

Capítulo 2

Modelos clásicos

En este segundo capítulo, resumimos los aspectos que más caracterizan a los Modelos Lineales Generalizados. Para ello, partimos de los puntos fundamentales del modelo lineal normal los cuales, por confrontación con los no lineales y con los no normales, llevarán naturalmente a justificar la elección de estos modelos frente a las transformaciones clásicas de la respuesta. De este modo, presentamos el modelo logístico como caso particular de los Modelos Lineales Generalizados y explicamos las bases que llevan a definir los diseños experimentales para esta clase de modelos.

2.1. Introducción

Si bien vivimos en un mundo profundamente *no lineal* —que también muchas veces puede ser etiquetado como *no normal*¹— la mayor parte de los modelos que se aplican al análisis de datos resultan ser, paradójicamente, más lineales que no lineales. Ante esta circunstancia, también suele darse muy a menudo la utilización de modelos en donde la respuesta tiene una distribución normal. Así, el *modelo lineal normal* suele ser un punto de partida muy utilizado para introducir extensiones y generalizaciones del modelo clásico. Dichas extensiones permiten ampliar los alcances hacia nuevos problemas que el original no es del todo adecuado, y particularmente a la hora de estudiar la búsqueda de modelos que no se ajusten a aquellas “condiciones clásicas”. En este capítulo, partiremos también del enfoque clásico y describiremos las notas más características de los modelos que más se aplican a los datos de naturaleza binaria.

En el contexto de nuestro trabajo, y sin pretender hacer una clasificación exhaustiva, diremos que podemos agrupar estos modelos en dos clases principales:

¹Incluso más, en el que no queda muy claro lo que significa “normal”, cuanto menos, en el sentido *lato* del término.

- Modelos para datos *normales*, y
- Modelos para datos *no normales*.

Mediante una breve síntesis sobre los detalles más destacados de éstos —y en particular, de la segunda clase de modelos— se podrá comprender mejor el contexto en que utilizaremos los modelos para datos binarios, de naturaleza inmarcesiblemente² no normal y no lineal. En el **Apéndice B**, resumimos las principales características sobre los modelos que resultan adecuados para los datos correspondientes.

Como comentario final, no realizaremos una presentación exhaustiva en este capítulo sobre los modelos que utilizaremos, ya que se trata solamente de presentar qué aspectos son clave en los modelos que utilizaremos luego para estudiar los diseños para *MSR*. La forma en que lo haremos será partiendo del siempre útil modelo normal, tanto en sus formas lineal como no lineal, para dar luego paso a una presentación de los *MLG*, que son los que hemos elegido para realizar este trabajo. Y por último, destacaremos los aspectos más importantes de los *MDB* como caso particular de los *MLG*, que serán aquellos de los que más sacaremos provecho.

2.2. Modelos normales

En general, la relación entre una respuesta y un conjunto de factores en el contexto de un modelo que siga una distribución normal puede expresarse mediante una relación de naturaleza vectorial que, abusando ligeramente de la notación, tendrá la siguiente forma³:

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \quad (2.1)$$

en donde:

- \mathbf{y} es la variable respuesta, un vector de dimensión $n \times 1$,
- \mathbf{x} representa la lista de los n vectores columna, cada uno de dimensión $k \times 1$, con los niveles de cada uno de los p factores de variabilidad considerados en el modelo⁴.

²Aunque la obvia excepción es el propio *modelo lineal normal*, que es un caso particular de los *MLG*.

³*Vid.* p.ej. BOX Y LUCAS (1959) y BOX Y DRAPER (1971), entre otros.

⁴Puesto que el tema de la formulación de la dimensionalidad suele ser dispar entre los principales textos de modelado estadístico, nos vemos conducidos a indicar la notación que utilizaremos en este trabajo. Una de las características que tienen los *MLG* es que al definir su predictor lineal, cada término tendrá un coeficiente β_j para cada variable explicativa x_j . Este hecho hace que tanto el vector de coeficientes, $\boldsymbol{\beta}$, como el de los factores, \mathbf{x} , tengan igual número de componentes. Es frecuente encontrar que cuando se considera un conjunto de k factores de variabilidad para un experimento, el número de parámetros suele ser no k sino $k + 1$ debido a que la primera columna de la matriz de datos

- $\boldsymbol{\beta}$ es el vector columna de coeficientes del modelo, que se suponen constantes desconocidas, cuya dimensión será $p \times 1$.
- $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta})$ es la llamada “expectation function”⁵, que será la relación funcional que vincula a los coeficientes con los factores, ya sea mediante un modelo lineal como no lineal. Abusando nuevamente de la notación pero ganando en facilidad de conceptualización, diremos que este vector será de la forma:

$$\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}) = [f(\mathbf{x}_1, \boldsymbol{\beta}), \dots, f(\mathbf{x}_n, \boldsymbol{\beta})]', \quad (2.2)$$

con:

$$\dim[\mathbf{f}(\mathbf{x}, \boldsymbol{\beta})] = n \times 1,$$

en donde la i -ésima componente del mismo representará la relación que existe entre la respuesta y los niveles de los factores considerados para la i -ésima condición experimental, cuyo valor esperado es:

$$E(y_i | \mathbf{x}_i) = f(\mathbf{x}_{ij}, \boldsymbol{\beta}), \quad \text{para: } i = 1, \dots, n \quad \text{y} \quad j = 1, \dots, p.$$

- $\boldsymbol{\varepsilon}$ es una componente aleatoria $DIIN(\mathbf{0}, \sigma^2 \mathbf{I})$, de dimensión $n \times 1$, de la que se supone también que σ^2 es constante y desconocida.

