = (X^T X)^{-1} X^T Y$, sea $A = (X^T X)^{-1} X^T$, entonces $\hat{\beta} = AY$. Demostremos que $\beta = CY$ no es mejor estimador insesgado que $\hat{\beta}$. Sabemos que $E[CY] = CE[Y] = CX\beta$ por ser β insesgado se cumple que $E[\beta] = \beta \implies CX\beta = \beta \implies CX\beta - \beta = \vec{0} \implies (CX - I)\beta = \vec{0}$, como $\beta = 0$, entonces $\implies CX - I = \vec{0} \implies CX = I$. Ahora como C es una matriz y Y un vector columna y por la propiedad $Var(AY) = AVar(Y)A^T$, tenemos que $Var(CY) = CVar(Y)C^T = C\sigma^2 C^T = \sigma^2 CC^T$. Luego como $Var(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ y $CX = I$ y $(CX)^T = X^T C^T = I$, podemos escribir $Var(\hat{\beta}) = CX\sigma^2(X^T X)^{-1}X^T C^T = \sigma^2 CX(X^T X)^{-1}X^T C^T$

por lo tanto

$$\begin{aligned} Var(\beta) - Var(\hat{\beta}) &= \sigma^2 CC^T - \sigma^2 CX(X^T X)^{-1}X^T C^T \\ &= \sigma^2 C(I - X(X^T X)^{-1}X^T)C^T \\ &= \sigma^2 CHC^T \end{aligned}$$

H es simétrica ($H^T = H$) e idempotente ($H^2 = H$), H es definida positiva (Una matriz A se dice que es semidefinida positiva si $Y^T AY \geq 0$ para todo vector $Y \neq 0$. Diremos que es definida positiva si $Y^T AY > 0$ para todo vector $Y \neq 0$), por tanto $CHC^T > 0$.

Entonces $Var(\beta) - Var(\hat{\beta}) > 0 \implies Var(\beta) > Var(\hat{\beta})$. Luego $\hat{\beta}$ es el mejor estimador insesgado de mínima varianza.

1.5. Diseños experimentales y matrices de información

El conjunto $\{x_1, \dots, x_N\}$ de elementos de \mathcal{X} (aunque algunos de los elementos pueden coincidir unos con otros) se llama un diseño exacto (o discreto) de tamaño N .

Tengamos sólo $n < N$ puntos distintos. Supongamos que x_i ocurre r_i veces entre los puntos $\{x_1, \dots, x_N\}$ para $i = 1, \dots, n$ así

$$N = \sum_{i=1}^n r_i$$

Asociamos $w_i = r_i/N$ con cada uno de los puntos x_i , $i = 1, \dots, n$. Así

$$\xi = \begin{pmatrix} x_1 & \dots & x_n \\ w_1 & \dots & w_n \end{pmatrix} \quad (1.2)$$

será llamada diseño exacto (discreto) o diseño n -puntos de tamaño N .

La matriz

$$M(\xi) = \sum_{i=1}^n f(x_i) f^\top(x_i) w_i \quad (1.3)$$

se llama la matriz de información de diseño ξ .

Por el teorema de Gauss-Markov, tenemos

$$Cov(\boldsymbol{\theta}) = \frac{\sigma^2}{N} M^{-1}(\xi)$$

para la matriz varianza de la estimación de mínimos cuadrados.

El diseño ξ es una medida de probabilidad discreta, definido por (1.2), que incluye los puntos del conjunto \mathcal{X} y los coeficientes de peso.

En muchas situaciones prácticas, es imposible de realizar estos diseños y tales diseños se deben considerar como aproximación de algunos diseños discretos.

Vamos a escribir un diseño, concentrado en un número finito de puntos, en la forma (1.2), donde los coeficientes $w_i = \xi(x_i)$ son números positivos arbitrarios tales que $\sum w_i = 1$. La matriz

$$M(\xi) = \sum_{i=1}^n f(x_i) f(x_i)^\top \xi(x_i) \quad (1.4)$$

se le llama la matriz de información del diseño aproximado.

Sea Ξ el conjunto de todos los diseños aproximados y \mathcal{M} el conjunto de matrices de información que les corresponden:

$$\mathcal{M} = \{M; M = M(\xi) \text{ para algunos } \xi \in \Xi\}$$

Sea Ξ_n el conjunto de diseños aproximados, concentrado en n puntos (con pesos distintos de cero).

Las propiedades básicas de matrices de información pueden enunciarse como un teorema.

