



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS MATEMÁTICAS

FACULTAD DE CIENCIAS

MODELOS LINEALES MIXTOS
GENERALIZADOS.

TESIS

QUE PARA OBTENER EL GRADO ACADÉMICO DE
MAESTRA EN CIENCIAS MATEMÁTICAS

PRESENTA

MARISELA GARCÍA PACHECO.

DIRECTORA DE TESIS: DRA. SILVIA RUIZ VELASCO ACOSTA.

MÉXICO, D.F.

SEPTIEMBRE, 2006.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**MODELOS LINEALES MIXTOS
GENERALIZADOS.**

Agradezco

A mi hijo: Luis Gerardo por la motivación para la realización de esta tesis.

A mis padres: Fernando García Y Emma Yolanda Pacheco y hermanos Josué Fernando, Emma Yolanda y Efrén Antonio. Por el apoyo incondicional en cualquier etapa de mi vida.

A mis sobrinos: Oscar y María Fernanda por los momentos de felicidad que me regalan.

A mis profesores por la paciencia y el apoyo que siempre recibí de ellos.

A mis amigos y compañeros por los grandes momentos de apoyo en mi estudio y en mi vida.

A Dios por todo lo que me regala día a día.

Índice general

1. LM	1
1.1. MODELOS LINEALES CLÁSICOS.	1
1.2. GLM	2
1.2.1. Modelos para datos continuos con varianza constante.	5
1.2.2. GLM para variables con respuesta binaria.	6
1.2.3. Modelo Poisson y Loglineal.	10
1.2.4. Ejemplos de Modelos probabilísticos para tablas de contingencia.	13
1.2.5. Formas de estimación.	15
1.3. LMM	20
1.3.1. Definición del modelo.	21
1.3.2. Ejemplos.	24
1.3.3. Formas de estimación.	28
1.3.4. Valores fijos y predictivos.	41
1.4. NLME	43
1.4.1. NLME	43
1.4.2. Estimación por verosimilitud.	45

2. GLMM	47
2.1. LA ESTRUCTURA DEL MODELO.	48
2.1.1. Consecuencias de tener efectos aleatorios.	49
2.2. MÉTODOS DE ESTIMACIÓN.	53
2.2.1. Estimación por máxima verosimilitud.	53
2.2.2. PQL	55
2.2.3. Estimación PQL	58
2.2.4. Quasi-verosimilitud Marginal.	60
2.3. MODELO MARGINAL CONTRA MODELO CONDICIONAL.	62
2.4. MÉTODOS DE BONDAD DE AJUSTE PARA GLMM. . . .	65
2.4.1. Desarrollo de la técnica.	67
3. APLICACIÓN.	71
3.1. MODELOS AJUSTADOS.	74
3.2. CONCLUSIÓN.	92
4. CONCLUSIÓN GENERAL.	95
A. RESULTADOS DE STATA.	97

Índice de figuras

1.1. Modelos GLM con respuesta binaria	11
3.1. Crecimiento de niños por edad y sexo	93

Índice de Tablas

1.1. Ejemplos de distribuciones con liga canónica.	5
3.1. Tabla de estadísticos de resumen para las covariables utilizadas en los modelos.	73
3.2. Tabla de percentiles y valores mínimos.	74
3.3. Modelos con una sola predicción para el efecto fijo.	75
3.4. Varianzas–covarianzas de los modelos con una sola predicción para el efecto fijo.	77
3.5. Modelo $\text{logit}(E(\text{crecio} t)) = \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t$. 78	
3.6. Modelo $\text{logit}(E(\text{crecio} t)) = \beta_{\text{hct36}} * \text{hct36} + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t$	79
3.7. Modelo $\text{logit}(E(\text{crecio} t)) = \beta_{\text{hct20}} * \text{hct20} + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t$	79
3.8. Modelo $\text{logit}(E(\text{crecio} t)) = \beta_{\text{hbg36}} * \text{hbg36} + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t$	80
3.9. Modelo $\text{logit}(E(\text{crecio} t)) = \beta_{\text{hct20}} * \text{hct20} + \beta_{\text{hbg36}} * \text{hbg36} + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t$	81
3.10. Modelo $\text{logit}(E(\text{crecio} t)) = \beta_{\text{lnlead18_36}} * \text{lnlead18_36} + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t$	82

- 3.11. Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{lead6}_36} * \text{lead6}_36 + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t. \dots \dots \dots 83$
- 3.12. Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{lnm28}_36} * \text{lnm28}_36 + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t. \dots \dots \dots 84$
- 3.13. Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{hct20}} * \text{hct20} + \beta_{\text{lnm28}_36} * \text{lnm28}_36 + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t. \dots \dots \dots 84$
- 3.14. Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{hgb36}} * \text{hgb36} + \beta_{\text{lnm28}_36} * \text{lnm28}_36 + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t. \dots \dots \dots 85$
- 3.15. Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{hct20}} * \text{hct20} + \beta_{\text{hgb36}} * \text{hgb36} + \beta_{\text{lnm28}_36} * \text{lnm28}_36 + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t. 86$
- 3.16. Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{lnlead}} * \text{lnlead} + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t. \dots \dots \dots 87$
- 3.17. Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{hct20}} * \text{hct20} + \beta_{\text{lnlead}} * \text{lnlead} + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t. \dots \dots \dots 88$
- 3.18. Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{hct20}} * \text{hct20} + \beta_{\text{hgb36}} * \text{hgb36} + \beta_{\text{lnlead}} * \text{lnlead} + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t. \dots \dots 89$
- 3.19. Criterio de Información de Akaike (AIC) y Bayesiano (BIC). . . 91

INTRODUCCIÓN.

Los métodos de regresión pretenden explicar relaciones entre una variable dependiente y una o más explicativas o independientes, esto ha permitido estudiar el comportamiento de fenómenos de tipo social, cultural, económico, natural, etc. En la gama que hay en la teoría de regresión se han tenido algunas limitantes para poder realizar mejores análisis, como puede ser que los modelos no solo expliquen modelos con distribución normal o que dos características o comportamientos que a simple vista son diferentes de manejar puedan ser distinguibles dentro de las variables que se involucran en el modelo que explica un determinado fenómeno, entre otras situaciones. Por ello, se ha buscado dar respuesta a estas situaciones y entre las opciones con las que se cuenta están los modelos lineales generalizados mixtos (GLMM).

Como los modelos de regresión lineal no permiten tener en cuenta los efectos que pueden provocar el modo en que un individuo se comporta dentro de distintas situaciones como pueden ser escuelas, hospitales, países, etc. (en fenómenos sociales o culturales), o enfermedades, climas en fenómenos naturales, estos modelos de regresión simples pueden representar de forma incompleta o engañosa la explicación de los resultados. Así mismo un modelo de regresión lineal que no tenga en cuenta una estructura jerárquica de los datos, no diferenciará entre estos sistemas. El uso de modelos mixtos, puede reconocer la jerarquía anidada de forma natural en el fenómeno a estudiar,

permitiendo una variación relativa en los resultados medidos. Estas jerarquías se modelan como variables explicativas con un comportamiento aleatorio, de tal forma que permita distinguir los diferentes comportamientos. Gráficamente, se puede ver que en un modelo de regresión simple, la recta estimada esta dada de manera única, mientras que, en los modelos mixtos la recta estimada está determinada por la aleatoriedad de la ordenada al origen y la pendiente asociadas, es decir, ambos valores se pueden convertir en variables aleatorias que se estiman a partir de los datos.

Por otro lado, los modelos lineales generalizados (GLM) permiten relacionar variables no solo de tipo normal, sino también variables que pertenecen a la familia exponencial, como son distribuciones de tipo binomial, multinomial, poisson o más generalmente variables cuya varianza es función de la media. Esto permite dar otra interpretación a los modelos como se verá en alguna parte de la tesis.

De este modo, se crean los modelos GLMM, los cuales no solo permiten asociar o relacionar una distribución normal, sino variables con distribuciones pertenecientes a la familia exponencial o cuya varianza es función de la media, las variables pueden ser continuas o discretas y; al mismo tiempo con algún comportamiento aleatorio, para dar una mejor explicación de manera individual a cada comportamiento dentro de la variable dependiente.

RESUMEN.

El trabajo es una cronología acerca del surgimiento de los modelos GLMM. Se aborda de manera muy superficial el tema de regresión lineal y diseño de experimentos, los cuales son la base de todos los modelos que se explican después.

Se explican y se desarrollan ejemplos de los modelos (GLM), que surgen por la necesidad de expandir un modelo lineal cuando el valor esperado ya no es una función lineal de las covariables, pero si lo es alguna función del valor esperado. Estos modelos se aplican a distribuciones que pertenecen a la familia exponencial o de manera general a variables aleatorias cuya varianza es función de la media.

Por otro lado, se tienen los modelos lineales mixtos (LMM), en donde se tienen dos efectos: uno llamado fijo, que es el mismo que se usa en los modelos lineales y, otro llamado aleatorio, este tiene la propiedad de dar a cada individuo un comportamiento distinto. Aquí, el valor esperado es condicional, es decir, se considera como la esperanza de la variable respuesta dados los efectos aleatorios. Estos modelos surgen a partir de datos longitudinales principalmente.

También se da una breve descripción de los modelos de efectos mixtos no lineales (NLEM), estos modelos generalizan el hecho de tener una función real, diferenciable del parámetro y las covariables, donde el parámetro es una

función lineal de las covariables que pueden ser de efectos fijos y aleatorios. Finalmente, se desarrolla a los modelos lineales mixtos generalizados (GLMM). Estos modelos constan de efectos fijos y aleatorios, por lo que la función ligada en GLM se complica, pues la media es condicional dados los efectos aleatorios, esto ya no garantiza una media marginal de forma natural, por lo que los valores de la variable respuesta no pueden ser interpretados marginalmente. También la función de verosimilitud se ve afectada por los efectos aleatorios, lo cual dificulta la maximización, para subsanar este hecho se recurre a la estimación quasi-verosimil, lo que significa ignorar la distribución de los datos, garantizando al mismo tiempo que se cumplan las condiciones de la función de verosimilitud.

Debido a la complejidad que tiene el modelo dados los efectos aleatorios, es importante comparar o tener información auxiliar para hablar del modelo marginal, pues no siempre es de interés hablar de un individuo en particular, algunas veces se desea hablar de los valores promedios.

Una aplicación proveniente de una cohorte de niños a los cuales se les tomaron medidas antropométricas y del nivel de plomo (en escala logaritmica) con periodos de edad entre 42 y 120 meses, para indagar si el plomo afecta al crecimiento del niño.

Ajustando un modelo logit longitudinal con el individuo y el nivel de plomo en los últimos seis meses como efectos aleatorios, se ve que los niños con alto nivel de plomo tienen un crecimiento menor que los que reportan bajo nivel de plomo en la sangre.

Capítulo 1

INTRODUCCIÓN A LOS MODELOS LINEALES (LM).

1.1. MODELOS LINEALES CLÁSICOS.

Los elementos de un modelo lineal clásico son: el conjunto de observaciones denotado por un vector columna $\mathbf{y} = [y_1, \dots, y_n]^T$, las cuales representan una variable aleatoria; el conjunto de covariables o variables explicativas es una matriz \mathbf{X} de dimensión $n \times p$, donde cada renglón de \mathbf{X} se asocia a una observación y, cada columna a una covariable diferente; a cada covariable se la asocia un coeficiente o parámetro, desconocido; el conjunto de parámetros es un vector de dimensión p , generalmente denotado por $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^T$. De tal forma que la relación del vector de observaciones con las covariables es:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{donde } E(\boldsymbol{\epsilon}) = 0, \text{ y por lo tanto } E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}.$$

donde $\boldsymbol{\epsilon}$ es el vector de errores aleatorio definido para cada valor de \mathbf{y} .

En 1805 Legendre propuso estimar las β 's minimizando $\epsilon^T \epsilon = \sum \epsilon_i^2 = \sum (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^2$ sobre los valores de $\boldsymbol{\beta}$, a este método se le conoce como mínimos cuadrados. En 1809, Gauss introduce la distribución Normal con media cero y varianza constante para los errores, y mostró que los estimadores de $\boldsymbol{\beta}$ que se obtienen por el criterio de mínimos cuadrados son los que tienen varianza mínima dentro de la clase de estimadores insesgados, y corresponden a los estimadores máximo-verosímiles. Al suponer la distribución normal en los errores estamos suponiendo que la variable aleatoria se distribuye $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$.

Más adelante, analizando datos astronómicos se dió origen a la teoría de diseño de experimentos, la cual se relaciona principalmente con R. A. Fisher y sus colaboradores. En estos modelos se utiliza un diseño factorial que hace crecer la información por observación, pues se estudian varios factores de clasificación y varios experimentos simultáneamente a través del tiempo. Este modelo considera a la matriz \mathbf{X} formada por ceros y unos llamada matriz de incidencia, a diferencia del modelo de regresión que considera covariables continuas.

Las ideas de Fisher sobre el modelo factorial dan pauta a los modelos lineales generalizados, que incluyen, entre otros, al análisis de cierto tipo de conteo o de proporciones.

1.2. MODELOS LINEALES GENERALIZADOS (GLM).

Estos modelos son una generalización del modelo lineal clásico, donde el vector \mathbf{y} no necesariamente tiene distribución normal, sino que su distribución pertenece a la familia exponencial. Que tiene la forma:

$$f_Y(y; \theta, \phi) = \exp\{(y\theta - d(\theta))/a(\phi) + c(y, \phi)\}$$

para algunas funciones específicas a , d , c . Si ϕ es conocido, entonces f_Y tiene parámetro canónico θ .¹

Un vector de observaciones \mathbf{y} con n componentes se supone como una realización de una variable aleatoria \mathbf{Y} cuyas componentes son distribuidas independientemente con media $\boldsymbol{\mu}$. La *parte sistemática* del modelo toma una forma para el vector $\boldsymbol{\mu}$ en términos de un número $p < n$ de parámetros desconocidos β_1, \dots, β_p , de la siguiente forma:

$$\eta_i = g(\mu_i) = \sum_{j=1}^p x_{ij}\beta_j, \quad \text{con } \mu_i = E(Y_i), \quad i = 1, \dots, n$$

donde las β 's son los parámetros a ser estimados, x_{ij} es el valor de la j -ésima covariable (variable explicativa), para la observación i . A la función g se le llama *función liga* y es una función monótona diferenciable. En notación matricial, tenemos que:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

donde $\boldsymbol{\eta}$ es de $n \times 1$, \mathbf{X} es de $n \times p$ y $\boldsymbol{\beta}$ es de $p \times 1$. A la matriz \mathbf{X} se le denomina la matriz diseño, $\boldsymbol{\beta}$ es el vector de parámetros, y $\boldsymbol{\eta}$ es el predictor lineal,

Sea $\ell(\theta, \phi; y) = \log f_Y(y; \theta, \phi)$ la función log-verosímil considerada como función de θ y de ϕ para una y dada. La media y la varianza de \mathbf{y} pueden ser derivadas como sigue.

¹Para más información acerca de esta definición se puede ver la referencia [1].

Sabemos que:²

$$E\left(\frac{\partial \ell}{\partial \theta}\right) = 0 \quad y \quad E\left(\frac{\partial^2 \ell}{\partial \theta^2}\right) + E^2\left(\frac{\partial \ell}{\partial \theta}\right) = 0. \quad (1.1)$$

Así, tenemos que:

$$\ell(\theta; y) = \frac{y\theta - d(\theta)}{a(\phi)} + c(y, \phi) \quad \Rightarrow \quad \frac{\partial \ell}{\partial \theta} = \frac{y - d'(\theta)}{a(\phi)} \quad y \quad \frac{\partial^2 \ell}{\partial \theta^2} = -\frac{d''(\theta)}{a(\phi)}$$

Entonces por las ecuaciones (1.1) se tiene que:

$$E\left(\frac{\partial \ell}{\partial \theta}\right) = \frac{y - d'(\theta)}{a(\phi)} = \frac{E(y) - d'(\theta)}{a(\phi)} = 0 \quad \Rightarrow \quad E(y) = d'(\theta)$$

$$E\left(\frac{\partial^2 \ell}{\partial \theta^2}\right) + E^2\left(\frac{\partial \ell}{\partial \theta}\right) = E\left(-\frac{d''(\theta)}{a(\phi)}\right) + E^2\left(\frac{y - d'(\theta)}{a(\phi)}\right) = -\frac{d''(\theta)}{a(\phi)} + \frac{E(y - d'(\theta))^2}{a^2(\phi)} =$$

$$-\frac{d''(\theta)}{a(\phi)} + \frac{Var(y)}{a^2(\phi)} = 0 \quad \Rightarrow \quad Var(y) = a(\phi)d''(\theta)$$

Por lo tanto, $\mu = d'(\theta)$ y $Var(y) = a(\phi)d''(\theta) = a(\phi)\mu'(\theta)$

Cuando el parámetro canónico es igual al predictor lineal, la función ligo correspondiente se conoce como *liga canónica* y, en este caso se tiene una estadística suficiente de igual dimensión que β en el predictor lineal $\eta = \sum x_{ij}\beta_j$, es decir, si

$$\theta = \eta \quad \Rightarrow \quad X^T y$$

es una estadística suficiente para β , pues

$$f_Y(y; \theta, \phi) = \exp\{(y\theta - d(\theta))/a(\phi) + c(y, \phi)\} =$$

²Ver referencia [3]

$$\begin{aligned} & \exp\{(y\theta/a(\phi))\} \exp\{a(\phi)/d(\theta) + c(y, \phi)\} = \\ & \exp\{(yX\beta/a(\phi))\} \exp\{a(\phi)/d(X\beta) + c(y, \phi)\} = \\ & \exp\left\{\left(\sum y_i x_i\right)\beta/a(\phi)\right\} \exp\{a(\phi)/d(X\beta) + c(y, \phi)\} \end{aligned}$$

por lo tanto $X^T y$ es una estadística suficiente para β .³

Distribución	$\mu(\theta) = E(Y; \theta)$	Liga canónica: $\theta(\mu)$
Normal	θ	identidad: μ
Poisson	$\exp(\theta)$	$\log \mu$
Binomial	$\frac{e^\theta}{1+e^\theta}$	logit: $\log\left(\frac{\mu}{1-\mu}\right)$
Gamma	$-1/\theta$	recíproco: μ^{-1}
Gaussiana Inversa	$(-2\theta)^{-1/2}$	μ^{-2}

Tabla 1.1: Ejemplos de distribuciones con liga canónica.

1.2.1. Modelos para datos continuos con varianza constante.

Un modelo lineal se puede escribir como:

$$Y_i \sim N(\mu_i, \sigma^2) \quad \mu_i = \eta_i \quad \eta_i = \sum_{j=1}^p x_{ij}\beta_j$$

donde $Y_i \sim N(\mu_i, \sigma^2)$ son observaciones independientes idénticamente distribuidas, $\mu_i = \eta_i$ corresponde a la liga identidad, $\eta_i = \sum_{j=1}^p x_{ij}\beta_j$ es el

³La tabla (1.1) se puede consultar en la referencia [1] de la bibliografía.

predictor lineal basado en las covariables, y los vectores $\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\eta}$ tienen n componentes, $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, donde $\boldsymbol{\beta}$ es un vector de parámetros.

Este modelo, sólo tiene una componente para el error, el vector de errores $\boldsymbol{\epsilon}$ en donde todas las componentes son variables aleatorias independientes con distribución normal con media cero y varianza σ^2 .

Los modelos de regresión consideran covariables continuas. Cuando las variables son categóricas los datos se estudian a partir de modelos de Análisis de varianza (ANOVA) que son utilizados principalmente en diseños de experimentos y generalmente se caracterizan porque la matriz diseño no es de rango completo. La mezcla de covariables continuas y discretas se conoce como Análisis de covarianza y pueden ser cruzados o anidados, entre otros.