Queda claro que el valor esperado de la respuesta de acuerdo con el modelo expresado en la (2.1) será la misma “expectation function”, ya que:

$$E(\mathbf{y} | \mathbf{x}) = E[\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}] = E[\mathbf{f}(\mathbf{x}, \boldsymbol{\beta})] + E[\boldsymbol{\varepsilon}] = \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}).$$

La varianza de la respuesta será la misma que la de la componente aleatoria⁶, pues partimos de la hipótesis que la componente sistemática es de naturaleza constante:

$$\mathbf{V}(\mathbf{y} | \mathbf{x}) = \mathbf{V}[\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}] = \mathbf{V}(\boldsymbol{\varepsilon}),$$

es decir:

$$\mathbf{V}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I} \quad (2.3)$$

Si llamamos f_{ij} al elemento correspondiente a la i -ésima condición experimental ($i = 1, \dots, n$) y al j -ésimo elemento del vector de coeficientes $\boldsymbol{\beta}$, se define la *matriz de derivadas de $f(\mathbf{x}, \boldsymbol{\beta})$* , de n filas por $p = k + 1$ columnas, denotada por \mathbf{F} :

$$\mathbf{F} = \{f_{ij}\} = \left\{ \frac{\partial}{\partial \beta_j} [f(\mathbf{x}_i, \boldsymbol{\beta})] \right\}, \quad \text{para: } i = 1, \dots, n \quad \text{y} \quad j = 1, \dots, p.$$

\mathbf{X} es una columna de unos, que se corresponde con el término independiente del vector de coeficientes, es decir, con el parámetro β_0 . De este modo, en lo que sigue de este trabajo —salvo que indiquemos lo contrario— diremos que los k factores de variabilidad tienen asociado un vector de coeficientes de dimensión $p = k + 1$.

⁵O también *componente sistemática* o *parte estructural* del modelo. Vid., p. ej., MYERS *et al.* (2002), cap. 3.

⁶Vid. p. ej.: BATES y WATTS (1988), p. 33.

De forma explícita, esta matriz tendrá el siguiente aspecto⁷:

$$\mathbf{F} = \begin{bmatrix} \frac{\partial}{\partial \beta_0} [f(\mathbf{x}_1, \boldsymbol{\beta})] & \cdots & \frac{\partial}{\partial \beta_j} [f(\mathbf{x}_1, \boldsymbol{\beta})] & \cdots & \frac{\partial}{\partial \beta_k} [f(\mathbf{x}_1, \boldsymbol{\beta})] \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial}{\partial \beta_0} [f(\mathbf{x}_i, \boldsymbol{\beta})] & \cdots & \frac{\partial}{\partial \beta_j} [f(\mathbf{x}_i, \boldsymbol{\beta})] & \cdots & \frac{\partial}{\partial \beta_k} [f(\mathbf{x}_i, \boldsymbol{\beta})] \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial}{\partial \beta_0} [f(\mathbf{x}_n, \boldsymbol{\beta})] & \cdots & \frac{\partial}{\partial \beta_j} [f(\mathbf{x}_n, \boldsymbol{\beta})] & \cdots & \frac{\partial}{\partial \beta_k} [f(\mathbf{x}_n, \boldsymbol{\beta})] \end{bmatrix}, \quad \mathbf{F} \in \mathbb{R}^{n \times (k+1)}$$

Para el caso de una superficie de respuesta como la que estudiaremos en este trabajo, tendremos que la función f será la misma probabilidad de éxito π , que hemos elegido modelar mediante el modelo logístico.

Cuando se estima el vector de parámetros del modelo normal, ya sea utilizando el criterio de mínimos cuadrados como el de máxima verosimilitud, se comprueba que en ambos casos se obtiene un mismo conjunto de p ecuaciones, llamadas *ecuaciones normales*, que serán de la forma⁸:

$$\frac{1}{\sigma^2} \sum_{i=1}^n [y_i - f(\mathbf{x}_i, \boldsymbol{\beta})] \cdot \left\{ \frac{\partial}{\partial \beta_j} [f(\mathbf{x}_i, \boldsymbol{\beta})] \right\}_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{ML}} = 0, \quad \text{para } j = 1, \dots, p,$$

en donde $\hat{\boldsymbol{\beta}}_{ML} = (\hat{\beta}_{0ML}, \hat{\beta}_{1ML}, \dots, \hat{\beta}_{kML})'$ representa el vector de estimaciones máximo-verosímiles de los parámetros del modelo, el cual suele obtenerse mediante algoritmos iterativos⁹, que la mayor parte de los paquetes estadísticos pueden determinar con precisión razonable. Expresadas vectorialmente, las ecuaciones anteriores —llamadas frecuentemente “score equations”— quedarán como:

$$\frac{1}{\sigma^2} \mathbf{F}' (\mathbf{y} - \hat{\boldsymbol{\mu}}) = \mathbf{0},$$

en donde \mathbf{F}' es la transpuesta de la matriz de derivadas \mathbf{F} , y $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_n)'$ es el vector columna formado por las medias *estimadas* de cada condición experimental, cada una de las cuales resultará:

$$\hat{\mu}_i = E(y_i | \mathbf{x}_i) = f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_{ML}).$$