Teorema 1.5.1. *Propiedades de las matrices de información:*

- (i) *Cualquier matriz de información es definida no negativa (en particular es simétrica).*

Demostración:

La demostración fácilmente se concluye de la definición de matriz no negativa y del hecho de que para todo vector $z \in \mathbb{R}^m$ se cumple

$$z^\top M(\xi)z = \sum_{i=1}^n z^\top f(\mathbf{x}_i)f(\mathbf{x}_i)^\top z \xi(\mathbf{x}_i) = \sum_{i=1}^n \|z^\top f(\mathbf{x}_i)\|^2 \xi(\mathbf{x}_i) \geq 0$$

- (ii) *Si $n < m$, siendo m el número de parámetros, entonces $\det M(\xi) = 0$.*

Demostración:

Supongamos que ξ tiene en su soporte $k < m$ puntos: $\mathbf{x}_1, \dots, \mathbf{x}_k$. En el desarrollo del determinante aparecerán siempre al menos dos columnas iguales y por tanto el determinante ha de ser cero.

- (iii) *El conjunto \mathcal{M} es convexo.*

Demostración:

Es necesario comprobar que para cualquier $\lambda \in [0, 1]$ y para cualquier par de diseños ξ_1 y ξ_2 , la matriz

$$M = (1 - \lambda)M(\xi_1) + \lambda M(\xi_2)$$

pertenece al conjunto \mathcal{M} .

Definimos el diseño ξ según la fórmula

$$\xi = (1 - \lambda)\xi_1 + \lambda\xi_2$$

mostraremos que $M = M(\xi) \in \mathcal{M}$; en efecto,

$$\begin{aligned} M &= (1 - \lambda)M(\xi_1) + \lambda M(\xi_2) \\ &= (1 - \lambda) \sum_{i=1}^n f(\mathbf{x}_i)f(\mathbf{x}_i)^\top \xi_1(\mathbf{x}_i) + \lambda \sum_{i=1}^n f(\mathbf{x}_i)f(\mathbf{x}_i)^\top \xi_2(\mathbf{x}_i) \\ &= \sum_{i=1}^n f(\mathbf{x}_i)f(\mathbf{x}_i)^\top [(1 - \lambda)\xi_1(\mathbf{x}_i) + \lambda\xi_2(\mathbf{x}_i)] \\ &= M(\xi) \end{aligned}$$

Con ello queda demostrado.

1.6. Criterios de optimalidad

Llamemos al diseño ξ no singular si el $\det M(\xi) \neq 0$. Vamos a considerar sólo el caso de la estimación de todo el conjunto de parámetros. Aquí, sólo los diseños no singulares son de interés. El teorema de Gauss-Markov es válido para ello.

Típicamente, no existe un diseño $\hat{\xi}$ tal que la matriz

$$M^{-1}(\hat{\xi}) - M^{-1}(\xi),$$

es no negativa, donde ξ es un diseño arbitrario. Por lo tanto, algunas funciones de matrices de información, que tienen sentido estadístico, se utilizan como los criterios de optimalidad.

Consideremos algunos criterios de optimalidad de nuestro interés.

1.6.1. D – criterio

El criterio de diseño más usado en las aplicaciones es el de D -optimalidad, en el que la varianza generalizada de las estimaciones de los parámetros, o su logaritmo se reduce al mínimo.

El diseño $\xi = \arg \max_{\xi} |M(\xi)| = \arg \min_{\xi} |(M(\xi))^{-1}|$ se llama diseño D -óptimo.

El funcional

$$\Psi[M(\xi)] = -\text{Log}|M(\xi)|$$

se llama criterio D -óptimo; donde el funcional Ψ definido en el espacio de elementos de la matriz de información de $M \in \mathcal{M}$, se le denomina criterio de optimización, ver Kiefer(1974).

D -optimalidad, criterio del determinante, equivale a minimizar el volumen del elipsoide de variación de los estimadores lineales e insesgados de los parámetros desconocidos.

1.6.2. G – criterio

Para diseños continuos la varianza normalizada de la respuesta predicha es

$$d(\mathbf{x}, \xi) = \mathbf{f}^T(\mathbf{x})M^{-1}(\xi)\mathbf{f}(\mathbf{x})$$

El criterio G -óptimo es de la forma

$$\max_{\mathbf{x} \in \mathcal{X}} d(\mathbf{x}, \xi) \rightarrow \inf_{\xi}.$$

Tenga en cuenta que para el diseño discreto normado ξ ,

$$d(\mathbf{x}, \xi) = \frac{\sigma^2}{N} \text{Cov}(\mathbf{f}(\mathbf{x})^\top \boldsymbol{\theta});$$

es decir $d(\mathbf{x}, \xi)$ es igual (a la constante de precisión) a la varianza de un valor, predicha por el modelo en el punto \mathbf{x} .