1.2.2. GLM para variables con respuesta binaria.

En los modelos lineales generalizados donde la respuesta es binaria, la respuesta puede ser éxito o fracaso, es decir, $z = 1$ si el resultado es éxito y $z = 0$ si el resultado es fracaso, con $P(z = 1) = p$ y $P(z = 0) = 1 - p$. Para un patrón de covariables hay n variables aleatorias independientes Z_1, \dots, Z_n , definimos $Y = \sum_{j=1}^n Z_j$, tal que Y es el número de éxitos en los n ensayos, entonces Y tiene distribución $Bin(n, p)$, tal que:

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y} \quad \text{donde} \quad y = 0, 1, \dots, n.$$

Esta distribución pertenece a la familia exponencial, que se representa como

$$f(y; \theta, \phi) = \exp \left\{ y \log \left(\frac{p}{1-p} \right) + n \log(1-p) + \log \binom{n}{y} \right\}.$$

Finalmente, sean, Y_1, Y_2, \dots, Y_N v.a.i.i.d., con $Y_i \sim Bin(n_i, p_i)$, el logaritmo de la función de verosimilitud tiene la forma:

$$\ell(p_1, \dots, p_n; y_1, \dots, y_n) = \sum_{i=1}^N y_i \log\left(\frac{p_i}{1-p_i}\right) + n_i \log(1-p_i) + \log\binom{n_i}{y_i}.$$

En donde se define $P_i = \frac{Y_i}{n_i}$ como la proporción de éxitos en cada subgrupo i . Además, como $E(Y_i) = n_i p_i$, entonces $E(P_i) = p_i$, y modelando las probabilidades p_i como:

$$g(p_i) = \mathbf{x}_i^T \boldsymbol{\beta},$$

donde \mathbf{x}_i es un vector de variables explicativas, $\boldsymbol{\beta}$ es un vector de parámetros y g es la función liga.

El caso más simple es el modelo lineal

$$p = \mathbf{x}^T \boldsymbol{\beta},$$

el cual es usado en algunas aplicaciones prácticas, pero con la desventaja que aunque p es una probabilidad, el valor fijo de $\mathbf{x}^T \boldsymbol{\beta}$ puede tener valores en todos los reales. Para asegurar que p está en el intervalo $[0, 1]$, se construye una función de distribución acumulada como sigue:

$$p = \int_{-\infty}^t f(s) ds \quad \text{donde} \quad f(s) \geq 0 \quad \text{y} \quad \int_{-\infty}^{\infty} f(s) ds = 1.$$

La función de densidad $f(s)$ es llamada distribución de tolerancia.

Si la distribución de tolerancia $f(s)$ es la distribución uniforme sobre el intervalo $[c_1, c_2]$

$$f(s) = \begin{cases} \frac{1}{c_2 - c_1} & \text{si } c_1 \leq s \leq c_2 \\ 0 & \text{en otro caso} \end{cases}$$

entonces

$$p = \int_{c_1}^x f(s) ds = \frac{x - c_1}{c_2 - c_1} \quad \text{para } c_1 \leq x \leq c_2.$$

Esta ecuación tiene la forma $p = \beta_1 + \beta_2 x$, donde

$$\beta_1 = -\frac{c_1}{c_2 - c_1}, \quad \beta_2 = \frac{1}{c_2 - c_1}.$$

Este modelo lineal considera a la función ligo g como la identidad y condiciona a x, β_1, β_2 tal que $c_1 \leq x \leq c_2$.

Uno de los primeros modelos que surgió fue el modelo *probit* que usa la distribución normal como distribución de tolerancia, tal que:

$$p = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{1}{2}\left(\frac{s - \mu}{\sigma}\right)^2\right\} ds = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

donde Φ denota la función de probabilidad acumulativa para la distribución normal estándar. Así,

$$\Phi^{-1}(p) = \beta_1 + \beta_2 x, \quad \text{donde } \beta_1 = -\frac{\mu}{\sigma}, \beta_2 = \frac{1}{\sigma} \text{ y } g = \Phi^{-1}$$

El modelo probit tiene aplicaciones en biología y ciencias sociales. En biología, cuando $x = \mu$ este modelo se llama la dosis letal $LD(50)$ porque corresponde a la dosis donde puede esperarse que la mitad de las observaciones sea un éxito.

Otro modelo que proporciona resultados muy parecidos al modelo probit es el modelo *logit* o *logístico*. La distribución de tolerancia es:

$$f(s) = \frac{\beta_2 \exp\{\beta_1 + \beta_2 s\}}{[1 + \exp(\beta_1 + \beta_2 s)]^2}$$

tal que

$$p = \int_{-\infty}^x f(s) ds = \frac{\exp(\beta_1 + \beta_2 x)}{1 + \exp(\beta_1 + \beta_2 x)}$$

con función liga

$$\log\left(\frac{p}{1-p}\right) = \beta_1 + \beta_2 x$$

El término $\log\left(\frac{p}{1-p}\right)$ es algunas veces llamado función *logit* y tiene una interpretación natural como el logaritmo de momios. Este modelo es muy usado para datos binarios ($n_i = 1$) y es implementado en muchos paquetes estadísticos. La forma de la función $f(s)$ así, como p es similar a la del modelo probit, excepto en las colas de las distribuciones. La ventaja de este modelo es que la liga corresponde a la liga canónica y por lo tanto los parámetros tienen estadísticas suficientes de la misma dimensión. En este caso, si tenemos una variable explicativa indicadora (0, 1) el coeficiente β_2 para esa variable representa el logaritmo del cociente de momios al pasar x de cero a uno, manteniendo todas las demás variables constantes. Cuando x es continua este coeficiente representa el logaritmo del cociente de momios al incrementarse x en uno. De igual manera, $\exp\{\beta_2 q\}$ representa el cociente de momios al incrementarse x en q .

Si la función de distribución de tolerancia es

$$f(s) = \beta_2 \exp[(\beta_1 + \beta_2 s) - \exp(\beta_1 + \beta_2 s)]$$

entonces

$$p = 1 - \exp[-\exp(\beta_1 + \beta_2 x)]$$

y así la función liga es

$$\log[-\log(1 - p)] = \beta_1 + \beta_2 x$$

conocida como la función *doble logaritmo complementario*. Este modelo es similar al logístico y probit para valores de p cercanos a 0.5 pero diferente de ellos en valores de p cercanos a 0 ó 1, a diferencia del logístico y del probit no es simétrico alrededor de 0.5; y es más apropiado cuando la probabilidad de unos (o ceros) es mayor que la de ceros (o unos).

La función liga canónica para los modelos binomiales, como se mencionó anteriormente, es llamada logística. La figura (1.1) muestra la gráfica de los modelos anteriores.

1.2.3. Modelo Poisson y Loglineal.

La distribución Poisson se usa con frecuencia para modelar datos de conteo. Si Y es el número de ocurrencias, su función de distribución puede escribirse como

$$f(y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \dots$$

que vista como miembro de la familia exponencial, tiene la expresión

$$f(y) = \exp(y \ln \mu - \mu) / y!$$

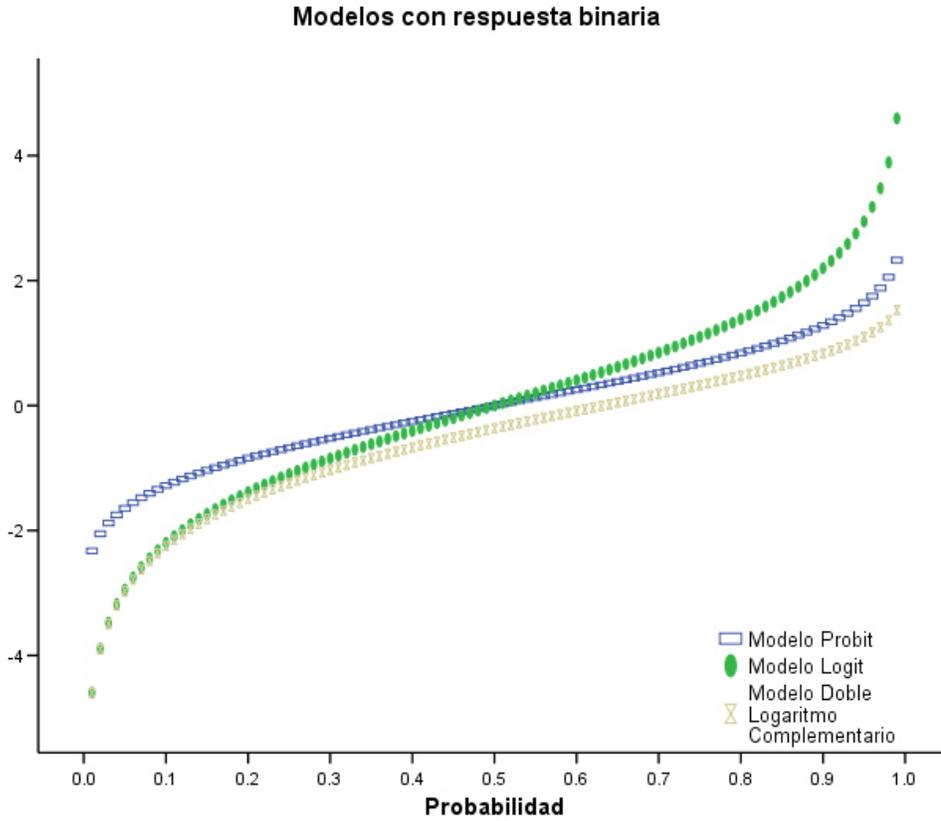


Figura 1.1: Modelos GLM con respuesta binaria

donde μ es el número promedio de ocurrencias, es decir, $E(Y) = \mu$ y $Var(Y) = \mu$. Por lo tanto

$$g(\mu_i) = x_i^T \beta \quad \rightarrow \quad \log(\mu_i) = x_i^T \beta$$

Sean Y_1, Y_2, \dots, Y_n v.a.i. con Y_i denotamos el número de eventos observados de un máximo posible n_i para la i -ésima covariable. El valor esperado de Y_i puede escribirse como

$$E(Y_i) = \mu_i = n_i \theta_i.$$

La dependencia de θ_i sobre las variables explicativas es generalmente modelada por

$$\theta_i = e^{x_i^T \beta}.$$

Por lo tanto, el modelo lineal generalizado es

$$E(Y_i) = \mu_i = n_i e^{x_i^T \beta}, \quad Y_i \sim \text{Poisson}(\mu).$$

La función liga natural es la función logaritmo

$$\log \mu_i = \log n_i + x_i^T \beta.$$

Esta ecuación difiere de las anteriores componentes lineales por el término $\log n_i$. Este se conoce como *offset* y es una constante conocida que se incorpora a los procedimientos de estimación con coeficiente uno. Los términos x_i y β son como antes la covariable y el parámetro respectivamente.

Cuando una variable explicativa binaria se denota como $x_j = 0$ si el factor está ausente y $x_j = 1$ si el factor está presente, β_j es la razón de cambio en el incremento en $E(Y_i)$ cuando x_j pasa de cero a uno con todas las demás constantes. Similarmente para variables explicativas continuas $\exp(\beta_j)$ representa un incremento en y_i cuando x_j se incrementa en uno y todas las demás permanecen constantes.

1.2.4. Ejemplos de Modelos probabilísticos para tablas de contingencia.

Sea el vector \mathbf{y} las frecuencias Y_i en N celdas de una tabla de clasificación cruzada, los modelos para tablas de contingencia son.

Modelo Poisson. Si las Y_i 's son variables aleatorias independientes con parámetros $E(Y_i) = \mu_i$ y distribución de probabilidad conjunta

$$f(\mathbf{y}; \boldsymbol{\mu}) = \prod_{i=1}^N \mu_i^{y_i} e^{-\mu_i} y_i!,$$

donde $\boldsymbol{\mu}$ es un vector de μ_i 's.

Modelo Multinomial. Si la única restricción es que la suma de las y_i 's sea n (observaciones clasificadas en N celdas), entonces la distribución multinomial podría usarse

$$f(\mathbf{y}; \boldsymbol{\mu} | n) = n! \prod_{i=1}^N \theta_i^{y_i} y_i!, \quad \text{donde} \quad \sum_{i=1}^N \theta_i = 1 \quad \sum_{i=1}^N y_i = n.$$

En este caso, $E(Y_i) = n\theta_i$. Para tablas de contingencia de dimensión dos, si j y k denotan renglón y columna, entonces la hipótesis considera comunmente que las variables renglón y columna son independientes, es decir,

$$\theta_{jk} = \theta_{j.} \theta_{.k}$$

donde $\theta_{j.}$ y $\theta_{.k}$ denotan la probabilidad marginal con $\sum_j \theta_{j.} = 1$ y $\sum_k \theta_{.k} = 1$. Esta hipótesis puede ser probada al comparar dos modelos lineales para el logaritmo de $\mu_{jk} = E(Y_{jk})$; a saber

$$\log \mu_{jk} = \log n + \log \theta_{jk}$$

y

$$\log \mu_{jk} = \log n + \log \theta_{j\cdot} + \log \theta_{\cdot k}.$$

Modelos de producto multinomial. Si hay más marginales fijas sobre el total n , entonces los productos apropiados de distribuciones multinomiales se pueden usar para modelar los datos. Por ejemplo, para una tabla de dimensión tres con J renglones, K columnas y L capas, si el total por renglón es fijo en cada capa, la probabilidad conjunta para las y'_{jkl} s es

$$f(\mathbf{y}|y_{j\cdot l}, j = 1, \dots, J, l = 1, \dots, L) = \prod_{j=1}^J \prod_{l=1}^L y_{j\cdot l}! \prod_{k=1}^K \theta_{jkl}^{y_{jkl}} y_{jkl}!$$

donde $\sum_k \theta_{jkl} = 1$ para cada combinación de j y l . En este caso, $E(Y_{jkl}) = y_{j\cdot l} \theta_{jkl}$.

Si únicamente los totales por capas son fijos, entonces

$$f(\mathbf{y}|y_{\cdot l}, l = 1, \dots, L) = \prod_{l=1}^L y_{\cdot l}! \prod_{j=1}^J \prod_{k=1}^K \theta_{jkl}^{y_{jkl}} y_{jkl}!$$

con $\sum_j \sum_k \theta_{jkl} = 1$ para $l = 1, \dots, L$. También $E(Y_{jkl}) = y_{\cdot l} \theta_{jkl}$.

Todos los modelos anteriores son equivalentes en verosimilitud, en todos los casos $E(Y_i)$ puede escribirse como producto de parámetros y otros términos. Además, de ser equivalente al modelo Poisson, la función logaritmo, contiene una componente lineal

$$\log E(Y_i) = \text{constante} + \mathbf{x}_i^T \boldsymbol{\beta}.$$

Los modelos lineales generalizados que tienen esta función liga se llaman modelos *log-lineales*.

1.2.5. Formas de estimación.

Estimación por máxima verosimilitud.

La función log-verosimilitud para la familia exponencial está dada por

$$\ell = \sum_{i=1}^n \frac{y_i \theta_i - d(\theta_i)}{a(\phi)} - \sum_{i=1}^n c(y_i, \phi).$$

Para obtener los estimadores máximo verosimiles de β , se obtienen la primera y segunda derivada de ℓ con respecto a β , de donde se tiene

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \frac{1}{a(\phi)} \sum \left[y_i \frac{\partial \theta_i}{\partial \beta} - \frac{\partial d(\theta_i)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta} \right] \\ &= \frac{1}{a(\phi)} \sum (y_i - \mu_i) \frac{\partial \theta_i}{\partial \beta} = \frac{1}{a(\phi)} \sum (y_i - \mu_i) \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta} \end{aligned}$$

por la regla de la cadena, y dado que $\frac{\partial d(\theta_i)}{\partial \theta_i} = \mu_i$

donde

$$\begin{aligned} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta} &= \left(\frac{\partial \mu_i}{\partial \theta_i} \right)^{-1} \frac{\partial \mu_i}{\partial g(\mu_i)} \frac{\partial g(\mu_i)}{\partial \beta} = \left(\frac{\partial^2 d(\theta_i)}{\partial \theta_i^2} \right)^{-1} \left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} \frac{\partial g(\mu_i)}{\partial \beta} \\ &= \frac{1}{v(\mu_i)} \left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} \frac{\partial \mathbf{x}_i^T \boldsymbol{\beta}}{\partial \beta} = \frac{1}{v(\mu_i) g_\mu(\mu_i)} \mathbf{x}_i^T \quad \text{donde } \frac{\partial^2 d(\theta_i)}{\partial \theta_i^2} = v(\mu_i) \text{ y } g_\mu(\mu_i) = \frac{\partial g(\mu_i)}{\partial \mu_i}. \end{aligned}$$

Entonces

$$\frac{\partial l}{\partial \beta} = \frac{1}{a(\phi)} \sum (y_i - \mu_i) \frac{1}{v(\mu_i)g_\mu(\mu_i)} x_i^T = \frac{1}{a(\phi)} \sum (y_i - \mu_i) w_i g_\mu(\mu_i) x_i^T$$

donde $w_i = (v(\mu_i)g_\mu^2(\mu_i))^{-1}$, que escrito en notación matricial puede escribirse como

$$\frac{\partial l}{\partial \beta} = \frac{1}{a(\phi)} \mathbf{X}^T \mathbf{W} \Delta (\mathbf{y} - \boldsymbol{\mu})$$

con $\mathbf{W} = \text{diag}\{w_i\}$ y $\Delta = \text{diag}\{g_\mu(\mu_i)\}$. Igualando a cero $\frac{\partial l}{\partial \beta}$, se tiene que

$$\mathbf{X}^T \mathbf{W} \Delta \mathbf{y} = \mathbf{X}^T \mathbf{W} \Delta \boldsymbol{\mu}, \quad (1.2)$$

donde \mathbf{W} , Δ , $\boldsymbol{\mu}$ son funciones del parámetro β desconocido. Típicamente estas funciones no son funciones lineales de β y así (1.2) no puede ser resuelta analíticamente.⁴

Las ecuaciones máximo verosímiles para β generalmente se resuelven por un método iterativo de mínimos cuadrados ponderado. Este se deriva como un ejemplo del uso del método de puntajes de Fisher. El puntaje de Fisher es un método iterativo para maximizar una verosimilitud y tiene la forma

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + \text{boldsymbolsymbol}I(\boldsymbol{\theta}^{(m)})^{-1} \left. \frac{\partial l}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(m)}}$$

donde (m) indica la m -ésima iteración, $\mathbf{I}(\boldsymbol{\theta})$ es la matriz de información de Fisher y $\boldsymbol{\theta}$ es el vector de parámetros de entrada.

Así, para el modelo tenemos que

⁴Esta información se encuentra en la referencia [3].

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \boldsymbol{\Delta} (\mathbf{y} - \boldsymbol{\mu})$$

donde \mathbf{W} , $\boldsymbol{\Delta}$, $\boldsymbol{\mu}$ son evaluadas en $\boldsymbol{\beta}^{(m)}$. Con

$$\text{Var}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}(\mathbf{y} - \boldsymbol{\mu})) = \text{Var}(\boldsymbol{\Delta}(\mathbf{y} - \boldsymbol{\mu})) = \text{diag}\{a(\phi)v(\mu)g_{\mu}^2(\mu_i)\} = a(\phi)\mathbf{W}^{-1}.$$

Por lo que una regresión ponderada de $\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}(\mathbf{y} - \boldsymbol{\mu})$ sobre \mathbf{X} usando pesos iguales a la inversa de su varianza da

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} [\mathbf{X}\boldsymbol{\beta}^{(m)} + \boldsymbol{\Delta}(\mathbf{y} - \boldsymbol{\mu})] = \boldsymbol{\beta}^{(m)} + (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \boldsymbol{\Delta} (\mathbf{y} - \boldsymbol{\mu})$$

que es la misma que la utilizada en la iteración.

Estimación por máxima quasi-verosimilitud.

Hay ocasiones en donde no se conoce una distribución de los datos y a pesar de ello se desea tener métodos inferenciales con los cuales trabajar como en el caso de máxima verosimilitud. Bajo esta idea se desarrolla la estimación por quasi-verosimilitud, para derivar una cantidad tipo verosimilitud cuya construcción requiere pocas suposiciones. Esto es, algo en donde sea fácil de imitar las propiedades que derivan del log-verosimilitud (o función de puntajes), más que la verosimilitud misma.