⁷Para facilitar la utilización posterior de las dimensiones del vector de factores \mathbf{x} , diremos que para aquellos casos en que se considere que la primera componente de dicho vector es 1, entonces tendremos que su dimensión es igual a $p = k + 1$, de las que k de ellas serán los niveles de los factores, x_j , con: $j = 1, \dots, k$. Así, cuando indiquemos que el vector \mathbf{x} posee $p = k + 1$ dimensiones haremos referencia a que su primera componente es un 1, y que el mismo consta de los k factores de variabilidad considerados: x_1, \dots, x_k . Asimismo, al referirnos al vector de parámetros $\boldsymbol{\beta}$, diremos también que el mismo tendrá también p dimensiones. Dada la naturaleza del predictor lineal para el modelo logístico —un parámetro para cada factor (de esto hablaremos más adelante)— tendremos que esta notación es coherente con la utilizada para el vector de factores: tendremos k componentes β_j , con: $j = 1, \dots, k$, más el término independiente, β_0 , lo que totaliza la cantidad de $p = k + 1$ dimensiones consideradas para aquél vector.

⁸Vid. BOX Y LUCAS (1959).

⁹Vid. p. ej.: BATES Y WATTS (1988), pp. 39 *et seq.*, y MYERS *et al.* (2002), pp. 73 *et seq.*

La naturaleza de la estimación máximo verosímil del vector $\boldsymbol{\beta}$ es del tipo asintóticamente insesgada, con distribución asintóticamente normal¹⁰, y con varianza proporcional a la matriz de derivadas, \mathbf{F} . En símbolos:

$$\widehat{\boldsymbol{\beta}}_{ML} \underset{\text{asint.}}{\sim} N \left[\boldsymbol{\beta}, \sigma^2 (\mathbf{F}'\mathbf{F})^{-1} \right], \quad (2.4)$$

en donde:

$$\mathbf{I}(\boldsymbol{\beta}) = \frac{\mathbf{F}'\mathbf{F}}{\sigma^2} \quad (2.5)$$

es la *Matriz de Información de Fisher* (que en adelante abreviaremos como *MIF*). Puede verse claramente que la *Matriz de Varianzas y Covarianzas* del vector de parámetros queda definida en la ecuación (2.4) la que, a su vez, resulta igual a la inversa de la *MIF* de acuerdo con la ecuación (2.5):

$$\mathbf{I}^{-1}(\boldsymbol{\beta}) = \mathbf{V}(\boldsymbol{\beta}) = \sigma^2 (\mathbf{F}'\mathbf{F})^{-1} \quad (2.6)$$

Apoyándonos nuevamente en un ligero abuso de la notación¹¹, diremos que la matriz $\mathbf{I}(\boldsymbol{\beta})$ será igual a la derivada segunda de la función log-verosimilitud de la distribución con respecto a $\boldsymbol{\beta}$, cambiada de signo, es decir:

$$(-1) \frac{\partial^2}{\partial \boldsymbol{\beta}^2} \{ \ln[L(\boldsymbol{\beta} | \mathbf{y})] \} = (-1) \frac{\partial^2}{\partial \boldsymbol{\beta}^2} [\ell(\boldsymbol{\beta})] = -\frac{\mathbf{F}'\mathbf{F}}{\sigma^2} = -\mathbf{I}(\boldsymbol{\beta}) \quad (2.7)$$

Otra forma equivalente de obtener la matriz asintótica de varianzas y covarianzas de $\boldsymbol{\beta}$ es calculando la varianza de las “score equations”¹², que también es llamada la *Matriz Hessiana* evaluada en el valor máximo verosímil de la estimación de $\boldsymbol{\beta}$, denotada como $\mathbf{H}(\widehat{\boldsymbol{\beta}}_{ML})$. En símbolos:

$$\mathbf{H}(\widehat{\boldsymbol{\beta}}_{ML}) = \mathbf{V} \left[\frac{1}{\sigma^2} \mathbf{F}'(\mathbf{y} - \widehat{\boldsymbol{\mu}}) \right] = \frac{\mathbf{F}'\mathbf{F}}{\sigma^2}$$

Cuando la varianza de la distribución normal es desconocida, y el algoritmo de estimación de los coeficientes haya convergido en un cierto valor $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{ML}$, la misma puede estimarse mediante la suma de cuadrados residual¹³, es decir:

$$\widehat{\sigma}^2 = \frac{1}{n-p} (\mathbf{y} - \widehat{\boldsymbol{\mu}})' (\mathbf{y} - \widehat{\boldsymbol{\mu}}), \quad \text{con } \nu = n - p \text{ grados de libertad,}$$

¹⁰ Vid. p. ej.: MYERS *et al.* (2002), pp. 86 *et seq.*

¹¹ *Nota bene*: el abuso de notación de la ecuación (2.7) está dado por la forma en como hemos escrito la derivada de la log-verosimilitud, puesto que el operador $\frac{\partial^2}{\partial \boldsymbol{\beta}^2}$ se refiere a una derivada segunda en la que $\boldsymbol{\beta}$ es de naturaleza vectorial y que armando todos los casos para las componentes de dicho vector, nos hace llegar a una matriz, que es la misma $\mathbf{I}(\boldsymbol{\beta})$. Hecha esta aclaración, nos ha parecido útil expresar esta segunda derivada mediante esta notación, cuyo significado no hemos definido rigurosamente pero sí hemos explicado a qué hace referencia.