1.6.3. D_s – criterio

Son apropiados cuando el interés es estimar un subconjunto de s parámetros de todo el p – vector $\boldsymbol{\beta}$. Por lo tanto sin perdida de generalidad, los términos del modelo se

pueden dividir en dos grupos

$$E(Y) = f^T(\mathbf{x})\boldsymbol{\beta} = f_1^T(\mathbf{x})\boldsymbol{\beta}_1 + f_2^T\boldsymbol{\beta}_2$$

donde $\boldsymbol{\beta}_1 \in \mathbb{R}^s$ y $\boldsymbol{\beta}_2 \in \mathbb{R}^{p-s}$. Los $\boldsymbol{\beta}_1$ son los parámetros de interés y los $p-s$ parámetros $\boldsymbol{\beta}_2$ suelen ser tratados como parámetros molestia.

Un ejemplo es cuando $\boldsymbol{\beta}_1$ corresponde a los factores experimentales y $\boldsymbol{\beta}_2$ corresponde a los parámetros de las variables de bloqueo (Ver ejemplos, capítulo 15 en *Optimum Experimental Designs, with SAS* de A.C. Atkinson, A.N. Donev y R.D. Tobias). Un segundo ejemplo es cuando los experimentos están diseñados para comprobar la bondad de ajuste de un modelo.

Para obtener expresiones para el criterio de diseño y función de varianza relacionadas, dividimos la matriz de información como

$$M(\xi) = \begin{pmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{12}^T(\xi) & M_{22}(\xi) \end{pmatrix}$$

La matriz de covarianza para la estimación de mínimos cuadrados de $\boldsymbol{\beta}_1$ es $M^{11}(\xi)$, la submatriz superior izquierda $s \times s$ de $M^{-1}(\xi)$. Se puede verificar, a partir de los resultados de la inversa de una matriz particionada (por ejemplo, Fedorov 1972, p. 24), que

$$M^{11}(\xi) = \left\{ M_{11}(\xi) - M_{12}(\xi)M_{22}^{-1}(\xi)M_{12}^T(\xi) \right\}^{-1}$$

En consecuencia, el diseño D_s -óptimo para $\boldsymbol{\beta}_1$ maximiza el determinante

$$|M_{11}(\xi) - M_{12}(\xi)M_{22}^{-1}(\xi)M_{12}^T(\xi)| = \frac{|M(\xi)|}{|M_{22}(\xi)|}.$$

Por otro lado, en la práctica se pueden utilizar teoremas que proporcionan herramientas para la construcción y el control de la optimización de un diseño. Si la atención se limita al parámetro $\boldsymbol{\beta}$, se considera el conocido teorema general de equivalencia (cf. Kiefer y Wolfowitz(1990); cf. Silvey (1980); cf. Pukelsheim(2006)).

1.6.4. Teorema de equivalencia

El siguiente resultado de Kiefer y Wolfowitz (1960) es de gran importancia en la teoría del diseño experimental óptimo.

Teorema 1.6.1. (*Kiefer-Wolfowitz. Teorema de equivalencia*) Para el modelo (1.1), existe un diseño D -óptimo bajo los supuestos clásicos de regresión y las siguientes condiciones son equivalentes:

- (i) ξ^* es un diseño D -óptimo.
- (ii) ξ^* es un diseño G -óptimo.
- (iii) $\max_{x \in \mathcal{X}} d(x, \xi^*) = m$.

Por otra parte, todos los D -óptimos diseños tienen la misma matriz de información, y la función de predicción de la varianza $d(x, \xi^*)$ alcanza su máximo en los puntos de cualquier diseño D -óptimo con soporte finito.

Vale la pena subrayar que el teorema es cierto para los diseños que sean D -óptimo en la clase de diseño aproximado.

Este teorema no sólo establece la equivalencia entre D y G -criterios, sino que también da la importante condición necesaria y suficiente de D -optimalidad: Diseño ξ^* es D -óptimo si y sólo si $\max_{x \in \mathcal{X}} d(x, \xi^*) = m$.

la demostración del teorema se puede encontrar en Kiefer y Wolfowitz (1960). Muchos análogos del teorema de Kiefer-Wolfowitz se pueden encontrar en Kiefer (1974), y aparece en forma más general en Whittle (1973).

Ejemplo 3

Sea $\eta(x_j, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3$; un modelo polinomial con $x \in [-1, 1]$.

En el caso D -óptimo, se verificará a continuación que el diseño

$$\xi^* = \begin{pmatrix} -1 & -\sqrt{5}/5 & \sqrt{5}/5 & 1 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

es un diseño D -óptimo.