Definiendo un análogo de la verosimilitud, excepto que diferenciamos con respecto a μ_i en vez de θ_i . Primero, sabemos que

$$E \left(\frac{\partial \log f_{Y_i}(y_i)}{\partial \mu_i} \right) = 0$$

Entonces, por la regla de la cadena, tenemos que

$$\begin{aligned} v^* &= \text{var} \left(\frac{\partial \log f_{Y_i}(y_i)}{\partial \mu_i} \right) = \text{var} \left(\frac{\partial \log f_{Y_i}(y_i)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \right) = \text{var} \left(\frac{\partial \log f_{Y_i}(y_i)}{\partial \theta_i} \right) \left(\frac{\partial \theta_i}{\partial \mu_i} \right)^2 = \\ &= -E \left(\frac{\partial^2 \log f_{Y_i}(y_i)}{\partial \theta_i^2} \right) \left(\frac{\partial \theta_i}{\partial \mu_i} \right)^2 = \frac{1}{a(\phi)} \frac{\partial^2 d(\theta_i)}{\partial \theta_i^2} \left(\frac{\partial \theta_i}{\partial \mu_i} \right)^2 = \frac{v(\mu_i)}{a(\phi)} \left(\frac{\partial \theta_i}{\partial \mu_i} \right)^2 = \\ &= \frac{v(\mu_i)}{a(\phi)} \left(\frac{\partial \mu_i}{\partial \theta_i} \right)^{-2} = \frac{v(\mu_i)}{a(\phi)} \left(\frac{\partial^2 d(\theta_i)}{\partial \theta_i^2} \right)^{-2} = \frac{v(\mu_i)}{a(\phi)} \frac{1}{(v(\mu_i))^2} = \frac{1}{a(\phi)v(\mu_i)} \end{aligned}$$

donde $\mu_i = d'(\theta_i)$, y $v(\mu_i) = \partial^2 d(\theta_i) / \partial \theta_i^2$. Por lo tanto,

$$\text{var} \left(\frac{\partial \log f_{Y_i}(y_i)}{\partial \mu_i} \right) = -E \left(\frac{\partial^2 \log f_{Y_i}(y_i)}{\partial \mu_i^2} \right) = \frac{1}{a(\phi)v(\mu_i)}.$$

Estas ecuaciones son análogas a las ecuaciones de máxima verosimilitud. Además, buscamos una cantidad en lugar de $\partial \log f_{Y_i}(y_i) / \partial \mu_i$, la cual cumple las propiedades anteriores.

Sea $q_i = \frac{y_i - \mu_i}{a(\phi)v(\mu_i)}$, donde $\text{var}(y_i) \propto v(\mu_i)$. Esto lo cumplen todas las distribuciones de la familia exponencial presentadas antes. Así, la contribución del log-verosimilitud de y_i es la integral con respecto a μ_i de $\frac{\partial \log f_{Y_i}(y_i)}{\partial \mu_i}$, definimos el log quasi-verosimilitud vía la contribución y_i hecha, como

$$\begin{aligned} Q_i &= \int_{y_i}^{\mu_i} \frac{y_i - \mu_i}{a(\phi)v(t)} dt \\ \text{donde } \frac{\partial Q_i}{\partial \mu_i} &= \frac{y_i - \mu_i}{a(\phi)v(\mu_i)} = q_i. \end{aligned}$$

Finalmente, los estimadores máximo quasi-verosímiles de β , se encuentran resolviendo el máximo de las ecuaciones quasi-verosímiles.

$$\frac{\partial}{\partial \beta} \sum Q_i = 0 \quad \Leftrightarrow$$

$$\sum \frac{\partial Q_i}{\partial \beta} = \sum \frac{\partial Q_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta} = \sum q_i \frac{\partial \mu_i}{\partial \beta} = \sum \frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \frac{\partial \mu_i}{\partial \beta} = 0$$

donde

$$\frac{\partial \mu_i}{\partial \beta} = \frac{\partial \mu_i}{\partial g(\mu_i)} \frac{\partial g(\mu_i)}{\partial \beta} = \left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} \frac{\partial x_i^T \beta}{\partial \beta} = \frac{x_i^T}{g_\mu(\mu_i)}$$

Entonces,

$$\sum \frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \frac{x_i^T}{g_\mu(\mu_i)} = \sum \frac{y_i - \mu_i}{a(\phi)v(\mu_i)g_\mu(\mu_i)} x_i^T = 0.$$

Es decir,

$$\frac{1}{a(\phi)} \mathbf{X}^T \mathbf{W} \Delta (\mathbf{y} - \boldsymbol{\mu}) = 0,$$

que es el mismo estimador que en el caso de máxima verosimilitud, si la distribución de la variable aleatoria es miembro de la familia exponencial.

En donde, Q_i es construido usando solo información acerca de como la varianza cambia con la media y nada más. Es frecuente, que cuando especificamos la relación media–varianza obtenemos las ecuaciones máximo quasi–verosimiles, las cuales son iguales a las correspondientes en la verosimilitud legítima, cuando tenemos distribuciones que pertenecen a la familia exponencial. Además, el estimador de máxima quasi–verosimilitud proporciona dos grados de robustez; primero, no necesita suponer una distribución y segundo, sólo tiene que especificar la relación media–varianza para un contraste proporcional, que puede ser estimado de los datos. Sin embargo, si $a(\phi)$ es desconocida debe ser estimada por ejemplo, vía un estimador de momentos

$$\hat{\alpha}(\phi) = \frac{1}{n-p} \sum \frac{(y_i - \hat{\mu}_i)^2}{v(\mu_i)}$$

donde n es el número de observaciones y p es la dimensión de β .

1.3. MODELOS LINEALES MIXTOS (LMM).

Muchos modelos estadísticos pueden ser expresados como modelos lineales que consideran efectos fijos, es decir, parámetros asociados con una población o con ciertos niveles repetidos de factores experimentales; y efectos aleatorios, es decir, asociados con unidades experimentales individuales extraídas aleatoriamente de una población. Un modelo con efectos fijos y aleatorios se llama *modelo de efectos mixtos*.⁵

Los modelos de efectos mixtos son usados para describir la relación entre una variable respuesta y alguna covariable en datos agrupados de acuerdo a uno o más factores de clasificación. Ejemplo de datos agrupados incluyen datos longitudinales, datos con medidas repetidas, datos multinivel y diseño por bloques. Los modelos de efectos mixtos presentan gran flexibilidad en la estructura de covarianza inducida por los grupos en los datos.

Los efectos fijos se usan para modelar la media de y , mientras que los efectos aleatorios gobiernan la estructura de varianza-covarianza de y . Así, los efectos aleatorios simplifican la difícil tarea de especificar los distintos elementos de $var(y)$. Dicho de otro modo, si no se consideran los efectos aleatorios, se pueden analizar los elementos de la varianza de y de forma, pero con los factores aleatorios podemos analizarlos con atributos de varianza y covarianza. Así, como estos efectos son diferentes, requieren un análisis diferente.

⁵Ver referencia [4]

1.3.1. Definición del modelo.

Los modelos lineales de efectos mixtos son modelos con efectos fijos y aleatorios en los que se tiene linealidad en ambos efectos dentro de la función del modelo. Los efectos aleatorios pueden ser considerados como un término del error entre las correlaciones de las observaciones dentro del mismo grupo.

Un solo nivel de agrupación.

Los modelos lineales de efectos mixtos para un solo nivel de agrupación expresan un vector \mathbf{y}_i respuesta n_i -dimensional para el i -ésimo grupo como

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M. \quad (1.3)$$

$$\text{donde} \quad \mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

$\boldsymbol{\beta}$ es un vector p -dimensional de efectos fijos, \mathbf{b}_i es un vector de efectos aleatorios q -dimensional, \mathbf{X}_i es una matriz conocida de variables con efectos fijos de tamaño $n_i \times p$, \mathbf{Z}_i es una matriz conocida de variables con efectos aleatorios de tamaño $n_i \times q$ y $\boldsymbol{\epsilon}_i$ es el vector de errores dentro del grupo de dimensión n_i con distribución normal multivariada. La suposición $\text{var}(\boldsymbol{\epsilon}_i) = \sigma^2 \mathbf{I}$ se puede relajar, de tal forma que se permite modelar varianzas no constantes o estructuras especiales de correlación dentro de los grupos. Los efectos aleatorios \mathbf{b}_i y los errores dentro del grupo $\boldsymbol{\epsilon}_i$ se suponen independientes para diferentes grupos e independientes dentro del grupo.

Como la distribución del vector de efectos aleatorios \mathbf{b}_i se supuso normal con media $\mathbf{0}$, está completamente caracterizado por su matriz de varianza-covarianza $\boldsymbol{\Psi}$, la cual debe ser simétrica y semidefinida positiva, es decir,

todos sus valores propios deben ser no negativos. Aunque se puede suponer que Ψ es definida positiva (valores propios estrictamente positivos), por ser un modelo indefinido, siempre puede ser re-expresado como un modelo definido positivo de baja dimensión.

Los efectos aleatorios \mathbf{b}_i son definidos con media $\mathbf{0}$ y por lo tanto cualquier media diferente de cero para un término en los efectos aleatorios debe ser expresado como parte de los términos de efectos fijos. Además, las columnas de \mathbf{Z}_i generalmente son un subconjunto de las columnas de \mathbf{X}_i .

Para efectos de cómputo es conveniente expresar la matriz de varianza-covarianza en la forma de un factor de precisión relativo, Δ , el cual es cualquier matriz que satisfaga

$$\frac{\Psi^{-1}}{1/\sigma^2} = \Delta^T \Delta.$$

Si Ψ es definida-positiva, entonces Δ existe aunque no necesariamente es única. Posiblemente se elige el factor Cholesky de $\sigma^2\Psi^{-1}$ como Δ . La matriz Δ es el *factor de precisión relativo* porque son factores de la *matriz de precisión* Ψ^{-1} de los efectos aleatorios, expresados en términos de la precisión $1/\sigma^2$ de ϵ_i .

Modelos lineales de efectos mixtos multinivel.

El modelo anterior puede ser extendido a efectos aleatorios de niveles múltiples anidados. En el caso de dos niveles anidados de efectos aleatorios el vector respuesta del nivel más interno de grupos es \mathbf{y}_{ij} , $i = 1, \dots, M$, $j = 1, \dots, M_i$ donde M es el número de grupos en el primer nivel y M_i es el número de grupos en el segundo nivel dentro del grupo i del primer nivel. El tamaño de \mathbf{y}_{ij} es n_{ij} , las matrices del modelo de efectos fijos son *boldsymbol* X_{ij} , $i = 1, \dots, M$, $j = 1, \dots, M_i$ de tamaño $n_{ij} \times p$. Usando efectos aleatorios

para el primer nivel \mathbf{b}_i de tamaño q_1 y efectos aleatorios para el segundo nivel \mathbf{b}_{ij} de tamaño q_2 con matrices $\mathbf{Z}_{i,j}$ de $n_i \times q_1$ y \mathbf{Z}_{ij} de $n_i \times q_2$ respectivamente, el modelo se escribe como

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{i,j}\mathbf{b}_i + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, M_i$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}_1), \quad \mathbf{b}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Psi}_2), \quad \boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Los efectos aleatorios del nivel 1 \mathbf{b}_i se suponen independientes para diferente i , los efectos aleatorios del nivel 2 \mathbf{b}_{ij} se suponen independientes para diferentes i ó j y son independientes de los efectos aleatorios del nivel 1 y los errores dentro del grupo $\boldsymbol{\epsilon}_{ij}$ se suponen independientes para diferentes i ó j , e independientes de los efectos aleatorios.

Si se consideran Q niveles de efectos aleatorios, el modelo tiene la misma estructura que en el caso de 2 niveles, con matriz de varianza-covarianza, $\boldsymbol{\Psi}_q$, $q = 1, \dots, Q$, para los efectos aleatorios del nivel q , y es simétrica, definida positiva, pudiendo ser diagonal o múltiple de la identidad. El número de niveles es el número de niveles anidados de efectos aleatorios.

Forma general del modelo extendido.

Extender el modelo lineal mixto de un nivel a una forma más general permite relajar la suposición de independencia del vector de errores dentro del grupo $\boldsymbol{\epsilon}_i$, a la heteroscedasticidad y correlación dentro de grupos, la cual es expresada como

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M \quad (1.4)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \Psi), \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{\Lambda}_i), \quad i = 1, \dots, M$$

donde las matrices $\mathbf{\Lambda}_i$ son definidas positivas parametrizadas por un conjunto de parámetros $\boldsymbol{\lambda}$, generalmente pequeños y fijos. Así, los $\mathbf{\Lambda}_i$ suponen independencia para diferentes grupos, e independencia de los efectos aleatorios \mathbf{b}_i .

Similarmente, la extensión a dos niveles considera a

$$\boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \sigma^2 \mathbf{\Lambda}_{ij}), \quad i = 1, \dots, M, \quad j = 1, \dots, M_i$$

donde $\mathbf{\Lambda}_{ij}$ son matrices definidas positivas parametrizadas por un vector $\boldsymbol{\lambda}$ fijo. De igual manera para un modelo multinivel con Q niveles de efectos aleatorios.

1.3.2. Ejemplos.

Diseño por bloques aleatorizado.

Un diseño por bloques aleatorizado es un tipo de experimento en el cual hay dos factores de clasificación: un factor experimental para el cual usamos efectos fijos y un factor por bloques para el cual usamos efectos aleatorios. Un ejemplo esquematizará el modelo.

Se realizó un experimento ergonómico, con un diseño por bloques aleatorizado. El experimento registra el esfuerzo requerido por nueve diferentes individuos para levantar cuatro tipos de taburetes. Queremos comparar estos cuatro tipos particulares de taburetes, así utilizamos efectos fijos para el factor *tipo*. Los nueve diferentes individuos representan una muestra de la población acerca de la cual deseamos hacer inferencia, así usamos efectos

aleatorios para modelar el factor *individuo*. Después de analizar los datos se observó una diferencia sistemática entre los tipos de taburetes sobre sus medidas, dándose que los promedios de la variable *tipo* aparecen ordenadas así, $T_1 \leq T_4 \leq T_3 \leq T_2$.

Un modelo con efectos fijos β_j para el factor *tipo* y efectos aleatorios b_i para el factor *individuo* podría ser escrito como

$$y_{ij} = \beta_j + b_i + \epsilon_{ij}, \quad i = 1, \dots, 9, \quad j = 1, \dots, 4$$

$$b_i \sim N(0, \sigma_b^2), \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

o equivalentemente

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i b_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 9,$$

donde

$$b_i \sim N(0, \sigma_b^2), \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

donde para $i = 1, \dots, 9$

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Z}_i = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{pmatrix}$$

La matriz de efectos fijos \mathbf{X}_i es una matriz de celdas promedios, ya que la j -ésima componente de $\boldsymbol{\beta}$ representa lo que podría ser el esfuerzo medio que representa el j -ésimo tipo de taburete si la población total fuera considerada.

Modelos de efectos mixtos para diseño por bloques, repetidos.

El modelo anterior puede verse como un diseño por bloques no repetido, pues una única observación está disponible en cada combinación de condiciones experimentales. En otros experimentos puede ser necesario repetir medidas. Por ejemplo, los datos de máquinas que dan la producción para la elegir aleatoriamente de seis trabajos probados sobre tres diferentes tipos de máquinas. Cada trabajo usa cada máquina tres veces, así tiene tres réplicas en cada conjunto de condiciones.

Tomando el factor *trabajo* como efecto aleatorio y el factor *máquina* como efecto fijo. Las repeticiones en este experimento nos permiten calcular la presencia de interacciones entre *trabajo* y *máquina*. Esto es, podemos direccionar la cuestión de si el efecto de cambiar de un tipo de máquina a otro es diferente para diferentes trabajos.

Como el *trabajo* representa una muestra aleatoria de la población de interés, cualquier término de interacción puede modelar diferencias entre los *trabajos* al cambiar de una *máquina* a otra, también se puede expresar a la variable *interacción* como un efecto aleatorio. El modelo incorporando el término *interacción* como efecto aleatorio es: b_{ij} , $i = 1, \dots, 6$ y $j = 1, \dots, 3$

$$y_{ijk} = \beta_j + b_i + b_{ij} + \epsilon_{ijk}, \quad i = 1, \dots, 6, \quad j = 1, \dots, 3, \quad k = 1, \dots, 3$$

$$b_i \sim N(0, \sigma_1^2), \quad b_{ij} \sim N(0, \sigma_2^2), \quad \epsilon_{ijk} \sim N(0, \sigma^2)$$

Este modelo presenta efectos aleatorios en dos niveles: el efecto b_i para el trabajo y el efecto b_{ij} para el tipo de máquina dentro de cada trabajo, es decir, la *interacción*.

También puede existir una relación de varianza–covarianza en el efecto aleatorio b_i , en donde el modelo sería el mismo, solo que $b_i \sim N(\mathbf{0}, \Psi)$, con Ψ definida positiva, simétrica, la cual representa a la matriz de varianza–covarianza de b_i .

Un modelo de análisis de covarianza.

Un modelo lineal, tal como

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i, \quad i = 1, \dots, N, \quad \epsilon_i \sim N(0, \sigma^2)$$

relaciona una variable respuesta continua y_i con una o más covariables continuas x_i .

El término análisis de covarianza designa un tipo de modelo que relaciona una respuesta continua con un factor de clasificación y una covariable continua. Si y_{ij} es la j -ésima observación en el i -ésimo grupo de datos y x_{ij} es el valor correspondiente de la covariable, un análisis de covarianza con un efecto aleatorio para la intersección podría ser:

$$y_{ij} = \beta_1 + b_i + \beta_2 x_{ij} + \epsilon_{ij}, \quad i = 1, \dots, M, j = 1, \dots, n_i,$$

$$b_i \sim N(0, \sigma_b^2), \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

El cual combina un modelo de análisis de efectos aleatorios de varianza con un modelo de regresión lineal.

Una aplicación común de modelos de efectos aleatorios de análisis de covarianza es modelar datos de curvas de crecimiento, o más generalmente conocidos como *datos longitudinales* o medidas repetidas.

Consideremos un conjunto de medidas de la distancia de la glándula pituitaria a la fisura pterygomaxial que se tomaron por dos años completos a once niñas entre 8 y 14 años de edad. Se desea ver si hay diferencias cualitativas entre su patrón de crecimiento para esta medida, para ello se toma el siguiente modelo

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 11,$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

con matrices

$$\mathbf{X}_i = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix}, \quad \mathbf{Z}_i = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

El vector de efectos fijos $\boldsymbol{\beta}$ bidimensional, consiste de la intersección media β_1 , para la población y la pendiente común o razón de crecimiento β_2 . Los vectores de efectos aleatorios \mathbf{b}_i , $i = 1, \dots, 11$ unidimensionales describen algún cambio en la intersección por cada niña. La matriz $\boldsymbol{\Psi} = \sigma_b^2$ de tamaño 1×1 , representa la varianza de las medidas para una edad fija.

1.3.3. Formas de estimación.

Estimación por máxima verosimilitud para LMM en un nivel.

Considerando el modelo (1.3) que tiene un solo nivel de efectos aleatorios. Los parámetros del modelo son $\boldsymbol{\beta}, \sigma^2$ y los parámetros determinados por

Δ el factor de precisión relativo. Usando $\boldsymbol{\theta}$ para representar el conjunto de parámetros que determinan a Δ .

La función de verosimilitud para el modelo (1.3) es la densidad de probabilidad dada por

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$$

donde L es la función de verosimilitud, p es la función de densidad de probabilidad y \mathbf{y} es la entrada del vector respuesta N -dimensional, $N = \sum_{i=1}^M n_i$.

Al ser los efectos aleatorios \mathbf{b}_i , $i = 1, \dots, M$ parte del modelo, debemos integrar la densidad condicional de los datos dados los efectos aleatorios con respecto a la densidad marginal de los efectos aleatorios para poder obtener la densidad marginal de los datos. Usando la independencia de \mathbf{b}_i y $\boldsymbol{\epsilon}_i$, esto se expresa como

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = \prod_{i=1}^M p(\mathbf{y}_i | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = \prod_{i=1}^M \int p(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) p(\mathbf{b}_i | \boldsymbol{\theta}, \sigma^2) d\mathbf{b}_i$$

donde la densidad condicional de \mathbf{y}_i , es una normal multivariada

$$p(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) = \frac{\exp\{-\|\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i\|^2 / 2\sigma^2\}}{(2\pi\sigma^2)^{n_i/2}}$$

y la densidad marginal de \mathbf{b}_i es también normal multivariada

$$p(\mathbf{b}_i | \boldsymbol{\theta}, \sigma^2) = \frac{\exp\{-\mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i\}}{(2\pi)^{q/2} \sqrt{|\boldsymbol{\Psi}|}} = \frac{\exp\{-\|\Delta \mathbf{b}_i\|^2 / 2\sigma^2\}}{(2\pi\sigma^2)^{q/2} \text{abs}|\Delta|^{-1}} \quad (1.5)$$

donde $|\cdot|$ denota el determinante de la matriz y abs es el valor absoluto. ⁶

⁶Consultar referencia [4].