¹² Vid. p. ej.: MYERS *et al.* (2002), p. 34.

¹³ Vid. p. ej.: CORNELL (1990), p. 71.

en donde el vector de valores esperados $\hat{\boldsymbol{\mu}} = E(\mathbf{y} \mid \mathbf{x})$ se ha evaluado en el valor que maximiza la log-verosimilitud.

Teniendo en cuenta la ecuación (2.6), a partir de esta estimación se llega a la *matriz asintótica estimada de varianzas y covarianzas* de $\boldsymbol{\beta}$, que será¹⁴:

$$\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}}_{ML}) = \widehat{\sigma}^2(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}. \quad (2.8)$$

Esta última expresión (2.8) resultará general para los dos casos básicos de normalidad, ya sea para modelos no lineales como lineales, ya que lo único que variará será la forma de calcular la matriz de derivadas \mathbf{F} . Veremos en el punto siguiente los resultados para el caso normal y lineal, en donde se simplifica el cálculo de dicha matriz.

2.2.1. Caso normal lineal

Cuando la forma de la componente sistemática del modelo (ecuación 2.2) es del tipo lineal, es decir:

$$f(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta},$$

las cantidades anteriormente vistas toman también aspecto lineal. Si consideramos $i = 1, \dots, n$ condiciones y $j = 1, \dots, k$ factores, se comprueba inmediatamente lo siguiente:

a. Componente sistemática, $f(\mathbf{x}, \boldsymbol{\beta})$:

$$f(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}$$

b. Matriz de derivadas, \mathbf{F} :

$$f_{ij} = \frac{\partial}{\partial \beta_j} [f(\mathbf{x}_i, \boldsymbol{\beta})] = \frac{\partial}{\partial \beta_j} \left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij} \right) = \mathbf{x}_i$$

De este modo, la matriz \mathbf{F} de derivadas será:

$$\mathbf{F} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} = [\mathbf{1}', \mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_k]' = \mathbf{X}$$

¹⁴ Vid. p. ej. MYERS *et al.* (2002), pp. 79 *et seq.*

Por lo tanto, la matriz de derivadas será directamente la matriz de diseño, \mathbf{X} , correspondiente a los niveles que tomarán los k factores (ó $k + 1$ términos) a lo largo de las n condiciones experimentales que se lleven a cabo.

c. **Estimador máximo verosímil del vector de parámetros, $\hat{\beta}_{ML}$:**

$$\hat{\beta}_{ML} \underset{\text{asint.}}{\sim} N \left[\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right]$$

d. **Matriz asintótica de varianzas y covarianzas de $\hat{\beta}_{ML}$, $\mathbf{V}(\hat{\beta}_{ML})$:**

$$\mathbf{V}(\hat{\beta}_{ML}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

e. **Matriz de Información de Fisher, $\mathbf{I}(\hat{\beta}_{ML})$:**

$$\mathbf{I}(\hat{\beta}_{ML}) = \frac{\mathbf{X}'\mathbf{X}}{\sigma^2}$$

Cuando las condiciones del problema muestren que no se cumple la hipótesis de homoscedasticidad, suele ser razonable suponer que existe una cierta matriz cuadrada Σ , conocida, de tal forma que ε es una componente aleatoria $DIIN(0, \sigma^2 \Sigma)$. Comparando la varianza con la ecuación (2.3), se tendrá que:

$$\mathbf{V}(\varepsilon) = \sigma^2 \Sigma. \quad (2.9)$$

En este caso, para obtener un estimador insesgado del vector de coeficientes, debe corregirse el caso homoscedástico mediante la siguiente expresión:

$$\hat{\beta}_{ML} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \mathbf{X}'\Sigma^{-1}\mathbf{y}, \quad (2.10)$$

cuya matriz de varianzas y covarianzas es:

$$\mathbf{V}(\hat{\beta}_{ML}) = \sigma^2 (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}. \quad (2.11)$$

Para el caso particular en donde la matriz Σ resultara diagonal, la misma se puede expresar por medio de una matriz de “pesos” \mathbf{W} . Se puede demostrar¹⁵ que esta matriz es la inversa de la matriz generalizada de varianzas Σ , es decir:

$$\mathbf{W} = \Sigma^{-1}$$

Con esta notación alternativa, las expresiones (2.10) y (2.11) se transforman en sus equivalentes:

$$\hat{\beta}_{ML} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{y}$$

¹⁵ Vid. WEISBERG (2005), DRAPER Y SMITH (1998) o SEARLE (1982), entre otros.

$$\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML}) = \sigma^2 (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$$

Cuando se da esta situación de independencia de las observaciones de la respuesta para cada valor dado de \mathbf{x}_i , es decir, cuando son independientes las observaciones $y_i \mid \mathbf{x}_i$, la matriz de “pesos”¹⁶ resulta:

$$\boldsymbol{\Sigma} = \text{diag}\{V(y_i \mid \mathbf{x}_i)\} = \text{diag}\{\sigma_i^2\}, \text{ para todo } i.$$

Si además se diera el caso adicional que las varianzas fuesen iguales entre sí, se tendría que $\mathbf{W} = \mathbf{I}$, con lo cual se estaría en el caso homoscedástico.