En efecto, bastará con mostrar que el diseño ξ^* verifica las condiciones del criterio de D -óptimidad. Primero note que su matriz de momento es

$$M(\xi^*) = \sum_{x \in \{-1, -\sqrt{5}/5, \sqrt{5}/5, 1\}} [1, x, x^2, x^3]^\top [1, x, x^2, x^3] (1/4)$$

$$= \begin{pmatrix} 1 & 0 & 3/5 & 0 \\ 0 & 3/5 & 0 & 13/25 \\ 3/5 & 0 & 13/25 & 0 \\ 0 & 13/25 & 0 & 63/125 \end{pmatrix}$$

Ahora

$$M^{-1}(\xi^*) = \begin{pmatrix} 13/4 & 0 & -15/4 & 0 \\ 0 & 63/4 & 0 & -65/4 \\ -15/4 & 0 & 25/4 & 0 \\ 0 & 65/4 & 0 & 75/4 \end{pmatrix}$$

Luego

$$d(x, \xi^*) = f^\top(x) M^{-1}(\xi^*) f(x)$$

$$= \frac{75}{4}x^6 - \frac{105}{4}x^4 + \frac{33}{4}x^2 + \frac{13}{4}$$

Por último

$\max d(x, \xi^*) = 4$ y dicha función tiene sus puntos críticos en los valores del diseño.

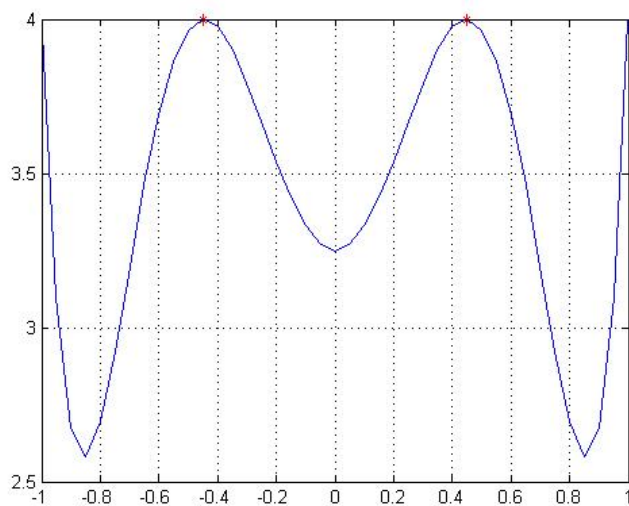


Figura 1.1 Ejemplo 3

Por lo tanto ξ^* es D -óptimo.

Diseños óptimos en la presencia de efectos de bloques aleatorios

Para un modelo lineal en la presencia de efectos de bloques aleatorios se describe la situación donde se tienen b bloques, cada uno con m_i observaciones. Por lo tanto, la j -ésima observación Y_{ij} al bloque i se puede escribir como

$$Y_{ij} = \gamma_i + \beta_0 + \mathbf{f}(\mathbf{x}_{ij})^\top \boldsymbol{\beta} + \epsilon_{ij} \quad (2.1)$$

donde \mathbf{x}_{ij} son los puntos experimentales, $j = 1, \dots, m_i$, $\mathbf{f} = (f_1, \dots, f_p)^\top$ es un conjunto de funciones (conocidas) de regresión y $\beta_0 \in \mathbb{R}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ son los parámetros desconocidos. f puede ser la función identidad para regresión lineal simple.

El término γ_i es el efecto del i -ésimo bloque aleatorio con $E(\gamma_i) = 0$ y $Var(\gamma_i) = \sigma_\gamma^2$. Los errores de observación aleatorio ϵ_{ij} se supone que son homoscedasticos, $E(\epsilon_{ij}) = 0$, $Var(\epsilon_{ij}) = \sigma^2$ y $Cov(\gamma_i, \epsilon_{ij}) = 0$. El análisis adicional dependerá del cociente de varianza $d = \sigma_\gamma^2 / \sigma^2$. Nos centraremos en los parámetros de la población $\theta = (\beta_0, \boldsymbol{\beta}^\top)^\top$. Asumiremos que el número de observaciones por bloque es constante, es decir, $m_i = m$.

Denotemos por $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^\top$ el vector de observaciones para el bloque i . La matriz de covarianza correspondiente $Cov(\mathbf{Y}_i) = \sigma^2 \mathbf{V}$ es completamente simétrica, $\mathbf{V} = \mathbf{I}_m + d \mathbf{1}_m \mathbf{1}_m^\top$, donde \mathbf{I}_m indica la matriz identidad $m \times m$ y $\mathbf{1}_m$ es un vector de longitud m con todas las entradas iguales a uno. El efecto fijo individual de la matriz de diseño $\mathbf{X}_i = (\mathbf{1}_m \mid \mathbf{F}_i)$ se puede descomponer en la primera columna de unos correspondiente a la intersección β_0 y la matriz de diseño para el vector de parámetros $\boldsymbol{\beta}$.