Entonces la función de verosimilitud es

$$\begin{aligned}
 L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) &= \prod_{i=1}^M \frac{abs|\Delta|}{(2\pi\sigma^2)^{n_i/2}} \int \frac{\exp\{-\|\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2 + \|\Delta\mathbf{b}_i\|^2\}/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \\
 &= \prod_{i=1}^M \frac{abs|\Delta|}{(2\pi\sigma^2)^{n_i/2}} \int \frac{\exp\{-\|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \quad (1.6)
 \end{aligned}$$

donde

$$\tilde{\mathbf{y}}_i = \begin{pmatrix} \mathbf{y}_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbf{X}}_i = \begin{pmatrix} \mathbf{X}_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbf{Z}}_i = \begin{pmatrix} \mathbf{Z}_i \\ \Delta \end{pmatrix}$$

son vectores y matrices de datos aumentados.

El exponente de la integral en (1.5) es la suma de cuadrados residual por lo que puede ser reescrito como

$$\begin{aligned}
 \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2 &= \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2 + \|\tilde{\mathbf{Z}}_i(\mathbf{b}_i - \hat{\mathbf{b}}_i)\|^2 = \\
 &= \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2 + (\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i), \quad (1.7)
 \end{aligned}$$

donde $\hat{\mathbf{b}}_i$ es el estimador obtenido por mínimos cuadrados

$$\hat{\mathbf{b}}_i = (\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i)^{-1} \tilde{\mathbf{Z}}_i^T (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta})$$

El primer sumando de (1.7) no depende de \mathbf{b}_i , por lo que solo hay que integrar el segundo término de (1.7) en la integral de (1.5), entonces

$$\frac{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}} \int \frac{\exp\{-(\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i =$$

$$\frac{1}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}} \int \frac{\exp\{-(\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2} \sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}} d\mathbf{b}_i = \frac{1}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}} \quad (1.8)$$

Finalmente, sustituyendo (1.7) y (1.8) en (1.5) se obtiene

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) &= \prod_{i=1}^M \frac{abs|\Delta|}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}} \frac{\exp\{-\sum_{i=1}^M \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \tilde{\mathbf{Z}}_i \hat{\mathbf{b}}_i\|^2 / 2\sigma^2\}}{(2\pi\sigma^2)^{N/2}} \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\frac{\sum_{i=1}^M \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \tilde{\mathbf{Z}}_i \hat{\mathbf{b}}_i\|^2}{2\sigma^2}\right) \prod_{i=1}^M \frac{abs|\Delta|}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}} \end{aligned} \quad (1.9)$$

De donde los estimadores de máxima verosimilitud para $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, σ^2 se obtienen maximizando esta última expresión. Sin embargo, esto es más simple si nos concentramos en la verosimilitud como función de $\boldsymbol{\theta}$ únicamente. Es decir, calculamos los estimadores condicionales $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ y $\hat{\sigma}^2(\boldsymbol{\theta})$ como los valores que maximizan a $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ para un $\boldsymbol{\theta}$ dado. Como las partes en donde se involucra a $\boldsymbol{\beta}$ y σ^2 en la verosimilitud, tienen la forma de un modelo de regresión lineal, así $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ y $\hat{\sigma}^2(\boldsymbol{\theta})$ se pueden determinar por la teoría de regresión clásica.

El estimador de mínimos cuadrados para $\boldsymbol{\beta}$ dependerá condicionalmente de $\hat{\mathbf{b}}_i$, y éste a su vez, dependerá de $\boldsymbol{\beta}$, además, se quiere determinar el valor por mínimos cuadrados conjunto como solución a mínimos cuadrados de

$$(\hat{\mathbf{b}}_1^T, \dots, \hat{\mathbf{b}}_M^T, \hat{\boldsymbol{\beta}}^T)^T = \underset{\mathbf{b}_1, \dots, \mathbf{b}_M, \boldsymbol{\beta}}{\operatorname{argmin}} \|\mathbf{y}_e - \mathbf{X}_e(\mathbf{b}_1, \dots, \mathbf{b}_M, \boldsymbol{\beta})^T\|^2,$$

donde

$$\mathbf{y}_e = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{0} \\ \mathbf{y}_2 \\ \mathbf{0} \\ \vdots \\ \mathbf{y}_M \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{X}_e = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{X}_1 \\ \Delta & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} & \mathbf{X}_2 \\ \mathbf{0} & \Delta & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_M & \mathbf{X}_M \\ \mathbf{0} & \mathbf{0} & \dots & \Delta & \mathbf{0} \end{pmatrix}$$

De donde se tiene

$$(\hat{\mathbf{b}}_1^T, \dots, \hat{\mathbf{b}}_M^T, \hat{\boldsymbol{\beta}}^T)^T = (\mathbf{X}_e^T \mathbf{X}_e)^{-1} \mathbf{X}_e^T \mathbf{y}_e$$

y

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{\|\mathbf{y}_e - \mathbf{X}_e(\hat{\mathbf{b}}_1^T, \dots, \hat{\mathbf{b}}_M^T, \hat{\boldsymbol{\beta}}^T)^T\|^2}{N}.$$

Sustituyendo este valor en la verosimilitud de (1.9) se tiene

$$L(\boldsymbol{\theta}) = L(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}, \hat{\sigma}^2(\boldsymbol{\theta})) = \frac{\exp(-N/2)}{(2\pi\hat{\sigma}^2(\boldsymbol{\theta}))^{N/2}} \prod_{i=1}^M \frac{abs|\Delta|}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}}$$

Nótese que sólo es necesario conocer la norma de los residuales del problema de mínimos cuadrados aumentado y no los valores de $\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_M, \boldsymbol{\beta}$. Para encontrar esta solución se usará la descomposición ortogonal-triangular de matrices rectangulares, pues son fáciles de calcular computacionalmente. También es conocida como descomposición QR, pues

$$\mathbf{X} = \mathbf{Q} \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix} = \mathbf{Q}_t \mathbf{R},$$

donde \mathbf{X} es una matriz de $n \times p$ ($n \geq p$) de rango p , \mathbf{Q} es una matriz ortogonal de $n \times n$, \mathbf{R} es una matriz triangular superior de tamaño $p \times p$ y \mathbf{Q}_t (\mathbf{Q} -truncada) consiste de las primeras p columnas de \mathbf{Q} . Como \mathbf{Q} es ortogonal, entonces $\mathbf{Q}^T \mathbf{Q} = \mathbf{Q} \mathbf{Q}^T = \mathbf{I}$ y así, $\mathbf{Q}_t^T \mathbf{Q}_t = \mathbf{I}$, además al ser \mathbf{Q} una generalización de una rotación o una reflexión en el plano, \mathbf{Q} ó \mathbf{Q}^T preservan normas de vectores sobre multiplicación, es decir,

$$\|\mathbf{Q}^T \mathbf{y}\|^2 = (\mathbf{Q}^T \mathbf{y})^T \mathbf{Q}^T \mathbf{y} = \mathbf{y}^T \mathbf{Q} \mathbf{Q}^T \mathbf{y} = \mathbf{y}^T \mathbf{y} = \|\mathbf{y}\|^2.$$

Si aplicamos esto al vector residual en el problema de mínimos cuadrados, se tiene

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \|\mathbf{Q}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2 = \|\mathbf{Q}^T \mathbf{y} - \mathbf{Q}^T \mathbf{X}\boldsymbol{\beta}\|^2 =$$

$$\left\| \mathbf{c} - \mathbf{Q}^T \mathbf{Q} \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix} \boldsymbol{\beta} \right\|^2 = \left\| \mathbf{c} - \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix} \boldsymbol{\beta} \right\|^2 = \|\mathbf{c}_1 + \mathbf{R}\boldsymbol{\beta}\|^2 + \|\mathbf{c}_2\|^2,$$

donde $\mathbf{c} = (\mathbf{c}_1^T, \mathbf{c}_2^T) = \mathbf{Q}^T \mathbf{y}$ es el vector de residuales rotado. Las componentes \mathbf{c}_1 y \mathbf{c}_2 son de tamaño p y $n - p$ respectivamente.

Como \mathbf{X} es de rango p , la matriz \mathbf{R} de $p \times p$ es no singular y triangular superior. La solución de mínimos cuadrados $\hat{\boldsymbol{\beta}}$ es fácil de evaluar como la solución $\mathbf{R}\hat{\boldsymbol{\beta}} = \mathbf{c}_1$ y la suma de cuadrados residual es $\|\mathbf{c}_2\|^2$, la cual no depende de $\hat{\boldsymbol{\beta}}$.

Entonces descomponiendo la matriz aumentada $\tilde{\mathbf{Z}}_i$ como

$$\tilde{\mathbf{Z}}_i = \mathbf{Q}_{(i)} \begin{pmatrix} \mathbf{R}_{11(i)} \\ \mathbf{0} \end{pmatrix}$$

donde $\mathbf{Q}_{(i)}$ es de $(n_i + q) \times (n_i + q)$ y $\mathbf{R}_{11(i)}$ es de $q \times q$. Por lo que

$$\|\tilde{\mathbf{y}}_i + \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2 = \|\mathbf{Q}_{(i)}^T(\tilde{\mathbf{y}}_i + \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\mathbf{Z}}_i\mathbf{b}_i)\|^2 =$$

$$\|\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i\|^2 + \|\mathbf{c}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2,$$

donde la matriz $\mathbf{R}_{10(i)}$ es de $q \times p$, $\mathbf{R}_{00(i)}$ es de $n_i \times p$, el vector $\mathbf{c}_{1(i)}$ es de dimensión q y $\mathbf{c}_{0(i)}$ es un vector de dimensión n_i , se definen por

$$\begin{pmatrix} \mathbf{R}_{10(i)} \\ \mathbf{R}_{00(i)} \end{pmatrix} = \mathbf{Q}_{(i)}^T \tilde{\mathbf{X}}_i \quad \begin{pmatrix} \mathbf{c}_{1(i)} \\ \mathbf{c}_{0(i)} \end{pmatrix} = \mathbf{Q}_{(i)}^T \tilde{\mathbf{y}}_i$$

Sustituyendo esto en la integral de (1.6)

$$\begin{aligned} & \int \frac{\exp\{-(\|\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2 + \|\Delta\mathbf{b}_i\|^2)/2\sigma^2\}}{(2\pi\sigma^2)^{1/2}} d\mathbf{b}_i \\ &= \exp\left[\frac{\|\mathbf{c}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2}{-2\sigma^2}\right] \int \frac{\exp[\|\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i\|^2/2\sigma^2]}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \end{aligned} \quad (1.10)$$

Sea $\boldsymbol{\phi}_i = (\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i)/\sigma$ donde $\mathbf{R}_{11(i)}$ es no singular, entonces $d\boldsymbol{\phi}_i = \sigma^{-q} \text{abs}|\mathbf{R}_{11(i)}| d\mathbf{b}_i$, haciendo este cambio de variable, la integral es

$$\begin{aligned} & \int \frac{\exp[-\|\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i\|^2/2\sigma^2]}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i = \\ & \frac{1}{\text{abs}|\mathbf{R}_{11(i)}|} \int \frac{\exp(-\|\boldsymbol{\phi}_i\|^2/2)}{(2\pi)^{q/2}} d\boldsymbol{\phi}_i = \frac{1}{\text{abs}|\mathbf{R}_{11(i)}|} \end{aligned} \quad (1.11)$$

donde

$$\text{abs}|\mathbf{R}_{11(i)}| = \sqrt{(|\mathbf{R}_{11(i)}^T|)^2} = \sqrt{|\mathbf{R}_{11(i)}^T| |\mathbf{R}_{11(i)}|} = \sqrt{|\mathbf{R}_{11(i)}^T \mathbf{R}_{11(i)}|} =$$

$$= \sqrt{\left| \begin{pmatrix} \mathbf{R}_{11(i)}^T & \mathbf{0} \end{pmatrix} \mathbf{Q}_{(i)}^T \mathbf{Q}_{(i)} \begin{pmatrix} \mathbf{R}_{11(i)} \\ \mathbf{0} \end{pmatrix} \right|} = \sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}$$

Así, el valor de la integral es el mismo que en (1.6). Al ser $\mathbf{R}_{11(i)}$ triangular, su determinante es el producto de los elementos de la diagonal.

Sustituyendo (1.10) y (1.11) en (1.6) la función de verosimilitud es

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = \prod_{i=1}^M \frac{\exp[-\|\mathbf{c}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2/2\sigma^2]}{(2\pi\sigma^2)^{n_i/2}} \text{abs}\left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|}\right) =$$

$$\frac{\exp[-\sum_{i=1}^M \|\mathbf{c}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2/2\sigma^2]}{(2\pi\sigma^2)^{N/2}} \prod_{i=1}^M \text{abs}\left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|}\right)$$

Donde el término del exponente tiene la forma de la suma de cuadrados residual para $\boldsymbol{\beta}$ ponderada sobre todos los grupos. Formando una descomposición ortogonal–triangular nueva.

$$\begin{pmatrix} \mathbf{R}_{00(1)} & \mathbf{c}_{0(1)} \\ \vdots & \vdots \\ \mathbf{R}_{00(M)} & \mathbf{c}_{0(M)} \end{pmatrix} = \mathbf{Q}_{(0)} \begin{pmatrix} \mathbf{R}_{00} & \mathbf{c}_0 \\ \mathbf{0} & \mathbf{c}_{-1} \end{pmatrix}$$

La función de verosimilitud, tiene la forma

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-N/2} \exp\left[\frac{\|\mathbf{c}_{-1}\|^2 + \|\mathbf{c}_0 - \mathbf{R}_{00}\boldsymbol{\beta}\|^2}{-2\sigma^2}\right] \prod_{i=1}^M \text{abs}\left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|}\right) \quad (1.12)$$

Por lo que, para un θ dado, los valores de $\boldsymbol{\beta}$ y σ^2 que maximizan esta expresión son

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{R}_{00}^{-1} \mathbf{c}_0 \quad \text{y} \quad \hat{\sigma}^2(\boldsymbol{\theta}) = \frac{\|\mathbf{c}_{-1}\|^2}{N} \quad (1.13)$$

Evaluando estos valores en $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y})$ de (1.12) se tiene

$$L(\boldsymbol{\theta} | \mathbf{y}) = L(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}, \hat{\sigma}^2(\boldsymbol{\theta}) | \mathbf{y}) = \left(\frac{N}{2\pi \|\mathbf{c}_{-1}\|^2} \right)^{N/2} \exp \left(-\frac{N}{2} \right) \prod_{i=1}^M \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right)$$

que se denomina la verosimilitud perfil; y la log-verosimilitud perfil

$$\ell(\boldsymbol{\theta} | \mathbf{y}) = \log L(\boldsymbol{\theta} | \mathbf{y}) =$$

$$\frac{N}{2} [\log N - \log(2\pi) - 1] - N \log \|\mathbf{c}_{-1}\| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right) \quad (1.14)$$

La log-verosimilitud perfil se maximiza con respecto a $\boldsymbol{\theta}$, produciendo el estimador de máxima verosimilitud de $\hat{\boldsymbol{\theta}}$. El estimador de máxima verosimilitud para $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}^2$ se obtienen sustituyendo $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ en (1.13).

Aunque técnicamente los efectos aleatorios \mathbf{b}_i no son parámetros para el modelo estadístico frecuentemente se desea *estimar* sus valores. El modelo condicional de los efectos aleatorios, evaluado en el estimador condicional de $\boldsymbol{\beta}$, es el mejor predictor lineal insesgado o BLUP de \mathbf{b}_i , $i = 1, \dots, M$ y puede ser evaluado usando matrices de la descomposición ortogonal-triangular, como

$$\hat{\mathbf{b}}_i(\boldsymbol{\theta}) = \mathbf{R}_{11(i)}^{-1} (\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})),$$

así, si $\boldsymbol{\theta}$ es desconocido se reemplaza por su estimador de máxima verosimilitud $\hat{\boldsymbol{\theta}}$, produciendo el estimador BLUP $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$.

Estimación por verosimilitud restringida.

Algunas veces los estimadores de máxima verosimilitud de componentes de varianza tales como σ^2 y σ_b^2 tienden a sobreestimar los parámetros, por lo que se prefiere la estimación máximo verosimilitud restringida (o residual) REML para estas cantidades.

Hay varias maneras de definir el criterio de estimación REML. Una definición es suponer una distribución inicial uniforme para los efectos fijos $\boldsymbol{\beta}$ e integrarla fuera de la verosimilitud, es decir,

$$L_R(\boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = \int L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) d\boldsymbol{\beta}$$

Usando (1.12) y haciendo el cambio de variable

$$\boldsymbol{\phi}_i = (\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i)/\sigma$$

con diferencial $d\boldsymbol{\phi}_i = \sigma^{-q} \text{abs}|\mathbf{R}_{11(i)}| d\mathbf{b}_i$ obtenemos la log-verosimilitud-restringido

$$\ell_R(\boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = \log L_R(\boldsymbol{\theta}, \sigma^2 | \mathbf{y}) =$$

$$-\frac{N-p}{2} \log(2\pi\sigma^2) - \frac{\|\mathbf{c}_{-1}\|^2}{2\sigma^2} - \log \text{abs}|\mathbf{R}_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right)$$

esto produce el estimador condicional $\hat{\sigma}_R^2(\boldsymbol{\theta}) = \|\mathbf{c}_{-1}\|^2/(N-p)$ para σ^2 , por el cual se obtiene la función log-verosimilitud-restringida perfil

$$\ell_R(\boldsymbol{\theta} | \mathbf{y}) = \ell_R(\boldsymbol{\theta}, \hat{\sigma}_R^2(\boldsymbol{\theta}) | \mathbf{y}) =$$

$$\begin{aligned}
& -\frac{N-p}{2} \log(2\pi) - \frac{N-p}{2} \log(\sigma^2) - \frac{\|\mathbf{c}_{-1}\|^2}{2\sigma^2} - \log \text{abs}|\mathbf{R}_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right) = \\
& -\frac{N-p}{2} \log(2\pi) - \frac{N-p}{2} \log(\|\mathbf{c}_{-1}\|^2/(N-p)) - \frac{\|\mathbf{c}_{-1}\|^2}{2\|\mathbf{c}_{-1}\|^2/(N-p)} - \log \text{abs}|\mathbf{R}_{00}| \\
& \quad + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right) = \\
& \frac{N-p}{2} [\log((N-p)/2\pi) - 1] - \frac{N-p}{2} \log \|\mathbf{c}_{-1}\|^2 - \log \text{abs}|\mathbf{R}_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right) = \\
& \text{constante} - (N-p) \log \|\mathbf{c}_{-1}\| - \log \text{abs}|\mathbf{R}_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right) \quad (1.15)
\end{aligned}$$

donde $\text{constante} = \frac{N-p}{2} [\log((N-p)/2\pi) - 1]$

Los componentes de la función log-verosimilitud-restringida perfil en (1.15) son similares a la log-verosimilitud perfil (1.14) excepto que el logaritmo de la norma del vector residual tiene un múltiplo diferente y hay un término extra, el $\log \text{abs}|\mathbf{R}_{00}| = \log |\sum_{i=1}^M \mathbf{X}_i^T \Sigma_i^{-1} \mathbf{X}_i|/2$. Para evaluar el estimador de máxima verosimilitud restringida, se maximiza la log-verosimilitud-restringida perfil (1.15) con respecto a $\boldsymbol{\theta}$ únicamente y usando los resultados del estimador REML $\hat{\boldsymbol{\theta}}_R$ para obtener el estimador REML de σ^2 , $\hat{\sigma}_R^2(\hat{\boldsymbol{\theta}}_R)$. Similarmente, el estimador REML de BLUP para efectos aleatorios se obtienen reemplazando $\boldsymbol{\theta}$ por $\hat{\boldsymbol{\theta}}_R$ en $\hat{\mathbf{b}}_i(\boldsymbol{\theta}_R) = \mathbf{R}_{11(i)}^{-1}(\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}_R))$.