Ejemplos reales de estos casos se desarrollan de forma completa y descriptiva en obras tales como BATES Y WATTS (1986), MYERS *et al.* (2002), VENABLES Y RIPLEY (2002), entre otras.

2.3. Modelos no normales: introducción a los *MLG*

Una primera nota que aparece desde el punto de vista de los *diseños* en modelos no normales es que los mismos no tienen la ventaja de conducir —cuanto menos, fácilmente— a diseños óptimos, que ya hemos comentado anteriormente¹⁷.

En particular, cuando la respuesta tiene naturaleza discreta no podrán considerarse estrictamente los procedimientos aplicables al modelo lineal debido a la falta de cumplimiento de varios de sus puntos de partida. Por ejemplo¹⁸: (a) en el modelo de Bernoulli, la respuesta será del tipo dicotómico, pudiendo tomar valores 0 ó 1 solamente, y (b) en el caso del modelo binomial, la respuesta podrá tomar valores enteros y positivos solamente, ya que ésta es representativa del número de éxitos, y , observados en una cierta muestra de tamaño fijo m . En este tipo de modelos, resulta de particular interés el estudio de la probabilidad de éxito π y los diferentes valores que puede tomar el vector de factores $\mathbf{x} = (x_1, \dots, x_k)'$, de tal forma de encontrar relaciones funcionales que los vinculen entre sí, con funciones de la forma $\pi(\mathbf{x}, \boldsymbol{\beta})$.

En la práctica, no todo suele cuadrar de manera sencilla para el experimentador como para que cualquier sistema tenga un funcionamiento del tipo lineal y normal, cuya resolución ofrece elegantes y completas posibilidades. En opinión de muchos investigadores, una gran mayoría de casos puede englobarse en estos tipos de modelos.

¹⁶Esta matriz de “pesos”, $\mathbf{W} = \boldsymbol{\Sigma}^{-1}$, tendrá en su diagonal las inversas de las varianzas de cada condición experimental, $(\sigma_{ii}^2)^{-1}$, y fuera de ella, las inversas de las covarianzas correspondientes, dadas por $[\text{cov}(\sigma_i, \sigma_j)]^{-1}$, para $i \neq j$. Queda claro que en caso de independencia entre las observaciones, se tendrá que los elementos fuera de la diagonal serán todos iguales a cero.

¹⁷*Vid.*, p. ej., MYERS *et al.* (2002).

¹⁸En el **Apéndice B** comentamos un breve esquema conceptual de estas distribuciones.

Sin embargo, existen casos en donde los modelos anteriores ya no serán útiles. Por ejemplo, cuando se tienen datos que siguen distribuciones de probabilidad del tipo discreto, todo lo anterior no podrá aplicarse con relativa tanta facilidad.

En el caso que nos ocupa, los *MDB*, éstos tendrán una tipología netamente no normal, ya que las distribuciones que siguen los mismos será binomial o de Bernoulli, como caso particular de la primera. En el capítulo 1, hemos presentado algunas formas útiles que consideraremos para proseguir con el estudio de las superficies de respuesta, cuyos fundamentos se han ampliado y comentado en el **Apéndice B**.

Habiendo entonces introducido cuál es el contexto que conduce a considerar modelos para la probabilidad de éxito como variable de interés, llegamos al momento de presentar de forma casi natural a los *MLG*, enfoque que ofrece recursos muy valiosos y consistentes para tratar estos modelos. Un caso particular de ellos, constituyen los *MDB*. Por otro lado, la consideración de las transformaciones de la respuesta como recurso para subsanar los apartamientos de las hipótesis necesarias para aplicar los métodos derivados de los mínimos cuadrados o del de máxima verosimilitud, también impone discutir las posibilidades y los alcances de uno y otro enfoque.

En los puntos siguientes, presentaremos este planteamiento de una forma un poco más minuciosa, de tal modo que nos vaya conduciendo a justificar cuál será el camino que seguiremos para la consideración de modelos para datos binarios, que es el tema que nos preocupa en este trabajo.

2.3.1. Transformación de la respuesta

La alternativa de transformar la respuesta para paliar el fenómeno de la heteroscedasticidad del modelo binomial —ya sea considerando la respuesta como número de éxitos o bien como proporciones— es una posibilidad que ha sido ampliamente estudiada en trabajos muy autorizados¹⁹. Las posibilidades que ofrecen un gran conjunto de transformaciones no sólo se utilizan para estabilizar la varianza, sino que también sirven para transformar modelos no normales en normales, simplificar los cálculos, etc. No sólo interesa con las transformaciones estabilizar la varianza, sino que también se persigue conseguir normalidad en las observaciones para distintos niveles de \mathbf{x} y también intentar encontrar funciones $g(y)$ tales que $E[g(y) | \mathbf{x}]$ resulte lo más lineal

¹⁹Por citar algunas referencias representativas, mencionaremos los trabajos de BARTLETT (1947); ANSCOMBE (1948); FREEMAN Y TUKEY (1950) y BOX Y COX (1964), entre muchas otras. Existe también un par de referencias dedicadas al estudio de este problema, que son ATKINSON (1987) y CARROLL Y RUPERT (1988), que atacan con profundidad las implicancias y condiciones para el uso de transformaciones de la respuesta. Asimismo, también en los últimos diez años han aparecido publicaciones con aplicaciones concretas de transformaciones de otros modelos no normales, como por ejemplo BISGAARD Y FULLER (1994 y 1995). De igual modo, en LEWIS (1998) y especialmente en MONTGOMERY *et al.* (2001), AGRESTI (2002) y en DRAPER Y SMITH (1998), aparecen discusiones muy provechosas sobre el uso de las transformaciones de la respuesta.

posible. La nota más importante de este aspecto del uso de las transformaciones es que se pretendía lograr *todas las condiciones a la vez* mediante una *única transformación*.