La inversa de \mathbf{V} la podemos hallar mediante álgebra matricial

$$\begin{aligned}
 \mathbf{V}^{-1} &= (\mathbf{I}_m + d\mathbf{1}_m\mathbf{1}_m^\top)^{-1} \\
 &= \mathbf{I} - d^{\frac{1}{2}}\mathbf{1}_m(\mathbf{I} + d\mathbf{1}_m^\top\mathbf{I}\mathbf{1}_m)^{-1}d^{\frac{1}{2}}\mathbf{1}_m^\top \\
 &= \mathbf{I} - d\mathbf{1}_m(\mathbf{I} + dm\mathbf{I})^{-1}\mathbf{1}_m^\top \\
 &= \mathbf{I} - d\mathbf{1}_m(\mathbf{I}(1 + dm))^{-1}\mathbf{1}_m^\top \\
 &= \mathbf{I} - \frac{d}{1 + dm}\mathbf{1}_m\mathbf{1}_m^\top
 \end{aligned} \tag{2.2}$$

Luego, la matriz de información por bloque, utilizando (2.2) queda $\mathbf{X}_i^\top \mathbf{V}^{-1} \mathbf{X}_i = \mathbf{X}_i^\top \mathbf{X}_i - \frac{d}{1+md} \mathbf{X}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{X}_i$ que es proporcional a la inversa de la matriz de varianza-covarianza $Cov(\hat{\boldsymbol{\theta}}_i)$ si \mathbf{X}_i es de rango completo. Así $\boldsymbol{\theta}$ es estimado sobre una base por bloque

$$\begin{aligned}
 \hat{\boldsymbol{\theta}}_i &= (\mathbf{X}_i^\top \mathbf{V}^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i^\top \mathbf{V}^{-1} \mathbf{Y}_i \\
 &= (\mathbf{X}_i^\top \mathbf{X}_i)^{-1} \mathbf{X}_i^\top \mathbf{Y}_i
 \end{aligned}$$

Sobre la base de la población el mejor estimador lineal insesgado se puede calcular como $\hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^b \mathbf{X}_i^\top \mathbf{V}^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{V}^{-1} \mathbf{X}_i \hat{\boldsymbol{\theta}}_i$ si d es conocido. Entonces $Cov(\hat{\boldsymbol{\theta}}) = \sigma^2 \mathbf{M}_d^{-1}$, donde $\mathbf{M}_d = \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{V}^{-1} \mathbf{X}_i$ es la matriz de información sobre la base de la población. El subíndice d indica la dependencia del cociente de varianzas d . Como $\mathbf{M}_d = \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{X}_i - \frac{d}{1 + md} \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{X}_i$.

La matriz de información particionada de acuerdo a β_0 y $\boldsymbol{\beta}$, es

$$\begin{aligned}
M_d &= \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{X}_i - \frac{d}{1+md} \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{X}_i \\
&= \sum_{i=1}^b (\mathbf{1}_m \mid \mathbf{F}_i)^\top (\mathbf{1}_m \mid \mathbf{F}_i) - \frac{d}{1+md} \sum_{i=1}^b (\mathbf{1}_m \mid \mathbf{F}_i)^\top \mathbf{1}_m \mathbf{1}_m^\top (\mathbf{1}_m \mid \mathbf{F}_i) \\
&= \sum_{i=1}^b \begin{pmatrix} \mathbf{1}_m^\top \\ \mathbf{F}_i^\top \end{pmatrix} (\mathbf{1}_m \mid \mathbf{F}_i) - \frac{d}{1+md} \sum_{i=1}^b \begin{pmatrix} \mathbf{1}_m^\top \\ \mathbf{F}_i^\top \end{pmatrix} \mathbf{1}_m \mathbf{1}_m^\top (\mathbf{1}_m \mid \mathbf{F}_i) \\
&= \sum_{i=1}^b \begin{pmatrix} m & \mathbf{1}_m^\top \mathbf{F}_i \\ \mathbf{F}_i^\top \mathbf{1}_m & \mathbf{F}_i^\top \mathbf{F}_i \end{pmatrix} - \frac{d}{1+md} \sum_{i=1}^b \begin{pmatrix} m & \\ \mathbf{F}_i^\top \mathbf{1}_m & \end{pmatrix} (m \mid \mathbf{1}_m^\top \mathbf{F}_i) \\
&= \sum_{i=1}^b \begin{pmatrix} m & \mathbf{1}_m^\top \mathbf{F}_i \\ \mathbf{F}_i^\top \mathbf{1}_m & \mathbf{F}_i^\top \mathbf{F}_i \end{pmatrix} - \frac{d}{1+md} \sum_{i=1}^b \begin{pmatrix} m^2 & m \mathbf{1}_m^\top \mathbf{F}_i \\ \mathbf{F}_i^\top \mathbf{1}_m m & \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \end{pmatrix} \\
&= \frac{1}{1+md} \sum_{i=1}^b \begin{pmatrix} m + m^2 d & \mathbf{1}_m^\top \mathbf{F}_i + md \mathbf{1}_m^\top \mathbf{F}_i \\ \mathbf{F}_i^\top \mathbf{1}_m + md \mathbf{F}_i^\top \mathbf{1}_m & \mathbf{F}_i^\top \mathbf{F}_i + md \mathbf{F}_i^\top \mathbf{F}_i \end{pmatrix} \\
&\quad - \frac{1}{1+md} \sum_{i=1}^b \begin{pmatrix} dm^2 & dm \mathbf{1}_m^\top \mathbf{F}_i \\ d \mathbf{F}_i^\top \mathbf{1}_m m & d \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \end{pmatrix} \\
M_d &= \frac{1}{1+md} \left(\begin{array}{c|c} bm & \sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \\ \hline \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m & (1+md) \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - d \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \end{array} \right) \quad (2.3)
\end{aligned}$$