El criterio REML solo depende de $\boldsymbol{\theta}$ y σ . Sin embargo, también es útil evaluar la mejor estimación de $\boldsymbol{\beta}$ de (1.13) una vez que $\hat{\boldsymbol{\theta}}_R$ ha sido determinada usando el criterio REML.

Una diferencia importante entre la función de verosimilitud y la función de verosimilitud restringida es que la primera es invariante ante reparametrizaciones uno a uno de los efectos fijos (por ejemplo, un cambio en el contraste representado por una variable categórica), mientras que la última no. Cambiando las matrices \mathbf{X}_i resulta un cambio en $\log \text{abs}|\mathbf{R}_{00}|$ y un cambio correspondiente en $\ell_R(\boldsymbol{\theta}|\mathbf{y})$. Como una consecuencia, los modelos lineales de efectos mixtos con diferente estructura de efectos fijos apropiada usando REML no puede ser comparada sobre la base de su verosimilitud restringida. En particular, la razón de verosimilitud no es válida sobre estas circunstancias.

Niveles múltiples de efectos aleatorios.

La función de verosimilitud y la función de verosimilitud restringida para modelos multinivel con efectos mixtos puede calcularse usando las mismas técnicas descritas para el modelo de un nivel. Usando el modelo lineal de efectos mixtos, tenemos que la verosimilitud, se integra sobre ambos niveles de efectos aleatorios

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2 | \mathbf{y}) =$$

$$\prod_{i=1}^M \int \prod_{j=1}^{M_i} \left[\int p(\mathbf{y}_{ij} | \mathbf{b}_{ij}, \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) p(\mathbf{b}_{ij} | \boldsymbol{\theta}_2, \sigma^2) d\mathbf{b}_{ij} \right] p(\mathbf{b}_i | \boldsymbol{\theta}_1, \sigma^2) d\mathbf{b}_i \quad (1.16)$$

Como antes, podemos simplificar las integrales en (1.16) si aumentamos las matrices \mathbf{Z}_{ij} con $\boldsymbol{\Delta}_2$ y hacemos la descomposición ortogonal-triangular de estos arreglos aumentados. Esto nos permite evaluar las integrales interiores. Para evaluar las integrales exteriores iteramos este proceso, es decir,

$$\begin{pmatrix} \mathbf{Z}_{ij} & \mathbf{Z}_{i,j} & \mathbf{X}_{ij} & \mathbf{Y}_{ij} \\ \Delta_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} = \mathbf{Q}_{(ij)} \begin{pmatrix} \mathbf{R}_{22(ij)} & \mathbf{R}_{21(ij)} & \mathbf{R}_{20(ij)} & \mathbf{c}_{2(ij)} \\ \mathbf{0} & \mathbf{R}_{11(ij)} & \mathbf{R}_{10(ij)} & \mathbf{c}_{1(ij)} \end{pmatrix}$$

$$i = 1, \dots, M, \quad j = 1, \dots, M_i \quad (1.17)$$

La matriz $\mathbf{R}_{22(ij)}$ es una matriz triangular superior de dimensión $q_2 \times q_2$. Los otros arreglos en el primer renglón de la descomposición en (1.17) se usan sólo si el estimador condicional $\hat{\beta}(\boldsymbol{\theta})$ o el modo condicional $\hat{\mathbf{b}}_{ij}(\boldsymbol{\theta})$ y $\hat{\mathbf{b}}_i(\boldsymbol{\theta})$ es requerido. Los arreglos en el segundo renglón de (1.17): $\mathbf{R}_{11(ij)}$, $\mathbf{R}_{10(ij)}$, $\mathbf{c}_{1(ij)}$ tiene n_{ij} renglones cada uno.

Para evaluar la integral exterior en (1.16) formamos y descomponemos un arreglo aumentado:

$$\begin{pmatrix} \mathbf{R}_{11(i1)} & \mathbf{R}_{10(i1)} & \mathbf{c}_{1(i1)} \\ \vdots & \vdots & \vdots \\ \mathbf{R}_{11(iM_i)} & \mathbf{R}_{10(iM_i)} & \mathbf{c}_{1(iM_i)} \\ \Delta_1 & \mathbf{0} & \mathbf{0} \end{pmatrix} = \mathbf{Q}_{(i)} \begin{pmatrix} \mathbf{R}_{11(i)} & \mathbf{R}_{10(i)} & \mathbf{c}_{1(i)} \\ \mathbf{0} & \mathbf{R}_{00(i)} & \mathbf{c}_{0(i)} \end{pmatrix}$$

$$i = 1, \dots, M \quad (1.18)$$

La descomposición final, es como el caso de un nivel, donde de nuevo produce \mathbf{R}_{00} , \mathbf{c}_0 , \mathbf{c}_{-1} , así se obtiene la log-verosimilitud perfil para $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}_2$ como

$$\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) = \log L(\hat{\beta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \hat{\sigma}^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) | \mathbf{y}) =$$

$$\text{constante} - N \log \|\mathbf{c}_{-1}\| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta_1|}{|\mathbf{R}_{11(i)}|} \right) + \sum_{i=1}^M \sum_{j=1}^{M_i} \log \text{abs} \left(\frac{|\Delta_2|}{|\mathbf{R}_{22(ij)}|} \right)$$

Similarmente, el perfil de la log-verosimilitud restringida es

$$\begin{aligned} \ell_R(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) &= \log L_R(\hat{\boldsymbol{\beta}}_R(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \hat{\sigma}_R^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) | \mathbf{y}) = \\ & \text{constante} - (N - p) \log \|\mathbf{c}_{-1}\| - \log \text{abs} |\mathbf{R}_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta_1|}{|\mathbf{R}_{11(i)}|} \right) + \\ & \sum_{i=1}^M \sum_{j=1}^{M_i} \log \text{abs} \left(\frac{|\Delta_2|}{|\mathbf{R}_{22(ij)}|} \right) \end{aligned}$$

En general Q se anida en los efectos aleatorios.

1.3.4. Valores fijos y predictivos.

Los valores fijos, son valores que se predicen de las respuestas observadas sobre los modelos fijos, frecuentemente son de interés para verificar el modelo. Los valores predictivos para nuevas observaciones son una de las primeras cantidades de interés cuando hacemos inferencia de un modelo fijo.

En un modelo de efectos mixtos, los valores fijos y predictivos podrían obtenerse de diferentes niveles de anidamiento, o de todo el nivel poblacional. Las predicciones del nivel poblacional estiman el valor de la esperanza marginal de la respuesta. Por ejemplo, sea \mathbf{x}_h un vector de covariables de efectos fijos, el valor de la esperanza marginal del correspondiente vector respuesta y_h es

$$E(y_h) = \mathbf{x}_h^T \boldsymbol{\beta}$$

y los mejores predictores lineales insesgados (*BLUP*) del valor esperado poblacional, se obtienen sustituyendo el valor del estimador $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ en el valor esperado $\boldsymbol{\beta}$.

Los valores predictivos del k -ésimo nivel de anidamiento estiman la esperanza condicional de la respuesta, dados los efectos aleatorios de todos los niveles menores o iguales al k -ésimo nivel. Por ejemplo, $\mathbf{z}_{h(i)}$ denota un vector de covariables correspondientes a los efectos aleatorios asociados con el i -ésimo grupo del primer nivel de anidamiento, las predicciones del nivel-1 estiman

$$E(y_{h(i)}|\mathbf{b}_i) = \mathbf{x}_h^T \boldsymbol{\beta} + \mathbf{z}_{h(i)} \mathbf{b}_i$$

y su correspondiente *BLUP* para el valor esperado condicional dados los efectos aleatorios \mathbf{b}_i es

$$\hat{y}_{h(i)} = \mathbf{x}_h^T \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \mathbf{z}_{h(i)} \hat{\mathbf{b}}_i(\boldsymbol{\theta})$$

Similarmente, sea $\mathbf{z}_{h(i,j)}$ el vector de covariables asociado con el j -ésimo grupo del nivel-2 dentro del i -ésimo grupo del nivel-1, los valores predictivos del nivel-2 estiman

$$E(y_{h(i)}|\mathbf{b}_i, \mathbf{b}_{ij}) = \mathbf{x}_h^T \boldsymbol{\beta} + \mathbf{z}_{h(i)}^T \mathbf{b}_i + \mathbf{z}_{h(i,j)}^T \mathbf{b}_{ij}$$

y su *BLUP* correspondiente es

$$\hat{y}_{h(i,j)} = \mathbf{x}_h^T \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \mathbf{z}_{h(i)}^T \hat{\mathbf{b}}_i(\boldsymbol{\theta}) + \mathbf{z}_{h(i,j)}^T \hat{\mathbf{b}}_{ij}(\boldsymbol{\theta})$$

donde en todos los casos $\boldsymbol{\theta}$ es reemplazado por su EMV o su estimador REML, produciendo el estimador *BLUP* de los valores esperados.

1.4. MODELOS DE EFECTOS MIXTOS NO LINEALES (NLME).

Los modelos de efectos mixtos no lineales, son modelos de efectos mixtos en los cuales algunos o todos sus efectos fijos o aleatorios ocurren de manera no lineal en la función del modelo, en el que la esperanza condicional de la respuesta dado el efecto aleatorio es también una función no lineal de coeficientes. También pueden ser una extensión de los modelos de regresión no lineal para datos independientes, en los que los efectos aleatorios son incluidos en el coeficiente dentro del grupo, lo que induce correlación dentro de los grupos.

1.4.1. Modelos NLME con un nivel de agrupación.

Comunmente, las aplicaciones de los modelos NLME se dan para datos con medidas repetidas (propuestos por Lindstrom y Bates (1990) que se pueden pensar como un modelo jerárquico. Así, un nivel de j observaciones sobre el i -ésimo grupo es modelado como

$$y_{ij} = f(\boldsymbol{\phi}_{ij}, \boldsymbol{\nu}_{ij}) + \epsilon_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, n_i$$

donde M es el número de grupos, n_i el número de observaciones dentro del i -ésimo grupo, f es cualquier función real, diferenciable de un vector paramétrico $\boldsymbol{\phi}_{ij}$ específico del grupo y un vector de covariables $\boldsymbol{\nu}_{ij}$; y ϵ_{ij} es el vector de errores dentro del i -ésimo grupo con distribución normal. La función f no es lineal en al menos una componente del vector paramétrico $\boldsymbol{\phi}_{ij}$ específico del grupo, el cual es modelado como

$$\boldsymbol{\phi}_{ij} = \mathbf{A}_{ij}\boldsymbol{\beta} + \mathbf{B}_{ij}\mathbf{b}_i, \quad \mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}) \quad (1.19)$$

donde $\boldsymbol{\beta}$ es un vector p -dimensional de efectos fijos, \mathbf{b}_i es un vector de efectos aleatorios q -dimensional asociado con el i -ésimo grupo (sin variar j) con matriz de varianza-covarianza $\boldsymbol{\Psi}$. Las matrices \mathbf{A}_{ij} y \mathbf{B}_{ij} son de dimensión apropiada y dependen del grupo y posiblemente de valores de algunas covariables de la j -ésima observación, lo cual permite incorporar covariables de *tiempo de variación* en los efectos fijos o los efectos aleatorios para el modelo. Esto supone que observaciones correspondientes a diferentes grupos son independientes y que los errores dentro de grupos ϵ_{ij} son independientes y distribuidos normal con media cero y varianza σ^2 , e independientes de \mathbf{b}_i . La suposición de independencia y homoscedasticidad para los errores dentro de grupos se puede relajar.

Como f es cualquier función no lineal de ϕ_{ij} , la representación de los coeficientes ϕ_{ij} específicos del grupo se pueden elegir de tal modo que \mathbf{A}_{ij} y \mathbf{B}_{ij} sean siempre matrices de incidencia.

En forma matricial, el modelo está definido como

$$\mathbf{y}_i = \mathbf{f}_i(\boldsymbol{\phi}_{ij}, \boldsymbol{\nu}_{ij}) + \boldsymbol{\epsilon}_i, \quad (1.20)$$

$$\boldsymbol{\phi}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \quad \text{para } i = 1, \dots, M$$

donde

$$\begin{aligned} \mathbf{y}_i &= \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, & \boldsymbol{\phi}_i &= \begin{pmatrix} \phi_{i1} \\ \vdots \\ \phi_{in_i} \end{pmatrix}, & \boldsymbol{\epsilon}_i &= \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}, & \mathbf{f}_i(\boldsymbol{\phi}_i, \boldsymbol{\nu}_i) &= \begin{pmatrix} f(\phi_{i1}, \nu_{i1}) \\ \vdots \\ f(\phi_{in_i}, \nu_{in_i}) \end{pmatrix} \\ \boldsymbol{\nu}_i &= \begin{pmatrix} \nu_{i1} \\ \vdots \\ \nu_{in_i} \end{pmatrix}, & \mathbf{A}_i &= \begin{pmatrix} \mathbf{A}_{i1} \\ \vdots \\ \mathbf{A}_{in_i} \end{pmatrix}, & \mathbf{B}_i &= \begin{pmatrix} \mathbf{B}_{i1} \\ \vdots \\ \mathbf{B}_{in_i} \end{pmatrix} \end{aligned} \quad (1.21)$$

El modelo anterior, se puede extender a datos agrupados múltiples o factores anidados modificando el modelo para efectos aleatorios.

1.4.2. Estimación por verosimilitud.

Como los efectos aleatorios son cantidades no observadas, la estimación por máxima verosimilitud en modelos de efectos mixtos se basa en la densidad marginal de la respuesta y , la cual, para un modelo con Q niveles de anidamiento, se calcula como

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_Q) = \int p(\mathbf{y} \mid \mathbf{b}, \boldsymbol{\beta}, \sigma^2) p(\mathbf{b} \mid \boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_Q) d\mathbf{b}$$

donde $p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_Q)$ es la densidad marginal de \mathbf{y} , $p(\mathbf{y} \mid \mathbf{b}, \boldsymbol{\beta}, \sigma^2)$ es la densidad condicional de \mathbf{y} dado los efectos aleatorios \mathbf{b} , y la distribución marginal de \mathbf{b} es $p(\mathbf{b} \mid \boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_Q)$. Para el modelo NLME (1.20), expresando la matriz de varianza-covarianza de efectos aleatorios en términos del factor de precisión $\boldsymbol{\Delta}$, tal que $\boldsymbol{\Psi}^{-1} = \sigma^{-2} \boldsymbol{\Delta}^T \boldsymbol{\Delta}$, se obtiene la densidad marginal de \mathbf{y} como

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}) =$$

$$\frac{|\boldsymbol{\Delta}|^M}{(2\pi\sigma^2)^{(N+Mq)/2}} \prod_{i=1}^M \int \exp \left\{ \frac{\|\mathbf{y}_i - \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i)\|^2 + \|\boldsymbol{\Delta} \mathbf{b}_i\|^2}{-2\sigma^2} \right\} d\mathbf{b}_i \quad (1.22)$$

donde $\mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i) = \mathbf{f}_i[\boldsymbol{\phi}_i(\boldsymbol{\beta}, \mathbf{b}_i), \boldsymbol{\nu}_i]$.

Como la forma de la función f puede ser no lineal en los efectos aleatorios, la integral en (1.22) generalmente no tiene una forma cerrada. Esta integral algunas veces se resuelve por aproximación de Taylor, tomando la expansión

de primer orden de la función f alrededor del valor esperado de los efectos aleatorios, o alrededor del condicional (sobre Δ) de los efectos aleatorios.

Capítulo 2

MODELOS LINEALES MIXTOS GENERALIZADOS (GLMM).

Los modelos lineales mixtos generalizados logran unir las principales ventajas de los modelos lineales mixtos y los modelos lineales generalizados: una de ellas, que los datos no necesariamente se suponen normalmente distribuidos o que la media no necesariamente es una combinación lineal de los parámetros, pero si lo es alguna función de la media. Y la otra, es la aleatoriedad de los efectos mixtos, lo que permite particularizar a los individuos asociando una relación de varianza-covarianza. Así, el modelo mixto condiciona los efectos fijos con los aleatorios dando una asociación entre grupos o niveles de anidamiento, permitiendo que el modelo sea estudiado para dar un comportamiento distinto a los individuos, otorgando distinta variabilidad a individuos de distintos grupos, por un lado o dando distinta variabilidad a individuos del mismo grupo por otro. Es decir, mientras el modelo marginal solo da resultados a nivel promedio, el modelo condicional (mixto generali-

zado), permite dar resultados por individuo.

2.1. LA ESTRUCTURA DEL MODELO.

Sea y el vector respuesta que por lo general se considera de elementos independientes (condicionados en los efectos aleatorios), cada uno con una distribución con densidad perteneciente a la familia exponencial

$$y_i | b_i \sim f_{Y_i|b_i}(y_i | \mathbf{b})$$

$$f_{Y_i|b_i}(y_i | b_i) = \exp\{[y_i\theta_i - d(\theta_i)]/a(\phi) - c(y_i, \phi)\} \quad (2.1)$$

donde la media condicional de y_i está relacionada con θ_i vía la identidad $\mu_i = \frac{\partial d(\theta_i)}{\partial \theta_i}$. Esta transformación da el sentido lineal al modelo en ambos efectos, el fijo y el aleatorio, es decir,

$$E(y_i | b_i) = \mu_i$$

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b} \quad (2.2)$$

donde g es una función conocida, llamada la *función liga*, \mathbf{x}_i^T es el i -ésimo renglón de la matriz de observaciones para los efectos fijos, y $\boldsymbol{\beta}$ es el vector paramétrico de los efectos fijos. Además, tenemos que \mathbf{z}_i^T es el i -ésimo renglón de la matriz de efectos aleatorios. μ_i denota la media condicional de \mathbf{y} dado \mathbf{b} , en este caso, no es la media marginal. Falta, sólo asignar la distribución del vector de efectos aleatorios \mathbf{b} :

$$\mathbf{b} \sim f_B(\mathbf{b}) \quad (2.3)$$

Se nota que el hecho de la distribución condicional de \mathbf{y} dado \mathbf{b} es justamente la extensión del modelo lineal generalizado, pues μ_i representa la media condicional más que la media marginal, todo lo demás es similar al modelo lineal generalizado, por lo que muchas de las relaciones se derivan de los modelos lineales generalizados.

La varianza condicional de \mathbf{y} dado \mathbf{b} está dada por

$$\text{var}(y_i | b_i) = a(\phi)d''(\theta)$$

la cual tiene una dependencia con la media condicional μ_i .

2.1.1. Consecuencias de tener efectos aleatorios.

La distribución marginal contra la distribución condicional. El modelo lineal mixto generalizado se condiciona respecto a \mathbf{b} , por lo que es necesario derivar aspectos de la distribución marginal de \mathbf{y} para entender lo que ocurre con las suposiciones hechas a los datos observados.

La media de y puede derivarse fácilmente por la media condicional.

$$E(y_i) = E(E(y_i | \mathbf{b})) = E(\mu_i) = E(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b})).$$

lo cual es fácil de simplificar si g^{-1} es una función lineal.

Por ejemplo, si $g(\mu) = \log \mu$ y $g^{-1}(x) = \exp\{x\}$. Entonces

$$\begin{aligned} E(y_i) &= E(\exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}\}) = \\ &= \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\} E(\exp\{\mathbf{z}_i^T \mathbf{b}\}) = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\} M_{\mathbf{b}}(\mathbf{z}_i) \end{aligned}$$

donde $M_{\mathbf{b}}(\mathbf{z}_i)$ es la función generadora de momentos de \mathbf{b} evaluada en \mathbf{z}_i .