A este respecto, una primera crítica —y relativamente “grave”— que suele hacerse frecuentemente²⁰ en el uso de este recurso en general es que no existe seguridad que *una única transformación* pueda ser capaz de proporcionar todos los objetivos mencionados en el párrafo anterior. De acuerdo con MONTGOMERY (2001), el uso de las transformaciones presenta algunos inconvenientes, que pueden resumirse en los siguientes puntos:

- El uso de escalas transformadas de medición de la respuesta —como es el caso, por ejemplo, de la transformación logarítmica— suele llevar a interpretaciones que pueden resultar poco claras, sobre todo cuando las respuestas miden magnitudes físicas²¹.
- En ocasiones como en el caso de conteos, pueden aparecer problemas en cuanto a la definición de regiones de experimentación en el espacio de los factores. Por ejemplo, puede ocurrir que utilizando la transformación de la raíz cuadrada, haya ciertas regiones en donde la misma resulte negativa, con lo cual, entra en aparente contradicción con la función matemática utilizada.

El empleo de transformaciones podría bien verse como un recurso que lo hace en cierto modo “forzado”, debido a que intrínsecamente se trata de convertir un modelo con naturaleza propia hacia otro en el cual se cuenta con herramientas de estudio y análisis que ya gozan de un grado autorizado de madurez, como por ejemplo, los alcances del modelo lineal normal. Podría parecer, por lo tanto, que el precio que se paga por convertir un modelo intrínsecamente heteroscedástico en otro homoscedástico sea el de perder facilidad en la interpretación de los resultados, pese a que por otro lado se gane en facilidad de cálculo.

Yendo un poco más allá de la inconstancia de varianza, otras propiedades estadísticas deseables suelen exigírseles también a las transformaciones, como lo son la distribución normal de las observaciones $y_i \mid \mathbf{x}_i$ y la linealidad del valor esperado de la respuesta, es decir, que ésta resulte como combinación lineal de efectos. WU Y HAMADA (2000) señalan que para ciertas distribuciones (p. ej.: la distribución de Poisson), no pueden lograrse mantener las tres propiedades deseables precitadas todas

²⁰ Vid. p. ej. MONTGOMERY (2001).

²¹ *Ídem* anterior. El autor cita ejemplos que ofrecen cierta dificultad en la interpretación al análisis, como por ejemplo de la *raíz cuadrada del número de elementos defectuosos*; al *logaritmo de la resistividad*, etc.

a la vez. En MCCULLOCH Y SEARLE (2001), p. 78, los autores hacen una breve pero interesante discusión entre transformar y directamente aplicar una función que enlace el valor esperado de la respuesta con un predictor lineal —la función “link”— como ventaja de utilizar los *MLG* en lugar de modificar los modelos. Otros autores, como WU Y HAMADA (2000) señalan que las dificultades expuestas en el punto anterior desaparecen cuando se utiliza el enfoque de los *MLG*, dando importantes puntos de reflexión y destacando que el uso de transformaciones no será necesario de considerar, ya que los fundamentos de los *MLG* tienen en cuenta de forma natural los efectos que se le piden a las transformaciones de la respuesta.

Tanto desde un punto de vista de desarrollo teórico como de disponibilidad de software para efectuar sus cálculos, los *MLG* constituyen un punto de vista muy eficiente para el análisis de datos provenientes de un grupo interesante de distribuciones, llamado *familia exponencial*, de la cual, la binomial forma parte.

2.3.2. Los *MLG* como alternativa a transformar

Frente a los inconvenientes que pueden traer la consideración de transformaciones de la respuesta, una alternativa válida y más “natural” parece ser el uso de los *MLG*, que ofrece un enfoque unificado para modelos tanto normales como no normales. La esencia que tiene el uso de estos modelos radica en un conjunto de aspectos metodológicos que describimos en el **Apéndice B**, aunque resumimos aquí los principales²²:

- a. Los datos no siguen necesariamente una distribución normal,
- b. El valor esperado no es considerado necesariamente como una combinación lineal de parámetros, aunque sí lo es una determinada transformación no lineal del mismo. En otras palabras, los *MLG*, en cambio, suponen una función determinada del *valor esperado*, llamada función “link”.
- c. La varianza quedará expresada mediante una cierta función $\Upsilon(\mu)$, cuya forma dependerá de la distribución que se trate (ver expresión 2.13).