Si el interés está en los efectos fijos β solamente, entonces las reglas para invertir las correspondientes matrices de información parcial particionadas $M_{\beta,d}^{-1} = Cov(\hat{\beta})/\sigma^2$ es igual a

$$M_{\beta,d} = \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - \frac{d}{1+md} \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i - \frac{1}{bm} \frac{1}{1+md} \left(\sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \right) \left(\sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \right) \quad (2.4)$$

También consideramos los modelos límites para $d = 0$ y $d \rightarrow \infty$, respectivamente: Para $d = 0$ obtenemos los modelos de efectos fijos y sin interceptos de bloques

$$Y_{ij} = \beta_0 + \mathbf{f}(\mathbf{x}_{ij})^\top \beta + \epsilon_{ij} \quad (2.5)$$

Obviamente, M_d tiende a $M_0 = \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{X}_i$ para $d \rightarrow 0$. Del mismo modo, $M_{\beta,d}$ tiende a

$$M_{\beta,0} = \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - \frac{1}{bm} \left(\sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \right) \left(\sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \right) \quad (2.6)$$

Para $d \rightarrow \infty$ introducimos el modelo de efectos fijos con bloques fijos

$$Y_{ij} = \mu_i + \mathbf{f}(\mathbf{x}_{ij})^\top \beta + \epsilon_{ij}; \quad (\mu_i = \gamma_i + \beta_0) \quad (2.7)$$

Aquí, el vector de parámetros $(\mu_1, \dots, \mu_b, \beta_1, \dots, \beta_p)^\top$ tiene dimensión $b + p$ y la matriz de información correspondiente tiene la forma

$$M_\infty = \left(\begin{array}{ccc|c} & & & \mathbf{1}_m^\top \mathbf{F}_1 \\ & & & \vdots \\ & & & \mathbf{1}_m^\top \mathbf{F}_b \\ \hline & m\mathbf{I}_b & & \\ \hline \mathbf{F}_1^\top & \cdots & \mathbf{F}_b^\top \mathbf{1}_m & \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i \end{array} \right) \quad (2.8)$$

Para β la matriz de información parcial correspondiente se puede calcular

$$\mathbf{M}_{\beta,\infty} = \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - \frac{1}{m} \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \quad (2.9)$$

De ahí se obtiene el siguiente resultado, que establece la matriz de información parcial $\mathbf{M}_{\beta,d}$.

Lema1.

$$\mathbf{M}_{\beta,d} = \frac{1}{1+md} \mathbf{M}_{\beta,0} + \frac{md}{1+md} \mathbf{M}_{\beta,\infty} \quad (2.10)$$

Tenga en cuenta que la matriz de información parcial $\mathbf{M}_{\beta,d}$ tiende a $\mathbf{M}_{\beta,\infty}$ cuando d tiende a ∞ .

2.1. Aspectos del diseño

La calidad de los estimadores $\hat{\theta}$ y $\hat{\beta}$ depende de la configuración experimental \mathbf{x}_{ij} , $i = 1, \dots, b$, $j = 1, \dots, m$, a través de las matrices de información \mathbf{M}_d y $\mathbf{M}_{\beta,d}$, respectivamente. El objetivo en el diseño experimental es elegir los puntos de una región diseño \mathcal{X} con el fin de minimizar la covarianza $\text{Cov}(\hat{\theta})$ o $\text{Cov}(\hat{\beta})$ o partes de ella, lo cual es equivalente a maximizar las correspondientes matrices de información \mathbf{M}_d o $\mathbf{M}_{\beta,d}$ respectivamente. Como esas matrices no están completamente ordenadas, una optimización uniforme no es posible, en general. Por lo tanto, algunos funcionales de valores reales que ponen énfasis en las propiedades particulares de los estimadores se optimizarán. El criterio de diseño más usado es el D -criterio, que tiene como objetivo maximizar el determinante de la matriz de información \mathbf{M}_d . Esto es equivalente a minimizar el volumen de un elipsoide de confianza para θ bajo la suposición de normalidad.