Si $b_i \sim N(0, \sigma_b^2)$ y \mathbf{Z} tiene un 1 en cada renglón y cero en los demás elementos, entonces $M_{\mathbf{b}}(\mathbf{z}_i) = \exp\{\sigma_b^2/2\}$ y

$$E(y_i) = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\} \exp\{\sigma_b^2/2\} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \sigma_b^2/2\}$$

o

$$\log E(y_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma_b^2/2$$

La varianza de y esta dada por:

$$\text{var}(y_i) = \text{var}(E(y_i | \mathbf{b})) + E(\text{var}(y_i | \mathbf{b})) = \text{var}(\mu_i) + E(a(\phi)v(\mu_i)) =$$

$$\text{var}(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b})) + E(a(\phi)v(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b})))$$

donde se necesita conocer la forma de g o la distribución condicional de \mathbf{y} para poder simplificar esta última expresión.

Así, por ejemplo, tenemos la función log liga y suponemos que todos los elementos de \mathbf{y} dado \mathbf{b} , son independientes con distribución Poisson. La varianza condicional de y_i dado \mathbf{b} es $a(\phi)v(\mu_i) = \mu_i$, por lo tanto

$$\text{var}(y_i) = \text{var}(\mu_i) + E(\mu_i) = \text{var}(\exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}\}) + E(\exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}\}) =$$

$$E(\exp\{2(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b})\}) - [E(\exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}\})]^2 + E(\exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}\}) =$$

$$\exp\{2\mathbf{x}_i^T\boldsymbol{\beta}\} \left(M_{\mathbf{b}}(2\mathbf{z}_i) - [M_{\mathbf{b}}(\mathbf{z}_i)]^2 + \exp\{-\mathbf{x}_i^T\boldsymbol{\beta}\} M_{\mathbf{b}}(\mathbf{z}_i) \right)$$

Si de nuevo, $b_i \sim N(0, \sigma_b^2)$ y \mathbf{Z} tiene la misma forma que antes, entonces

$$\text{var}(y_i) = \exp\{2\mathbf{x}_i^T\boldsymbol{\beta}\} \left(\exp\{2\sigma_b^2\} - \exp\{\sigma_b^2\} \right) + \exp\{\mathbf{x}_i^T\boldsymbol{\beta}\} \exp\{\sigma_b^2/2\} =$$

$$\exp\{\mathbf{x}_i^T\boldsymbol{\beta} + \sigma_b^2/2\} \left(\exp\{\mathbf{x}_i^T\boldsymbol{\beta}\} [\exp\{3\sigma_b^2/2\} - \exp\{\sigma_b^2/2\}] + 1 \right) =$$

$$E(y_i) \left(\exp\{\mathbf{x}_i^T\boldsymbol{\beta}\} [\exp\{3\sigma_b^2/2\} - \exp\{\sigma_b^2/2\}] + 1 \right) \quad (2.4)$$

donde el término en corchetes en (2.4) es mayor que 1 y solo sería cero si $\sigma_b^2 = 0$, lo que hace a la varianza más grande que la media. Por lo tanto, aunque la distribución condicional de y_i dado \mathbf{b} es Poisson, la distribución marginal de y_i no lo es. En efecto, bajo estas suposiciones siempre se tiene sobredispersión comparado con la distribución Poisson. En este sentido se piensa que los efectos aleatorios son además una manera para modelar o atribuir sobredispersión en una fuente particular. Es decir, los efectos aleatorios surgen del diseño, la muestra viene en conglomerados o grupos (escuelas, individuos, laboratorios, etc.), que se consideran aleatorios o son una muestra aleatoria. Matemáticamente equivale a sobredispersión.

Covarianza y correlación. El uso de efectos aleatorios introduce una correlación entre las observaciones, las cuales tienen cualquier efecto aleatorio en común. Lo mismo ocurre en los modelos lineales mixtos generalizados. Suponiendo independencia lineal de los elementos de \mathbf{y} tenemos:

$$\text{cov}(y_i, y_j) = \text{cov}(E(y_i | \mathbf{b}), E(y_j | \mathbf{b})) + E[\text{cov}(y_i, y_j | \mathbf{b})] =$$

$$\text{cov}(\mu_i, \mu_j) + E(0) = \text{cov}(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}), g^{-1}(\mathbf{x}_j^T \boldsymbol{\beta} + \mathbf{z}_j^T \mathbf{b}))$$

Si, por ejemplo, tenemos una función logaritmo como liga, esto se evalúa como

$$\text{cov}(y_i, y_j) = \text{cov}(\exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}\}, \exp\{\mathbf{x}_j^T \boldsymbol{\beta} + \mathbf{z}_j^T \mathbf{b}\}) =$$

$$\exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_j^T \boldsymbol{\beta}\} \text{cov}(\exp\{\mathbf{z}_i^T \mathbf{b}\}, \exp\{\mathbf{z}_j^T \mathbf{b}\}) =$$

$$\exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_j^T \boldsymbol{\beta}\} \text{cov}(M_{\mathbf{b}}(\mathbf{z}_i + \mathbf{z}_j) - M_{\mathbf{b}}(\mathbf{z}_i)M_{\mathbf{b}}(\mathbf{z}_j))$$

de nuevo, si $\mathbf{b} \sim N(0, \mathbf{I}\sigma_b^2)$ y \mathbf{Z} igual que antes, entonces

$$\text{cov}(y_i, y_j) = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_j^T \boldsymbol{\beta}\} \left[\exp\{\sigma_b^2\} (\exp\{\mathbf{z}_i^T \mathbf{z}_j \sigma_b^2\} - 1) \right] \quad (2.5)$$

lo cual es igual a cero si $\mathbf{z}_i^T \mathbf{z}_j = 0$, lo que significa que las dos observaciones no forman parte de los efectos aleatorios. Y es positiva si, $\mathbf{z}_i^T \mathbf{z}_j = 1$.

De (2.4) y (2.5), cuando $\mathbf{z}_i^T \mathbf{z}_j = 1$ se puede calcular la correlación como

$$\text{corr}(y_i, y_j) = \frac{e^{2\sigma_b^2} - e^{\sigma_b^2}}{\sqrt{(e^{2\sigma_b^2} - e^{\sigma_b^2} + e^{-\mathbf{x}_i^T \boldsymbol{\beta} + \sigma_b^2/2})(e^{2\sigma_b^2} - e^{\sigma_b^2} + e^{-\mathbf{x}_j^T \boldsymbol{\beta} + \sigma_b^2/2})}}$$

$$= \frac{1}{\sqrt{(1 + \eta e^{-\mathbf{x}_i^T \boldsymbol{\beta}})(1 + \eta e^{-\mathbf{x}_j^T \boldsymbol{\beta}})}}$$

donde

$$\eta = \frac{1}{e^{3\sigma_b^2/2} - e^{\sigma_b^2/2}}.$$

2.2. MÉTODOS DE ESTIMACIÓN.

2.2.1. Estimación por máxima verosimilitud.

De (2.1), (2.2) y (2.3) se puede escribir la función de verosimilitud como

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = \int \prod_i f_{Y_i|\mathbf{b}}(y_i | \mathbf{b}) f_{\mathbf{B}}(\mathbf{b}) d\mathbf{b} \quad (2.6)$$

donde se integra sobre la distribución q -dimensional de \mathbf{b} .

Para los efectos fijos, las ecuaciones de verosimilitud son difíciles numericamente, pero de (2.6) se pueden escribir de manera muy general como

$$\ell = \log L = \log \int f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b}) f_{\mathbf{B}}(\mathbf{b}) d\mathbf{b} = \log f_{\mathbf{Y}}(\mathbf{y})$$

así

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} \int f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b}) f_{\mathbf{B}}(\mathbf{b}) d\mathbf{b} / f_{\mathbf{Y}}(\mathbf{y}) =$$

$$\int \left[\frac{\partial}{\partial \boldsymbol{\beta}} f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b}) \right] f_{\mathbf{B}}(\mathbf{b}) d\mathbf{b} / f_{\mathbf{Y}}(\mathbf{y}),$$

donde $f_{\mathbf{B}}(\mathbf{b})$ no contiene a β . Además

$$\frac{\partial}{\partial \beta} f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b}) = \left(\frac{1}{f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b})} \frac{\partial f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b})}{\partial \beta} \right) f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b}) =$$

$$\frac{\partial \log f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b})}{\partial \beta} f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b})$$

entonces

$$\frac{\partial \ell}{\partial \beta} = \int \frac{\partial \log f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b})}{\partial \beta} f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b}) f_{\mathbf{B}}(\mathbf{b}) d\mathbf{b} / f_{\mathbf{Y}}(\mathbf{y}) =$$

$$\int \frac{\partial \log f_{\mathbf{Y}|\mathbf{b}}(\mathbf{y} | \mathbf{b})}{\partial \beta} f_{\mathbf{B}|\mathbf{y}}(\mathbf{b} | \mathbf{y}) d\mathbf{b} \quad (2.7)$$

Usando (2.6), la cual da la derivada del log-verosimil para GLM, en (2.7) tenemos

$$\frac{\partial \ell}{\partial \beta} = \int \mathbf{X}^T \mathbf{W}^* (\mathbf{y} - \boldsymbol{\mu}) f_{\mathbf{B}|\mathbf{y}}(\mathbf{b} | \mathbf{y}) d\mathbf{b} =$$

$$\mathbf{X}^T E[\mathbf{W}^* | \mathbf{y}] - \mathbf{X}^T E[\mathbf{W}^* \boldsymbol{\mu} | \mathbf{y}],$$

donde $\mathbf{W}^* = \text{diag}\{[a(\phi)v(\mu_i)g_\mu(\mu_i)]^{-1}\}$

La ecuación de verosimilitud para β es, por lo tanto

$$\mathbf{X}^T E[\mathbf{W}^* | \mathbf{y}] = \mathbf{X}^T E[\mathbf{W}^* \boldsymbol{\mu} | \mathbf{y}]$$

lo cual es similar al caso GLM, solo que W^* y $W^* \mu$ son los valores esperados condicionales dado y .¹

¹Para cualquier duda se puede consultar la referencia [3].

Para los parámetros de los efectos aleatorios, consideremos φ como los parámetros de los efectos aleatorios, tal que

$$\frac{\partial \ell}{\partial \varphi} = \int \frac{\partial \log f_{\mathbf{B}}(\mathbf{b})}{\partial \varphi} f_{\mathbf{B}|\mathbf{y}}(\mathbf{b} | \mathbf{y}) d\mathbf{b} = E \left[\frac{\partial \log f_{\mathbf{B}}(\mathbf{b})}{\partial \varphi} | \mathbf{y} \right]$$

lo cual ya no es posible simplificar, a menos que se especifique la forma para la distribución de los efectos aleatorios.

2.2.2. Quasi-verosimilitud Penalizada (PQL).

Supongamos que b tiene una distribución normal multivariada con media 0 y matriz de covarianza $\Sigma = \Sigma(\varphi)$, que depende de un vector desconocido φ de componentes de varianza. Por lo general, la distribución condicional de y_i es binomial, Poisson o hipergeométrica, aunque también se puede estimar junto con φ como un parámetro en la matriz de covarianzas de la distribución marginal de \mathbf{y} .

La integral de la función quasi-verosimilitud (ql) que se usa para estimar β y θ está definida como

$$e^{ql(\beta, \theta)} \propto |\Sigma|^{-1/2} \int \exp \left[-\frac{1}{2\phi} \sum_{i=1}^n d_i(y_i, \mu_i) - \frac{1}{2} \mathbf{b}^T \Sigma^{-1} \mathbf{b} \right] d\mathbf{b} \quad (2.8)$$

donde $d_i(y, \mu_i) = -2 \int_y^{\mu_i} \frac{y-u}{a_i v(u)} du$ denota la devianza para el i -ésimo individuo.²

Condicionando sobre \mathbf{b} , las observaciones son extraídas de la familia exponencial con función de varianza $v(\cdot)$ por lo que la devianza es conocida como $2\phi\{\ell(\mathbf{y}; \mathbf{y}, \phi) - \ell(\mathbf{y}; \mu, \phi)\}$ donde $\ell(\mathbf{y}; \mu, \phi)$ es la verosimilitud condicional de \mathbf{y}

²Consultar la referencia [6].

dada su media $\boldsymbol{\mu}$. En este caso $ql(\boldsymbol{\beta}, \boldsymbol{\theta})$ representa la función log-verosimilitud verdadera de los datos.

Escribiendo (2.8) como $c|\boldsymbol{\Sigma}|^{-1/2} \int e^{-k(\mathbf{b})} d\mathbf{b}$, y aplicando el método de Laplace a ésta integral, entonces definiendo k' como el vector q -dimensional de las primeras derivadas parciales de k con respecto a b y k'' como la matriz de $q \times q$ de las segundas derivadas parciales de k con respecto a b , tenemos que:

$$ql(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \log |k''(\tilde{\mathbf{b}})| - k(\tilde{\mathbf{b}})$$

y donde $\tilde{\mathbf{b}} = \tilde{\mathbf{b}}(\boldsymbol{\beta}, \boldsymbol{\theta})$ denota la solución de

$$k'(\mathbf{b}) = -\sum_{i=1}^n \frac{(y_i - \mu_i) \mathbf{z}_i}{a_i(\phi) v(\mu_i) g'(\mu_i)} + \boldsymbol{\Sigma}^{-1} \mathbf{b} = 0$$

lo cual maximiza a $k(b)$, y

$$\begin{aligned} k''(\mathbf{b}) &= \sum_{i=1}^n \frac{\mathbf{z}_i \mathbf{z}_i^T}{a_i(\phi) v(\mu_i) [g_\mu(\mu_i)]^2} + \boldsymbol{\Sigma}^{-1} + \mathbf{R} \\ &\approx \mathbf{Z}^T \mathbf{W} \mathbf{Z} + \boldsymbol{\Sigma}^{-1}, \end{aligned} \quad (2.9)$$

donde $\mathbf{W} = \text{diag}\{a_i(\phi) v(\mu_i) [g_\mu(\mu_i)]^2\}^{-1}$ es como antes la matriz diagonal de dimensión $n \times n$.

$$\mathbf{R} = -\sum_{i=1}^n (y_i - \mu_i) \mathbf{z}_i \frac{\partial}{\partial \mathbf{b}} \left[\frac{1}{a_i(\phi) v(\mu_i) g_\mu(\mu_i)} \right]$$

tiene media 0 y probabilísticamente es una función de orden n , menor que (2.9). Note que $\mathbf{R} = 0$ cumple con la función liga canónica, pues $g_\mu(\mu) = v^{-1}(\mu)$. Combinando (2.8) y (2.9) e ignorando \mathbf{R} , tenemos

$$ql(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\mathbf{I} + \mathbf{Z}^T \mathbf{W} \mathbf{Z} \boldsymbol{\Sigma}| - \frac{1}{2\phi} \sum_{i=1}^n d_i(y_i, \mu_i) - \frac{1}{2} \tilde{\mathbf{b}}^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{b}}, \quad (2.10)$$

donde $\tilde{\mathbf{b}}$ maximiza la suma de los dos últimos términos.

Suponiendo que las iteraciones ponderadas de GLM varían lentamente como una función de la media, e ignorando el primer término de la ecuación (2.10), elegimos a $\boldsymbol{\beta}$ tal que maximice el segundo término. Por lo tanto, $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}) = (\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\mathbf{b}}(\boldsymbol{\theta}))$, donde $\hat{\mathbf{b}}(\boldsymbol{\theta}) = \tilde{\mathbf{b}}(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))$, maximiza conjuntamente a PQL

$$-\frac{1}{2\phi} \sum_{i=1}^n d_i(y_i, \mu_i) - \frac{1}{2} \mathbf{b}^T \boldsymbol{\Sigma}^{-1} \mathbf{b} \quad (2.11)$$

Diferenciando con respecto a $\boldsymbol{\beta}$ y \mathbf{b} se llega a las ecuaciones de puntaje (*score*) para los parámetros de la media:

$$\sum_{i=1}^n \frac{(y_i - \mu_i) \mathbf{x}_i}{a_i(\phi) v(\mu_i) g_\mu(\mu_i)} = 0 \quad (2.12)$$

y

$$\sum_{i=1}^n \frac{(y_i - \mu_i) \mathbf{z}_i}{a_i(\phi) v(\mu_i) g_\mu(\mu_i)} = \boldsymbol{\Sigma}^{-1} \mathbf{b} \quad (2.13)$$

Definimos el vector trabajo \mathbf{Y} como $Y_i = \eta_i + (y_i - \mu_i) g_\mu(\mu_i)$, la solución de (2.12) y (2.13) vía el método de puntajes de Fisher se expresa como la solución iterada al sistema

$$\begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{Z} \boldsymbol{\Sigma} \\ \mathbf{Z}^T \mathbf{W} \mathbf{X} & \mathbf{I} + \mathbf{Z}^T \mathbf{W} \mathbf{Z} \boldsymbol{\Sigma} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{Y} \\ \mathbf{Z}^T \mathbf{W} \mathbf{Y} \end{pmatrix}$$

donde $\mathbf{b} = \boldsymbol{\Sigma} \boldsymbol{\nu}$. Resolviendo para $\boldsymbol{\beta}$ en

$$(\mathbf{X}^T \mathbf{v}^{-1} \mathbf{X})\boldsymbol{\beta} = \mathbf{X}^T \mathbf{v}^{-1} \mathbf{y},$$

donde $\mathbf{V} = \mathbf{W}^{-1} + \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}^T$, entonces

$$\tilde{\mathbf{b}} = \boldsymbol{\Sigma}\tilde{\mathbf{v}} = \boldsymbol{\Sigma}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \quad (2.14)$$

lo cual sugiere que la aproximación para la covarianza de $\hat{\boldsymbol{\beta}}$ es la matriz $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$, esto es exacto teóricamente para el modelo lineal cuando $\boldsymbol{\theta}$ es conocido. El error estándar para $\hat{\mathbf{b}}$ puede calcularse de (2.14), pero se ignora la variabilidad adicional dada al estimar $\boldsymbol{\theta}$.

2.2.3. Estimación de las componentes de varianza para PQL.

Sustituyendo el valor máximo de (2.10) en (2.11) y \mathbf{W} en $(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\mathbf{b}}(\boldsymbol{\theta}))$ se genera una aproximación para la función quasi-verosimilitud perfil para inferir sobre $\boldsymbol{\theta}$. Además, aproximando la estimación del vector \mathbf{y} , la matriz ponderada \mathbf{W} , las matrices de diseño \mathbf{X} y \mathbf{Z} . También ignorando que \mathbf{W} depende de $\boldsymbol{\theta}$ y reemplazando la devianza $\sum d_i(y_i, \mu_i)$ por la ji-cuadrada de Pearson $\sum (y_i - \mu_i)^2/[a_i v(\mu_i)]$, tenemos que

$$q\ell(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (2.15)$$

Tomando los grados de libertad de tal modo que $\hat{\boldsymbol{\beta}}$ sea considerado en la forma cuadrática de (2.15), entonces

$$q\ell_1(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (2.16)$$

esto corresponde a la verosimilitud perfil conjugada de Cox y Reid (1987), y tiene las siguientes propiedades:

- a) β y θ son parámetros ortogonales y
- b) la matriz de información para $\hat{\beta}(\theta)$ es $\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}$.

Definiendo $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$ y diferenciando (2.16) con respecto a las componentes de θ se obtienen las ecuaciones de estimación para los parámetros de varianza:

$$-\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_j} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) - \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_j} \right) \right] = 0 \quad (2.17)$$

La matriz de información de Fisher \mathbf{J} , tiene componentes

$$J_{jk} = -\frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_j} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_k} \right) \quad (2.18)$$

Nótese que como se ignora la dependencia de $\mathbf{W} = \text{diag}\{a_i(\phi)v(\mu_i)[g_\mu(\mu_i)]^2\}^{-1}$ con θ , calcular $\frac{\partial \mathbf{V}}{\partial \theta_j}$ en (2.17) y (2.18) de $q\ell_1$ no se puede considerar como función objetivo para resolver las ecuaciones.

Observaciones sobre la teoría asintótica.