Con el ánimo puesto en aprovechar la amplia autonomía metodológica que ofrecen los *MLG*, como así también motivados en que la cantidad de trabajos dedicados a sus aplicaciones al diseño de experimentos es relativamente pequeña, es que preferimos utilizar el enfoque que proporcionan los *MLG* frente al de las transformaciones de la respuesta, el cual utilizaremos en lo que sigue de esta tesis. En las últimas publicaciones relacionadas con este tópico parece haber una cierta tendencia —aunque, ciertamente,

²² *Vid.*, p. ej., MCCULLOCH Y SEARLE (2001).

no muy clara del todo— a preferir el uso de los *MLG* frente al de las transformaciones en el estudio de aplicaciones industriales²³. Ello no obstante, hay algunos modelos y condiciones especiales de los mismos en donde no resulta útil este enfoque, como por ejemplo, cuando la varianza de las componentes aleatorias de una distribución normal depende de uno o más factores mediante funciones de la forma $\sigma^2 = c^2(a + bx)^{2d}$, siendo a , b , c y d valores escalares²⁴.

2.3.3. Características fundamentales de los *MLG*

Siguiendo el mismo esquema que para los modelos anteriores, considerando $i = 1, \dots, n$ niveles de los factores y $j = 1, \dots, k$ factores, podemos destacar los siguientes aspectos característicos de los *MLG*:

a. Componente sistemática:

$$g[E(\mathbf{y} \mid \mathbf{x})] = \beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}$$

b. Matriz de derivadas, \mathbf{F} :

$$f_{ij} = \frac{\partial}{\partial \beta_j} \{g[E(\mathbf{y} \mid \mathbf{x})]\} = \frac{\partial}{\partial \beta_j} (\beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x})$$

c. Estimador máximo verosímil del vector de parámetros, $\hat{\boldsymbol{\beta}}_{ML}$:

$$\hat{\boldsymbol{\beta}}_{ML} \underset{\text{asint.}}{\sim} N \left[\boldsymbol{\beta}, [a(\phi)]^2 (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \right]$$

d. Matriz asintótica de varianzas y covarianzas de $\hat{\boldsymbol{\beta}}_{ML}$, $\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML})$:

$$\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML}) = [a(\phi)]^2 (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$$

e. Matriz de Información de Fisher, $\mathbf{I}(\hat{\boldsymbol{\beta}}_{ML})$:

$$\mathbf{I}(\hat{\boldsymbol{\beta}}_{ML}) = \frac{\mathbf{X}'\mathbf{W}\mathbf{X}}{[a(\phi)]^2}$$

Resulta bien conocido que los estimadores máximo-verosímiles del vector de parámetros tendrán propiedades del tipo asintóticas. En efecto, sea $\hat{\boldsymbol{\beta}}$ el valor final resultante de aplicar el algoritmo iterativo para la estimación del vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$,

²³Entre los más recientes: GREGO (1993); KHURI (1993 y 2001); MYERS *et al.* (1994); HAMADA Y NELDER (1997); LEWIS (1998); AGRESTI (2002) y LEWIS *et al.* (1999, 2001A y 2001B), entre otros.

²⁴*Vid.* p. ej., <http://www.zoo.ufl.edu/bolker/emd-2000/glm.html>, del Prof. Ben Bolker (University of Florida), en donde el autor hace una breve discusión de los alcances de los *MLG* en condiciones especiales.

para el caso particular del modelo logístico, y si las suposiciones del modelo han sido las correctas —incluyendo la elección de la función “link”— se demuestra²⁵ asintóticamente, que $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, es decir, será un estimador asintóticamente insesgado.

Puesto que en el caso más general no se cumple la hipótesis de varianza constante de las observaciones, se considera aquí también una matriz $\boldsymbol{\Sigma}^{-1}$ de varianzas y covarianzas de las observaciones. Ya que además las observaciones resultan independientes entre sí, se podrá escribir:

$$\boldsymbol{\Sigma}^{-1} = \mathbf{W} = \text{diag} \{ [V(y_i | \mathbf{x}_i)]^{-1} \} \quad (2.12)$$

en donde $V(y_i | \mathbf{x}_i)$ refiere a la varianza de la i -ésima observación binomial independiente, la que a su vez, depende del valor que tome su valor esperado $E(y_i | \mathbf{x}_i)$. Podemos expresar esto indicando que la varianza estará conectada con el valor esperado mediante una cierta función $\Upsilon(\cdot)$, que depende de la distribución de la que se trate, de modo que:

$$V(y_i | \mathbf{x}_i) = \Upsilon[E(y_i | \mathbf{x}_i)]. \quad (2.13)$$

En el caso que abordamos, la distribución de Bernoulli, se considera una función del tipo:

$$\Upsilon(\pi) = \pi(1 - \pi),$$

con lo cual, la función que vincula el valor esperado de la respuesta con la varianza será:

$$V(y_i | \mathbf{x}_i) = E(y_i | \mathbf{x}_i) \cdot [1 - E(y_i | \mathbf{x}_i)] = \pi(\mathbf{x}_i, \boldsymbol{\beta}) \cdot [1 - \pi(\mathbf{x}_i, \boldsymbol{\beta})]$$

2.3.4. Los *MLG* y el diseño de experimentos

En esta sección no abordaremos los detalles particulares sobre esta clase de modelos, puesto que existen excelentes referencias dedicadas al estudio de los *MLG*²⁶. Antes bien, haremos referencia a algunas propiedades generales de los mismos, que utilizaremos puntualmente para los datos binarios.

En casi toda la bibliografía que aborda el estudio de los *MLG* y sus aplicaciones siempre se hace un correcto hincapié en la definición del predictor lineal en función de

²⁵ Vid. p. ej. KRZANOWSKI (1998).