Si el interés está en los efectos β solamente, D_β -optimalidad se define en términos de la determinante de la inversa $\mathbf{M}_{\beta,d}^{-1}$ de la correspondiente matriz de información parcial. Como se ve a continuación,

$$\begin{aligned}
\det(\mathbf{M}_d) &= \left| \frac{1}{1+md} \left(\begin{array}{c|c} bm & \sum_{i=1}^b \mathbf{1}_m^T \mathbf{F}_i \\ \hline \sum_{i=1}^b \mathbf{F}_i^T \mathbf{1}_m & (1+md) \sum_{i=1}^b \mathbf{F}_i^T \mathbf{F}_i - d \sum_{i=1}^b \mathbf{F}_i^T \mathbf{1}_m \mathbf{1}_m^T \mathbf{F}_i \end{array} \right) \right| \\
&= \left(\frac{1}{1+md} \right)^{p+1} |bm| \left| (1+md) \sum_{i=1}^b \mathbf{F}_i^T \mathbf{F}_i - d \sum_{i=1}^b \mathbf{F}_i^T \mathbf{1}_m \mathbf{1}_m^T \mathbf{F}_i \right. \\
&\quad \left. - \sum_{i=1}^b \mathbf{F}_i^T \mathbf{1}_m (bm)^{-1} \sum_{i=1}^b \mathbf{1}_m^T \mathbf{F}_i \right| \\
&= \left(\frac{1}{1+md} \right)^{p+1} |bm| \left| \sum_{i=1}^b \mathbf{F}_i^T \mathbf{F}_i + md \sum_{i=1}^b \mathbf{F}_i^T \mathbf{F}_i - d \sum_{i=1}^b \mathbf{F}_i^T \mathbf{1}_m \mathbf{1}_m^T \mathbf{F}_i \right. \\
&\quad \left. - \frac{1}{bm} \sum_{i=1}^b \mathbf{F}_i^T \mathbf{1}_m \sum_{i=1}^b \mathbf{1}_m^T \mathbf{F}_i \right| \\
&= \left(\frac{1}{1+md} \right)^{p+1} |bm| \left| \sum_{i=1}^b \mathbf{F}_i^T \mathbf{F}_i - \frac{1}{bm} \sum_{i=1}^b \mathbf{F}_i^T \mathbf{1}_m \sum_{i=1}^b \mathbf{1}_m^T \mathbf{F}_i \right. \\
&\quad \left. + md \left(\sum_{i=1}^b \mathbf{F}_i^T \mathbf{F}_i - \frac{1}{m} \sum_{i=1}^b \mathbf{F}_i^T \mathbf{1}_m \mathbf{1}_m^T \mathbf{F}_i \right) \right| \\
&= \left(\frac{1}{1+md} \right)^{p+1} |bm| |\mathbf{M}_{\beta,0} + md \mathbf{M}_{\beta,\infty}| \\
&= \left(\frac{1}{1+md} \right)^{p+1} |bm| \left| (1+md) \left(\frac{1}{1+md} \mathbf{M}_{\beta,0} + \frac{md}{1+md} \mathbf{M}_{\beta,\infty} \right) \right| \\
&= \left(\frac{1}{1+md} \right)^{p+1} |bm| |(1+md) \mathbf{M}_{\beta,d}| \\
&= \left(\frac{1}{1+md} \right)^{p+1} |bm| (1+md)^p |\mathbf{M}_{\beta,d}| \\
&= \frac{bm}{1+md} \det(\mathbf{M}_{\beta,d})
\end{aligned}$$

sujeta por la fórmula para el determinante de matrices particionadas. Por lo tanto, D y D_β — optimalidad coinciden también en modelos de interceptos aleatorios, un hecho bien conocido en el ajuste de efectos fijos.

Lema2. *Un diseño (x_{ij}) es D -óptimo si y sólo si es D_β -óptimo.*

Si tenemos en cuenta los diseños que son uniformes en todos los bloques, es decir, en que los parámetros experimentales son los mismos para cada bloque, $x_{ij} \equiv x_j$, entonces la situación se simplifica radicalmente. En este caso las matrices de diseños individuales coinciden, $\mathbf{F}_i = \mathbf{F}_1$ y $\mathbf{X}_i = \mathbf{X}_1$, respectivamente, y \mathbf{X}_1 tiene que ser de rango columna completa para permitir estimabilidad de $\boldsymbol{\theta}$. Además, $\hat{\boldsymbol{\theta}} = \frac{1}{b} \sum_{i=1}^b \hat{\boldsymbol{\theta}}_i$ se reduce a la media de los valores ajustados de forma individual para los parámetros.

La matriz de covarianza estandarizada \mathbf{M}_d^{-1} se descompone de forma aditiva en la matriz correspondiente \mathbf{M}_0^{-1} para el modelo de efectos fijos y sin intercepciones individuales y la variabilidad de la intersección aleatoria (véase, por ejemplo, Entholzner y otros., (2005)). Para la matriz de información reducida observamos

$$\mathbf{M}_{\beta,0} = b \left(\mathbf{F}_1^\top \mathbf{F}_1 - \frac{1}{m} \mathbf{F}_1^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_1 \right) = \mathbf{M}_{\beta,\infty} \quad (2.11)$$

y, en consecuencia, por el Lema 1 $\mathbf{M}_{\beta,d} = \mathbf{M}_{\beta,0}$ es independiente de d . Así, el diseño D -óptimo para el modelo de efectos fijos y sin intersecciones individuales es D -y D_β -óptimo para cada $d \geq 0$ visto en el Lema 2.