La derivación de la quasi-verosimilitud penalizada en (2.11) y la modificación a la quasi-verosimilitud perfil (2.16) tienen varios ajustes y aproximaciones, por las cuales no es posible dar una buena justificación, más bien, es una motivación para aproximar las ecuaciones (2.12),(2.13) y (2.17). Primero las ecuaciones son ecuaciones REML bajo el modelo lineal teórico, en donde los vectores \mathbf{y} observados coinciden y $\mathbf{W} = \mathbf{I}$. Al menos, \mathbf{y} puede ser particionada en K componentes independientes donde K crece con n , lo cual

no garantiza que la teoría asintótica estándar se aplique. Por otro lado, los argumentos involucran aproximaciones de incrementos en la devianza. Esto ocurre, por ejemplo, cuando el denominador de proporciones binomiales, o la media de observaciones Poisson crecen.

2.2.4. Quasi-verosimilitud Marginal.

El modelo marginal.

La estructura de regresión dentro del modelo (2.1) es condicional sobre los valores de los efectos aleatorios \mathbf{b} . Cuando los efectos aleatorios se desean estimar para cada persona, esto significa que $\boldsymbol{\beta}$ representa efectos covariables del nivel del individuo. Como generalmente se desean estimadores de efectos de covariables sobre promedios, es más apropiado especificar los GLM en términos de la media marginal como ³

$$E(y_i) = \mu_i = h(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (2.19)$$

Sin embargo, la media marginal definida así generalmente no coincide con la media marginal calculada a partir de la media condicional de $y_i \mid b_i$.

Aunque se podría pensar de (2.19) como una forma derivada de una primera aproximación del modelo GLM, la cual es válida en el límite como componente de dispersión que se aproxima a cero. Escribiendo el modelo como

$$y_i = \mu_i + \epsilon_i \quad \text{con} \quad \text{var}(y_i) = a_i(\phi)v(\mu_i), \quad \mathbf{b} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$$

de donde se obtendría

³Consultar la referencia [6].

$$y_i \approx h(\mathbf{x}_i^T \boldsymbol{\beta}) + h'(\mathbf{x}_i^T \boldsymbol{\beta}) z_i^T \mathbf{b} + \epsilon_i,$$

donde $\mathbf{V}_0 = \text{diag}\{a_i(\phi)v(\mu_i)\}$ y $\boldsymbol{\Delta} = \text{diag}\{g_\mu(\mu_i)\}$, la primera aproximación de la varianza correspondiente es

$$\text{var}(\mathbf{y}) = \mathbf{V}_0 + \boldsymbol{\Delta}^{-1} \mathbf{Z} \boldsymbol{\Sigma} \mathbf{Z}^T \boldsymbol{\Delta}^{-1} \quad (2.20)$$

Estos modelos marginales fueron investigados para diseños de datos longitudinales por Zeger (1988). Mostró que la media marginal del modelo (2.1) con distribución normal en los efectos aleatorios se pueden expresar casi siempre como en la forma (2.19), al menos aproximadamente, pero con valores alternados para las variables o coeficientes de regresión.

Estimación de efectos fijos y efectos aleatorios.

Para $\boldsymbol{\theta}$ fijo, estimamos los coeficientes de regresión $\boldsymbol{\beta}$ en el modelo marginal usando las ecuaciones de quasi-verosimilitud apropiadas para resultados dependientes. Denotando el vector de la media marginal $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$, la ecuación de estimación generalizada (GEE)

$$U(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}^T} \text{var}^{-1}(\mathbf{y})(\mathbf{y} - \boldsymbol{\mu}) = 0$$

toma la forma

$$\mathbf{X}^T (\boldsymbol{\Delta} \mathbf{V}_0 \boldsymbol{\Delta} + \mathbf{Z} \boldsymbol{\Sigma} \mathbf{Z}^T)^{-1} \boldsymbol{\Delta} (\mathbf{y} - \boldsymbol{\mu}) = 0$$

Sea $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ denota el vector de predictores lineales $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$. La función de puntaje de Fisher conduce a la regresión IWLS del vector respuesta $\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\Delta}(\mathbf{y} - \boldsymbol{\mu})$ sobre \mathbf{X} con matriz ponderadora

$$\mathbf{V}^{-1} = (\Delta \mathbf{V}_0 \Delta + \mathbf{Z} \Sigma \mathbf{Z}^T)^{-1} = (\mathbf{W}^{-1} + \mathbf{Z} \Sigma \mathbf{Z}^T)^{-1},$$

donde $\mathbf{W} = \text{diag}\{a_i(\phi)v(\mu_i)[g_\mu(\mu_i)]^2\}^{-1}$ como en GLM. La iteración se hace equivalente al estimar $\boldsymbol{\beta}$ en la teoría normal asociada al modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \quad \text{con} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{W}^{-1}) \quad \text{y} \quad \mathbf{b} \sim N(\mathbf{0}, \Sigma).$$

La diferencia esencial entre las ecuaciones estimadas MQL para el modelo marginal y las ecuaciones PQL del modelo condicional es que estos incorporan el término de efectos aleatorios $\mathbf{z}_i^T \mathbf{b}$ en el predictor lineal.

2.3. MODELO MARGINAL CONTRA MODELO CONDICIONAL.

En un modelo marginal, la regresión de la variable respuesta sobre las variables explicativas es modelada separada de la correlación dentro de los individuos. En la regresión se modela la esperanza marginal, $E(y_{ij})$ como una función de las variables explicativas. Esto es, la respuesta promedio sobre la subpoblación que parte de un valor común de x . La esperanza marginal es cuando modelamos un estudio transversal. ⁴ Un modelo marginal tiene las siguientes suposiciones:

- (1) $E(y_{ij}) = \mu_{ij}$, es la esperanza marginal de la respuesta que depende de las variables explicativas, \mathbf{x}_{ij} , $g(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta}$ con g llamada la función *liga*.

⁴Consultar la referencia [5].

- (2) La varianza marginal depende de la media marginal.
- (3) La correlación entre y_{ij} y y_{ik} es una función de la media marginal y de parámetros adicionales α , es decir, $Corr(y_{ij}, y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha)$; donde ρ es una función conocida.

Por otro lado, en un modelo de efectos aleatorios la variable respuesta supone una función lineal de variables explicativas con coeficientes de regresión que varían de un individuo a otro. Los modelos de efectos aleatorios GLM especifican lo siguiente:

- (1) Dado b_i , las respuestas y_{i1}, \dots, y_{in_i} son mutuamente independientes y cumplen ser un modelo GLM con densidad $f(y_{ij}|b_i)$ perteneciente a la familia exponencial, además, z_{ij} es un subconjunto de x_{ij} .
- (2) Los efectos aleatorios b_i , son mutuamente independientes con distribución multivariada f .

La idea básica del modelo de efectos aleatorios es que tiene una componente heterogénea natural a través del individuo y junto con sus coeficientes de regresión se puede representar por una distribución de probabilidad. La correlación entre las observaciones para una persona se calcula por medio de las variables no observada, b_i . Un modelo de este tipo es conocido como un modelo de variable latente.

El modelo de efectos aleatorios es más útil cuando se quiere inferir acerca de los individuos más que en el promedio poblacional. El coeficiente de regresión β representa los efectos de las variables explicativas sobre el cambio del individuo. Esto es, el contraste al modelo marginal, los coeficientes describen el efecto de las variables explicativas sobre el promedio poblacional.

Por ejemplo, en un estudio clínico de nacimientos prematuros en mujeres alcohólicas, sea $y_{ij} = 1$ si el j -ésimo niño de la i -ésima mujer nace prematuramente y $y_{ij} = 0$ de otro modo. Suponiendo un único predictor x_{ij} dado por el número de tragos de alcohol por día. La aproximación marginal modelaría la media marginal de y_{ij} directamente suponiendo, por ejemplo, un modelo de regresión logística

$$\text{logit}(E(y_{ij})) = \text{logit}(P(y_{ij} = 1)) = \alpha + \beta x_{ij}.$$

Este modelo, explicaría el logit de la probabilidad de nacimientos prematuros, promediados en una población de mujeres. Nótese que si el modelo tuviera datos correlacionados, no sería competente suponer las observaciones independientes.

Por otro lado, el condicionamiento dado por los factores aleatorios para las mujeres, especifica el modelo condicional dado por

$$\text{logit}(E[y_{ij} | \mathbf{b}]) = \alpha + \beta x_{ij} + \mathbf{b}_i$$

donde \mathbf{b}_i representa el efecto aleatorio mujer. Esto corresponde a modelar la probabilidad condicional de un nacimiento prematuro por cada mujer. Además, los efectos aleatorios pueden tomar en cuenta a la variable no medible (mujer).

Teóricamente, es posible calcular la distribución marginal de \mathbf{y} (al menos de manera conceptual), a partir de la distribución de \mathbf{b} y la distribución condicional de $\mathbf{y} | \mathbf{b}$. Pero no es posible recuperar la distribución marginal de \mathbf{b} y la distribución condicional de $\mathbf{y} | \mathbf{b}$ a partir de la distribución marginal de \mathbf{y} . Lo que podría parecer favorable para el modelo condicional.

Sin embargo, en algunos casos la distribución marginal puede dar respuesta a preguntas comunes. Por ejemplo, en el consumo de alcohol, una pregunta in-

2.4. MÉTODOS DE BONDAD DE AJUSTE PARA GLMM. 65

interesante sería ¿en cuánto se reduce la incidencia de nacimientos prematuros, sobre el promedio del consumo de alcohol de la mujer? En este caso, la dificultad sería especificar la distribución condicional de $\mathbf{y} \mid \mathbf{b}$ y la distribución marginal de \mathbf{b} . La ventaja del modelo marginal y el sesgo de las ecuaciones que aproximan la estimación generalizada, es lo que prevalece en este caso.

Distinguir el modelo condicional del marginal es sencillo teóricamente, pero en la práctica frecuentemente es difícil. Por ejemplo, sería interesante pensar en *"la influencia del consumo de alcohol sobre nacimientos prematuros"*, lo cual no necesariamente especifica el tipo de modelo que hay que construir. El modelo condicional podría responder a como afecta el alcohol psicológicamente a cada mujer en forma individual. La distinción entre los modelos condicional y marginal es importante en la práctica.

2.4. MÉTODOS DE BONDAD DE AJUSTE PARA GLMM.

Un artículo reciente de Zhiying Pan y D. Y. Lin (Diciembre 2005) publicado en *Biometrics*⁵ usa métodos basados en sumas acumulativas de los residuales sobre covariables o valores predictivos de la variable respuesta para comparar cada proceso observado de manera gráfica y analítica con el número de realizaciones simuladas de una distribución nula. Los métodos particularmente son útiles para ver la forma funcional de una covariable o la función liga.

Aunque los modelos mal especificados pueden tener efectos adversos sobre la inferencia estadística, en los modelos lineales generalizados mixtos el uso de técnicas post-ajuste no ha sido explorado cuidadosamente. Las gráficas de residuales son rutinariamente usadas en modelos de regresión para respues-

⁵Ver referencia [7]

tas independientes. Difícilmente se determina si las observaciones muestrales reflejan un modelo mal especificado o fluctuaciones aleatorias. Este tipo de gráficas es aún más cuestionable con respuestas dependientes debido a la correlación de los residuales. Además, tales gráficas no son informativas para datos binarios porque todos los puntos ocurren en alguna de las líneas que están sobre una de las dos curvas de acuerdo a dos posibles valores de la respuesta.

En el artículo se desarrollan procesos objetivos e informativos para calcular lo adecuado de GLMM tomando sumas acumulativas de residuales sobre covariables o valores de predicción. Se aproxima la distribución nula de esos procesos de sumas acumulativas con procesos Gaussianos con media cero, cuyas distribuciones pueden ser generadas por simulación de Monte Carlo. Esto permite comparar los residuales observados con una realización de la distribución nula. Una inspección objetiva, es calcular el valor de p para la prueba supremo de Kolmogorov, basado en un gran número de realizaciones de una distribución nula. Los métodos de simulación se han desarrollado para los GML con respuestas indepenientes, para el modelo proporcional con datos sensoriales y para GLM con respuesta dependiente. El desarrollo del procedimiento de este tipo de técnica en GLMM es cuestionable porque la existencia de efectos aleatorios no solo complica la teoría, tal como se ve en las distribuciones asintóticas de procesos de sumas acumulativas y para la consistencia de las pruebas, sino también computacionalmente.

En el artículo se desarrollan nuevas técnicas para GLMM basadas en sumas acumulativas de residuales. Estos métodos pueden usarse para examinar componentes individuales de la parte determinista de un modelo. Los métodos que se han propuesto no requieren especificar una hipótesis alternativa y son informativos acerca de la naturaleza del modelo mal clasificado. Aunque se proponen distintas pruebas todas incluyen a las componentes aleatorias.

2.4.1. Desarrollo de la técnica.

Considerando un estudio longitudinal, en el cual se toman medidas repetidas de una variable respuesta de una muestra aleatoria de n individuos. Para $i = 1, \dots, n$, y $j = 1, \dots, M_i$, sea \mathbf{y}_{ij} la respuesta del i -ésimo individuo sobre la j -ésima repetición, y sea \mathbf{x}_{ij} y \mathbf{z}_{ij} las matrices de covariables asociadas con los efectos fijos y aleatorios, respectivamente. El modelo GLMM toma la forma

$$g\{E(\mathbf{y}_{ij}|\mathbf{b}_i)\} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i, \quad i = 1, \dots, n; \quad j=1, \dots, M_i \quad (2.21)$$

donde g es la función liga conocida y diferenciable.

La media marginal de \mathbf{y}_{ij} está dada por

$$E(\mathbf{Y}_{ij}) = E\{E(\mathbf{y}_{ij}|\mathbf{b}_i)\} = \int g^{-1}(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i)f_B(\mathbf{b}_i)d\mathbf{b}_i \quad (2.22)$$

Sea $m_{ij}(\boldsymbol{\theta}) = m_{ij} = E(\mathbf{y}_{ij})$ dada en (2.22). Dado $\hat{\boldsymbol{\theta}}$, el valor de predicción de la respuesta es $\widehat{m}_{ij} = m_{ij}(\hat{\boldsymbol{\theta}})$, y los residuales están definidos como $e_{ij} = \mathbf{y}_{ij} - \widehat{m}_{ij}$, los cuales se obtienen por integración numérica.

Esto es difícil de interpretar en gráficas de residuales individuales debido a la variabilidad desconocida que existe dentro del modelo. En contraste, es posible inspeccionar la distribución de cierto conjunto de residuales, para ello se usa una suma acumulativa de residuales con respecto a los valores de covariables o de predicción. Éstas sumas son casos especiales de las siguientes clases de procesos estocásticos:

$$T(\mathbf{x}) = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^{M_i} I(\mathbf{x}_{ij} \leq \mathbf{x})e_{ij}, \quad (2.23)$$

$$T_g(r) = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^{M_i} I(\widehat{m}_{ij} \leq r) e_{ij}, \quad (2.24)$$

donde $\mathbf{x} = (x_1, \dots, x_p)^T \in \mathbf{R}^p$, $r \in \mathbf{R}$, $I(\mathbf{x}_{ij} \leq \mathbf{x}) = I(x_{1ij} \leq x_1, \dots, x_{pij} \leq x_p)$, y x_{kij} es la k -ésima componente de \mathbf{x}_{ij} ($k = 1, \dots, p$).

Se define la hipótesis nula H_0 como: el modelo (2.21) es adecuado. Así, suponemos que las componentes aleatorias, incluyendo los efectos aleatorios y el error, son especificadas correctamente. Sobre H_0 , los procesos de sumas acumulativas son fluctuantes alrededor del 0, por lo tanto los valores grandes de $\sup_x |T(\mathbf{x})|$ o de $\sup_r |T_g(r)|$ podrían indicar un modelo mal especificado.

Se puede mostrar que $T(\mathbf{x})$ converge en distribución a un proceso gaussiano con media cero cuando H_0 no se rechaza. Definiendo

$$\widehat{T}(\mathbf{x}) = n^{-1/2} \sum_{i=1}^n \left\{ \sum_{j=1}^{M_i} I(\mathbf{x}_{ij} \leq \mathbf{x}) e_{ij} + \boldsymbol{\eta}'(\mathbf{x}; \hat{\boldsymbol{\theta}}) I^{-1}(\hat{\boldsymbol{\theta}}) \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} G_i, \quad (2.25)$$

donde (G_1, \dots, G_n) son variables aleatorias con distribución normal estándar y son independientes de los datos $(\mathbf{y}_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij})$, y

$$\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta}) = -n_{-1} \sum_{i=1}^n \sum_{j=1}^{M_i} I(\mathbf{x}_{ij} \leq \mathbf{x}) \frac{\partial m_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (2.26)$$

La distribución condicional de $\widehat{T}(\mathbf{x})$ dada los datos $(\mathbf{y}_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij})$ es la misma en el límite que la distribución no condicional de $T(\mathbf{x})$ bajo H_0 . Para obtener la distribución nula de $T(\mathbf{x})$, se obtienen una gran cantidad de valores de $\widehat{T}(\mathbf{x})$ generando una muestra aleatoria normal de (G_1, \dots, G_n) , para valores fijos de $(\mathbf{y}_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij})$ como valores observados. Además para muestras grandes, la distribución de $T_g(r)$ se puede aproximar por $\widehat{T}_g(r)$, donde $\widehat{T}_g(r)$ se obtiene de $\widehat{T}(\mathbf{x})$ reemplazando $I(\mathbf{x}_{ij} \leq \mathbf{x})$ en (2.25) y (2.26) por $I(m_{ij} \leq r)$.

2.4. MÉTODOS DE BONDAD DE AJUSTE PARA GLMM. 69

Para la forma funcional de la k -ésima componente del vector \boldsymbol{x} , consideremos el siguiente proceso

$$T_k(x) = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^{M_i} I(x_{kij} \leq x) e_{ij}, \quad k = 1, \dots, p \quad (2.27)$$

donde $x \in \mathbb{R}$. $T_k(x)$ es un caso especial de $T(\boldsymbol{x})$ dado en (2.23) con $x_l = \infty$ para toda $l \neq k$. Además, la distribución nula de $T_k(x)$ se puede aproximar por el proceso $\widehat{T}_k(x)$, el cual es un caso especial de $\widehat{T}(\boldsymbol{x})$ dado en (2.25) con $x_l = \infty$ para toda $l \neq k$.

La anormalidad de un proceso observado t_k de T_k se puede ver en una gráfica de t_k con pocas realizaciones del proceso T_k . Complementando esto con una prueba del tipo Kolmogorov para la estadística $S_k = \sup_k |T_k(x)|$, esta prueba llamada S_k por lo general, es consistente contra malas especificaciones de la forma funcional de x_k .

Capítulo 3

APLICACIÓN.

Los datos que se presentan forman parte de un estudio de plomo realizado en una cohorte de niños entre 1997–2000 y seguidos durante diez años. Los datos pertenecen a una muestra aleatoria de niños del Distrito Federal nacidos en el Hospital de Perinatología de los cuales se tienen medidas de la madre en diferentes periodos de gestación y lactancia y de los niños a partir de los tres años seis meses de edad, hasta los diez años tomadas cada seis meses de variables antropométricas, así como el nivel de plomo en diferentes etapas de gestación y en la sangre de los niños. En este ejemplo se busca, si existe, alguna relación entre el nivel plomo en la sangre y el crecimiento de niños. En este caso, la variable que se examinará es *crecio*, la cual es de tipo dicotómica y toma el valor 1 si el individuo no creció en los últimos seis meses y 0 si creció.

La base de datos consta de 1842 observaciones y 47 variables. Para el análisis solo se definen las covariables que resultaron significativas para los modelos.

subno es el número de registro con el que se identifica a cada niño en la base de datos.

sex con valores 1 si es niño y 2 si es niña.

year dado en meses con valores entre 42 y 120 meses. Especifica el periodo, en meses, en que se tomaron medidas a los niños.

hgb36 hemoglobina en la sangre de la madre a las 36 semanas de gestación.

hct20 porcentaje de eritrocitos en la sangre de la madre a las 20 semanas de gestación.

hct36 porcentaje de eritrocitos en la sangre de la madre a las 36 semanas de gestación.

lead6_36 el logaritmo del promedio de la concentración de plomo en la sangre en los primeros tres años de vida.

lnm28_36 el logaritmo del promedio de la concentración del plomo en la sangre de la mamá durante el tercer trimestre de embarazo.

lnlead18_36 el logaritmo del promedio de la concentración de plomo en la sangre en niños entre 18 y 36 meses de edad.

lnlead el logaritmo de la concentración de plomo en la sangre en los últimos seis meses.