²⁶ Dentro de las obras dedicadas a esta clase de modelos, las siguientes se han ganado la mayor aceptación en los últimos tiempos: NELDER Y WEDDERBURN (1972); MCCULLAGH Y NELDER (1989); FIRTH (1991); MYERS *et al.* (2002); DOBSON (2002) y MYERS *et al.* (2004). Igual consideración tienen las obras de: LINDSEY (1997); MCCULLOCH Y SEARLE (2001), FAHRMEIR Y TUTZ (2001), entre otros. Asimismo, MYERS y MONTGOMERY (1997) y MCCULLOCH (1990) constituyen una muy clara introducción al estudio de estos modelos, entre muchos otros artículos, como así también las obras de OLSSON (2002) y HARDIN Y HILBE (2001), entre otros.

un conjunto de variables regresoras y de otro de parámetros, lo cual queda expresado en:

$$g[E(\mathbf{y} \mid \mathbf{x})] = \beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}. \quad (2.14)$$

Sin embargo, no queda siempre del todo claro si los distintos niveles que toman las \mathbf{x} provienen de estudios observacionales o de experimentos diseñados.

Una primera nota que podemos indicar para el presente trabajo es que daremos al predictor lineal una naturaleza netamente “experimental”. Con esto, pretendemos dejar claro que el experimentador posee el control de los niveles que tomarán los factores de variabilidad del proceso, cuyos niveles variará de acuerdo a lo que el buen sentido y la metodología adecuada le vayan informando. Así, excluimos de los alcances de nuestro trabajo el del estudio de los estudios observacionales, en los que los factores han sido asignados sin una planificación adecuada. Por lo tanto, podemos indicar que este enfoque de integrar los *MLG* con el diseño de experimentos es nuestro punto de partida para intentar avanzar en el conocimiento de estos problemas.

Para comenzar a enlazar los *MLG* con el diseño de experimentos, en el sentido en que las observaciones de las variables respuesta se efectuarán para distintas condiciones experimentales —dadas cada una de ellas por el nivel que tome un cierto vector de factores $\mathbf{x} = (x_1, \dots, x_k)'$ — nos referiremos a la i -ésima observación de la respuesta mediante la notación:

$$y_i \mid \mathbf{x}_i,$$

siendo \mathbf{x}_i un valor particular que toma \mathbf{x} , representativo de una de las n posibles condiciones experimentales que se consideren para el experimento.

En el caso que nos ocupa, que es la relación entre la probabilidad de éxito π considerada como respuesta y un grupo de factores de variabilidad dado por $\mathbf{x} = (x_1, \dots, x_k)'$, interesará utilizar modelos capaces de describir cómo cambia π con los cambios ocurridos en \mathbf{x} . Tomando, por ejemplo, el caso del modelo lineal completo de segundo orden, y siguiendo la notación que describen las ecuaciones (1.3), tendremos que el predictor lineal será de la forma de la ecuación:

$$g[\pi(\mathbf{x}, \boldsymbol{\beta})] = \beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}, \quad (2.15)$$

en donde:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} \\ \frac{1}{2}\beta_{12} & \beta_{22} \end{bmatrix}$$

Considerando entonces el modelo logístico —y el link logit— la probabilidad de éxito quedará definida para cada una de las condiciones experimentales como:

$$\pi(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \mathbf{x}'_i\mathbf{b} + \mathbf{x}'_i\mathbf{B}\mathbf{x}_i)}{1 + \exp(\beta_0 + \mathbf{x}'_i\mathbf{b} + \mathbf{x}'_i\mathbf{B}\mathbf{x}_i)}. \quad (2.16)$$

Por tanto, si se define de forma particular $g[E(y_i | \mathbf{x}_i)]$, es decir, $g[\pi(\mathbf{x}_i, \boldsymbol{\beta})]$, mediante el link logit²⁷:

$$g[\pi(\mathbf{x}_i, \boldsymbol{\beta})] = \ln \left[\frac{\pi(\mathbf{x}_i, \boldsymbol{\beta})}{1 - \pi(\mathbf{x}_i, \boldsymbol{\beta})} \right] = \ln \left[\frac{\frac{\exp(\eta_i)}{1 + \exp(\eta_i)}}{1 - \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}} \right]$$

y si se desarrolla algebraicamente el logaritmo²⁸, se llegará a que:

$$\text{logit}[\pi(\mathbf{x}_i, \boldsymbol{\beta})] = \beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}, \quad (2.17)$$

que permite linealizar la expresión (2.16), la que a su vez no tiene impuestas restricciones en cuanto a los valores que puede tomar la función $\pi(\mathbf{x}, \boldsymbol{\beta})$. Esto permite aprovechar las bondades del modelo lineal.

Los *MLG* proporcionan un enfoque unificado para modelar una variedad relativamente importante de distribuciones de probabilidad, tanto de datos discretos como continuos²⁹. Aquí hemos presentado algunos aspectos básicos para los modelos binarios, que se ampliarán en el **Apéndice B**, dentro del contexto de los *MLG*.

²⁷La utilización de otros “links” para este modelo y una discusión sobre su grado de adecuación puede verse por ejemplo en MCCULLAGH Y NELDER (1989) o en COLLETT (2003). En el **Apéndice B** diremos unas palabras más a este respecto.

²⁸Vid. p. ej. KLEINBAUM (1994).

²⁹Vid. p. ej. OLSSON (2002), p. 36.