Ejemplo de aplicación del modelo propuesto

Consideremos el modelo de regresión cuadrático en dos variables sin interacciones

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \epsilon_i; \quad (x_{1i}, x_{2i}) \in [-1, 1] \times [-1, 1].$$

El diseño $\boldsymbol{\xi}$ el cual asigna iguales pesos $\frac{1}{9}$ a las cuatro esquinas $(\pm 1, \pm 1)$, a los cuatro puntos centrales de los lados $(0, \pm 1)$; $(\pm 1, 0)$ y al punto central $(0, 0)$ de la región experimental. Este diseño es D -óptimo para este modelo. Si el diseño $\boldsymbol{\xi}$ es bloqueado como sigue

$$\xi_1 = \begin{pmatrix} (-1, 0) & (0, 1) & (1, -1) \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} (-1, -1) & (0, 0) & (1, 1) \\ 1/3 & 1/3 & 1/3 \end{pmatrix},$$

$$\xi_3 = \begin{pmatrix} (-1, 1) & (1, 0) & (0, -1) \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

Entonces por el lema 2, ξ es D -óptimal para el correspondiente modelo en la presencia de efectos de bloques con respuesta

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_{11} x_{1ij}^2 + \beta_{22} x_{2ij}^2 + \gamma_i + \epsilon_{ij},$$

en la j -ésima corrida sobre el bloque i , ($i = 1, 2, 3$; $j = 1, 2, 3$) con pesos dados por $\xi(i, (x_{1ij}, x_{2ij})) = \frac{1}{3}\xi_i(x_{1ij}, x_{2ij})$. Además este diseño D -óptimal por bloques no depende del cociente de varianza d .

Conclusiones y trabajos futuros

En el presente trabajo se desarrolla la matriz de información de los criterios fijos en un modelo de regresión lineal en la presencia de efectos de bloques aleatorios como una combinación convexa de las matrices de información de los modelos límites cuando la varianza del efecto de bloque es cero o tiende a infinito.

Por otro lado, se muestra que los diseños óptimos para modelos de efectos fijos también son óptimos para modelos en la presencia de efectos de bloques aleatorios siempre que los bloques sean uniformes.

D y D_{β} -optimalidad coinciden también en modelos en la presencia de efectos de bloques aleatorios, un hecho ya conocido en los escenarios con efectos fijos.

Para trabajos futuros se pueden considerar modelos en la presencia de efectos de bloques aleatorios donde los parámetros de regresión interactúan con diferentes grupos o tratamientos.

Bibliografía

- [1] Atkins and Cheng. (1998). Optimal regression designs in the presence of random block effects. *JSPI* 77, 321-335.
- [2] Cheng. Optimal regression designs under random blocks-effects models. (1995). *Statistica Sinica* 5, 485-497.
- [3] Debusho and Haines. (2007). V- and D-optimal population designs for the simple linear regression model with a random intercept term. *Journal of Statistical Planning and inference* 138, 1116-1130.
- [4] Entholzner, M., Benda, N., Schwabe, R. (2005). A note on designs for estimating population parameters. *Biometrical Letters* 42, 25-41.
- [5] Fedorov, V.V. (1972) *Theory of Optimal Experiments*. Academic Press, New York.
- [6] Fedorov, V.V. (1997). *Model-Oriented Design of Experiments*. Springer, New York.
- [7] Goos, Peter. (2002) *The Optimal Design of Blocked and Split-Plot*. Springer, New York.
- [8] GraBhoff, U and Schwabe, R. (2003). On the analysis of paired observations. *Statistics and Probability Letters* 65, 7-12.
- [9] Norell, L. (2006). Optimal designs for maximum likelihood estimators in the one-way random model. U.U.D.M. Report 2006:24. Department of Mathematics, Uppsala University.
- [10] Schmelter, T. and Schwabe, R. (2008). On optimal designs in random intercept models. *Tetra Mountains Mathematical Publications*. 39, 145-53.
- [11] Schwabe, R. (1996). *Optimum Designs for Multi-Factor Models*. Springer, New York.

- [12] Silvey, S.D. (1980). Optimal Designs. Chapman & Hall, London.
- [13] Van Breukelen, G.J.P., Candel, M.J.J.M. and Berger, M.P.F. (2008). Relative efficiency of unequal cluster sizes for variance component estimation in cluster randomized and multicentre trials. *Statistical Methods in Medical Research*, 17, 439-58.