Los estadísticos de resumen para las covariables continuas definidas anteriormente se presentan en la tabla (3.1).

El estudio se realiza a partir de un modelo longitudinal que incluye efectos aleatorios en la siguiente forma:

$$g\{E(\mathbf{y}_{ij}|\mathbf{X}_i\mathbf{b}_i)\} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + b_{i0} + b_{i1}\mathbf{z}_{ij} \quad i = 1, \dots, M \quad j = 1, \dots, M_i \quad (3.1)$$

variable	Media	EE	Mínimo	Máximo
hgb36	122.03	13.08	78	155
hct20	38.57	3.20	31.3	47.1
hct36	37.55	3.58	27	48
lead6_36	2.32	0.47	1.27	3.78
lnm28_36	2.08	0.71	0	3.50
lnlead18_36	2.37	0.49	1.20	3.75
lnlead	1.91	0.53	-0.69	4.11

Tabla 3.1: Tabla de estadísticos de resumen para las covariables utilizadas en los modelos.

donde b_{i0} es la intersección aleatoria y b_{i1} la pendiente aleatoria. Para el caso de datos binarios el modelo supone que $\mu_{ij} = g^{-1}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + b_{i0} + b_{i1} t_{ij})$ y se aproxima a 0 o a 1 cuando $t_{ij} \rightarrow \infty$. La función g está dada como

$$\log \left(\frac{E(\mathbf{y}_{ij} | \mathbf{X}_i \mathbf{b}_i)}{1 - E(\mathbf{y}_{ij} | \mathbf{X}_i \mathbf{b}_i)} \right)$$

para un modelo logístico, que se interpreta como el logaritmo de la razón de momios al incrementarse x en q unidades. Entonces, $\exp\{\beta q\}$ representa la razón de momios al incrementarse x en q unidades. Donde la razón de momios mide el grado de asociación entre dos variables, es decir, compara las probabilidades del efecto entre dos poblaciones, la que tiene y la que no tiene la causa. En nuestro caso la causa es si el niño no creció en los últimos seis meses.

3.1. Modelos ajustados.

La siguiente tabla (3.2), presenta los valores del percentil 25 y el valor mínimo para las variables de tipo continuo, donde el incrementar su valor en una unidad en la razón de momios no es realista, para el modelo logístico.

variable	percentil 25	mínimo	c = diferencia
<i>lnlead18_36</i>	2.0742	1.1945	0.8797
<i>lead6_36</i>	2.0245	1.2712	0.7532
<i>hgb36</i>	113	78	35
<i>hct20</i>	36.4	31.3	5.1
<i>hct36</i>	34.7	27	7.7
<i>lnm28_36</i>	1.6094	0	1.6094
<i>lnlead</i>	1.6094	-0.6931	2.3026

Tabla 3.2: Tabla de percentiles y valores mínimos.

Los modelos fueron ajustados con la subrutina GLLAMM de STATA 9.0.¹ A continuación se presentan los modelos que contienen una sola variable en el efecto fijo y el efecto aleatorio es de la forma $b_{i0} + b_{1i}t_{ij} = b_{i0} + b_{lnlead_i}t_{ij}$, es decir, los individuos y el nivel de plomo en los últimos seis meses son los efectos aleatorios, esto porque supone que el plomo puede afectar de manera diferente a cada individuo, por lo cual se define como un modelo de efecto aleatorio de un nivel de anidamiento. Solo se presentan las variables que fueron significativas en el modelo, *sex* es marginalmente significativa con significancia del 5%, pero la interacción de *sex * year* no es significativa.

Como en este tipo de modelos no se tiene una bondad de ajuste confiable, en la aplicación se construyen primero los modelos con una variable explicativa, los valores de estos se presentan en la tabla (3.3). Además, en otra tabla se

¹Consultar la referencia [12].

despliegan sus varianzas y covarianzas para estos modelos. Posteriormente se presentan los modelos con más de una variable explicativa, esto para ver si se percibe algún cambio importante en los valores de las varianzas de los parámetros debido a otras variables. Finalmente, se tabulan para todos los modelos (con una o más variables explicativas), los valores que se obtiene del Criterio de Información de Akaike (AIC) y el Bayesiano (BIC) para decidir cual es el mejor modelo GLMM que explica los datos.

<i>Covariable</i>	valor de $c\beta$	$\exp(c\beta)$	z	p -value
<i>lnlead18_36</i>	0.8316	2.2970	3.04	0.002
<i>lead6_36</i>	0.6609	1.9366	2.66	0.008
<i>hgb36</i>	1.2477	3.4823	2.41	0.016
<i>hct20</i>	-0.6295	0.5329	-2.20	0.028
<i>hct36</i>	0.7469	2.1105	2.00	0.046
<i>year</i>	0.0158	1.0160	2.78	0.005
<i>sex</i>	0.6181	1.8554	1.90	0.058
<i>lnm28_36</i>	0.6114	1.8430	1.57	0.116
<i>lnlead</i>	0.8105	2.2491	1.02	0.307

Tabla 3.3: Modelos con una sola predicción para el efecto fijo.

La razón de momios de la variable *crecio* cuando el *logaritmo del promedio de la concentración de plomo de niños de año y medio a tres años de vida (lnlead18_36)* se incrementa en 0.88 unidades es de 129.70%. Es decir, los niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses.

La razón de momios de la variable *crecio* cuando *el logaritmo del promedio de la concentración del plomo en los primeros tres años de vida (lead6_36)*

se incrementa en 0.75 unidades es de 93.66 % más. También en este caso, los niños con mayores concentraciones de plomo en los primeros tres años de vida tienen mayor riesgo de dejar de crecer por seis meses.

La razón de momios de la variable *crecio* cuando *hemoglobina a las 36 semanas de gestación (hgb36)* se incrementa en 35 unidades es de 248.23 %. Por lo que al incrementarse la hemoglobina los niños tienen mayor riesgo de dejar de crecer por seis meses.

Mientras que la razón de momios de la variable *crecio* al incrementarse el porcentaje de eritrocitos en la sangre de la madre a las 20 semanas de gestación (*hct20*) 5.1 unidades es 53.29 % menor, es decir, es un factor protector, a mayor nivel menor el riesgo de dejar de crecer.

La razón de momios de la variable *crecio* cuando el porcentaje de eritrocitos en la sangre de la madre a las 36 semanas de gestación (*hct36*) se incrementa en 7.7 unidades es de 111.05 %. Es decir, el incremento en riesgo de dejar de crecer al incrementarse *hct36* es de 111 %.

La razón de momios de la variable *crecio* cuando (*year*) se incrementa en un año es de 1.60 % más. Es decir, al pasar un año de vida los niños tuvieron mayor riesgo de no crecer.

La razón de momios de los niños que no crecieron al cambiar de niño a niña la variable *sex* es 85.54 % más, es decir, las niñas tienen menor riesgo de dejar de crecer.

La razón de momios de la variable *crecio* cuando *el logaritmo del promedio de la concentración de plomo en la madre durante el tercer trimestre de embarazo (lnm28_36)* se incrementa en 1.61 unidades es de 84.30 % más, por lo que de nuevo, los niños que sus madres tuvieron mayores concentraciones de plomo durante el tercer trimestre de embarazo tienen mayor riesgo de dejar de crecer por seis meses.

La razón de momios de la variable *crecio* cuando *el logaritmo del promedio de la concentración de plomo en los últimos seis meses (lnlead)* se incrementa en 2.30 unidades es de 124.91 %. Es decir, al incrementarse la concentración de plomo en los últimos seis meses los niños incrementan el riesgo de dejar de crecer.

<i>Covariable</i>	$\sigma_{subno}^2(\sigma_{\sigma_{subno}^2})$	$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}^2})$	$cov(subno, lnlead)[\sigma_{cov}]$
<i>lnlead18_36</i>	5.1165(3.3968)	0.6523(0.5685)	-1.8268(1.3827)
<i>lead6_36</i>	4.6598(3.1870)	0.5424(0.5027)	-1.5897(1.2606)
<i>hgb36</i>	2.2073(3.3933)	0.2045(0.5033)	-0.6719(1.3115)
<i>hct20</i>	4.2757(3.7809)	0.5254(0.6563)	-1.4987(1.5775)
<i>hct36</i>	1.3612(2.8763)	0.0999(0.3464)	-0.3687(1.0053)
<i>year</i>	0.7601(1.3842)	0.0180(0.1028)	-0.1168(0.4344)
<i>sex</i>	4.6146(3.5068)	0.5377(0.5936)	-1.5752(1.4485)
<i>lnm28_36</i>	4.5273(3.7890)	0.4459(0.5623)	-1.4208(1.4720)
<i>lnlead</i>	5.1040(4.1717)	0.5975(0.6493)	-1.7463(1.6448)

Tabla 3.4: Varianzas–covarianzas de los modelos con una sola predicción para el efecto fijo.

La tabla (3.4), muestra los valores de las varianzas y covarianzas de los modelos de una sola variable explicativa como efecto fijo, en ella se percibe que las varianzas para la intercepción aleatoria (*subno*) son más altas que las varianzas para el efecto aleatorio (*lnlead*), esto significa que existe una variabilidad en el crecimiento entre los individuos que no está explicada por las variables en el modelo y, el hecho que σ_{lnlead}^2 sea mayor a cero indica que los niveles de plomo afectan al crecimiento de manera diferente. Además, los modelos con menor varianza en la intercepción aleatoria (*subno*) y en la pendiente aleatoria (*lnlead*) son en los que la variable explicativa explica más la varianza por el modelo.

Con estas variables se ajustan modelos con predictores múltiples para ver si hay alguna relación significativa entre ellas. En adelante todos los modelos contienen las variables *year* y *sex*. Como ya se mencionó, la interacción entre *year* y *sex* no fue significativa.

	<i>year</i>	<i>sex</i>
β	0.0148	0.5004
$exp(\beta)$	1.0149	1.6494
z	2.57	1.58
p -value	0.010	0.114
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	1.8198(2.4737)	
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.1438(0.3552)	
$cov(subno, lnlead)[\sigma_{cov}]$	-0.5116(0.9673)	

Tabla 3.5: Modelo $logit(E(crecio|t)) = \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

En la tabla (3.5) se ve que el cambio en la razón de momios de la variable *crecio* al pasar un año (12 meses) cuando *sex* se mantiene constante es el 1.49% más. Mientras que la razón de momios de la variable *crecio* al cambiar de niño a niña y manteniendo *year* constante es del 64.94% más. Este modelo corrobora que, los niños con edades más altas tienen mayor riesgo de dejar de crecer por seis meses que las niñas.

La razón de momios de la variable *crecio* al incrementarse en 7.7 unidades *hct36* y manteniendo constantes *year* y *sex* es de 129.49% más. Mientras que la razón de momios de *crecio* al pasar un año es de 1.21% más, al mantener constantes *hct36* y *sex*. Y la razón de momios de *crecio* al cambiar de niño a niña es de 86.16% más, al mantener constantes *hct36* y *year*, lo cual da un riesgo más severo en niños que en niñas. También este modelo reporta que los niños con mayores concentraciones de *hct36* tienen mayor riesgo de dejar de

	7.7*hct36	year	sex
$c * \beta$	0.8307	0.0121	0.6214
$exp(c * \beta)$	2.2949	1.0122	1.8616
z	2.26	1.64	1.50
p -value	0.024	0.101	0.133
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	0.9837(2.9110)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.0762(0.3845)		
$cov(subno, lnlead)[\sigma_{cov}]$	-0.2738(1.0747)		

Tabla 3.6: Modelo $logit(E(crecio|t)) = \beta_{hct36} * hct36 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

crecer por seis meses. En los valores de las varianzas se ve una disminución con respecto al modelo cuya efecto fijo es $hct36$, esto significa que la influencia de año y sexo explican con menor dispersión los datos.

	5.1*hct20	year	sex
$c * \beta$	-0.6633	0.0161	0.6456
$exp(c * \beta)$	0.5151	1.0162	1.9071
z	-2.31	2.45	1.77
p -value	0.021	0.014	0.077
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	1.1363(1.0805)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.1640(0.2112)		
$cov(subno, lnlead)[\sigma_{cov}]$	-0.4317(0.4785)		

Tabla 3.7: Modelo $logit(E(crecio|t)) = \beta_{hct20} * hct20 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

En la tabla (3.7) se reporta que la razón de momios de la variable $crecio$ al incrementarse en 5.1 unidades $hct20$ y manteniendo constantes $year$ y sex

es de 51.51% menos. Mientras que la razón de momios de *crecio* al pasar un *año* es de 1.62% más, al mantener constantes *hct20* y *sex*. Y la razón de momios de *crecio* al cambiar de niño a niña es de 90.71% más, al mantener constantes *hct20* y *year*, lo cual da un daño más severo en niños que en niñas. Del modelo se interpreta que los niños aumentan el riesgo de no crecer al mantener constante *hct20*, pero *hct20* sigue siendo un factor protector. Las varianzas se reportan menores con respecto al modelo con efecto fijo dado por *hct20* en la tabla (3.4).

	35 * <i>hgb36</i>	<i>year</i>	<i>sex</i>
$c * \beta$	1.3504	0.0110	0.7438
$exp(c * \beta)$	3.8591	1.0111	2.1039
z	2.58	1.42	1.63
p -value	0.010	0.155	0.103
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	0.7813(1.1431)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.1026(0.2005)		
$cov(subno, lnlead)[\sigma_{cov}]$	-0.2831(0.4781)		

Tabla 3.8: Modelo $logit(E(crecio|t)) = \beta_{hgb36} * hgb36 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

La razón de momios de la variable *crecio* al incrementarse en 35 unidades *hgb36* y manteniendo constantes *year* y *sex* es de 285.91% más. Mientras que la razón de momios de *crecio* al pasar un *año* es de 1.11% más, al mantener constantes *hct36* y *sex*. Y la razón de momios de *crecio* al cambiar de niño a niña es de 110.39% más, al mantener constantes *hct36* y *year*, lo cual da un daño más severo en niños que en niñas. También este modelo reporta que los niños crecieron menos y que al incrementarse la hemoglobina los niños tienen mayor riesgo de dejar de crecer por seis meses. En este caso las varianzas disminuyen bastante con respecto al modelo con variable *hct36* como única

variable explicativa.

	5.1*hct20	35*hgb36	year	sex
$c * \beta$	-1.1181	1.8260	0.0116	0.8677
$exp(c * \beta)$	0.3269	6.2091	1.0116	2.3815
z	-2.56	2.75	1.28	1.61
p -value	0.010	0.006	0.199	0.108
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$		1.6191(1.5672)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$		0.2475(0.2882)		
$cov(subno, lnlead)[\sigma_{cov}]$		-0.6330(0.6677)		

Tabla 3.9: Modelo $logit(E(crecio|t)) = \beta_{hct20} * hct20 + \beta_{hgb36} * hgb36 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

En la tabla (3.9) se tiene que la razón de momios de la variable *crecio* al incrementarse en 5.1 unidades *hct20* y manteniendo constantes *hgb36*, *year* y *sex* es de 32.69 % menos, lo cual sigue siendo un factor protector. La razón de momios de la variable *crecio* al incrementarse en 35 unidades *hgb36* y manteniendo constantes *hct20*, *year* y *sex* es de 520.91 % más. Mientras que la razón de momios de *crecio* al pasar un *año* es de 1.16 % más, al mantener constantes *hct20*, *hct36* y *sex*. Y la razón de momios de *crecio* al cambiar de niño a niña es de 138.15 % más, al mantener constantes *hct20*, *hct36* y *year*, lo cual da un daño más severo en niños que en niñas. Este modelo reporta que los niños con mayores concentraciones de hemoglobina en la sangre de las madres a las 36 semanas de gestación tienen mayor riesgo de dejar de crecer por seis meses y que al incrementarse la hemoglobina los niños aumentan ese riesgo de manera significativa. En este modelo los valores de la varianza se dispersan más con respecto al modelo con efectos fijos *hct20*, *year*, *sex*, y con el modelo *hgb36*, *year*, *sex* presentado en la tabla (3.8).

La razón de momios de la variable *crecio* al incrementarse en 0.88 unidades

	0.8797*lnlead18_36	year	sex
$c * \beta$	0.7471	0.0130	0.4956
$\exp(c * \beta)$	2.1109	1.0131	1.6415
z	2.76	2.25	1.60
p -value	0.006	0.025	0.110
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	4.1954(3.4775)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.5615(0.6014)		
$cov(subno, lnlead)[\sigma_{cov}]$	-1.5349(1.4413)		

Tabla 3.10: Modelo $logit(E(crecio|t)) = \beta_{lnlead18.36} * lnlead18_36 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

el logaritmo del promedio de la concentración de plomo de niños entre año y medio y tres años de vida ($lnlead18_36$) y manteniendo constantes $year$ y sex es de 111.09 % más. Mientras que la razón de momios de $crecio$ al pasar un año es de 1.31 % más, al mantener constantes $lnlead18_36$ y sex . Y la razón de momios de $crecio$ al cambiar de niño a niña es de 64.15 % más, al mantener constantes $lnlead18_36$ y $year$. El modelo reporta que los niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses. La diferencia en sus varianzas no cambio mucho con respecto al modelo dado por $lnlead18_36$, esto sugiere que el nivel de plomo durante el tercer trimestre del embarazo tiende a variar entre las madres de los niños sin importar el sexo y la edad.

La razón de momios de $crecio$ cuando al incrementarse en 0.75 unidades el logaritmo del promedio de la concentración del plomo en los primeros tres años de vida ($lead6_36$) es del 79.45 % más, manteniendo $year$ y sex constantes. Mientras que la razón de momios de $crecio$ al pasar un año es de 1.33 % más, al mantener constantes el logaritmo del promedio de la concentración del plomo en los primeros tres años de vida ($lead6_36$) y sex . Por otro lado

	0.7532*lead6_36	year	sex
$c * \beta$	0.5847	0.0132	0.4950
$\exp(c * \beta)$	1.7945	1.0133	1.6405
z	2.39	2.28	1.59
p -value	0.017	0.023	0.112
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	0.9779(0.7991)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.1276(0.1383)		
$cov(subno, lnlead)[\sigma_{cov}]$	-0.3533(0.3317)		

Tabla 3.11: Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{\text{lead6}_36} * \text{lead6}_36 + \beta_{\text{year}} * \text{year} + \beta_{\text{sex}} * \text{sex} + b_0 + b_{\text{lnlead}} * t$.

la razón de momios de *crecio* al cambiar de niño a niña es de 64.05 % más, al mantener constantes *el logaritmo del promedio de la concentración del plomo en los primeros tres años de vida (lead6_36)* y *year*; aquí también se tiene que el plomo afecta más a niños que a niñas y que los niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses, sin importar el sexo. En este caso las varianzas si sufren un cambio drástico en cuanto a los datos en el modelo de una variable explicativa de la tabla (3.4).

La razón de momios de *crecio* cuando al incrementarse en 1.61 unidades *el logaritmo del promedio de la concentración del plomo en la mamá durante el tercer trimestre de embarazo (lnm28_36)* es del 60.17 % más, manteniendo *year* y *sex* constantes. Mientras que la razón de momios de *crecio* al pasar un *año* es de 1.91 % más, al mantener constantes *lnm28_36* y *sex*. Por otro lado, la razón de momios de *crecio* al cambiar de niño a niña es de 76.76 % más, al mantener constantes *lnm28_36* y *year*; por lo que de nuevo se tiene que los niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses. Aquí la varianza solo disminuye en un 50 %

	1.6094*lnm28_36	year	sex
$c * \beta$	0.4710	0.0189	0.5696
$exp(c * \beta)$	1.6017	1.0191	1.7676
z	1.27	2.99	1.63
p -value	0.204	0.003	0.104
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	2.3495(3.1730)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.1841(0.4372)		
$cov(subno, lnlead)[\sigma_{cov}]$	-0.6577(1.2130)		

Tabla 3.12: Modelo $logit(E(crecio|t)) = \beta_{lnm28.36} * lnm28_36 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

con respecto al valor dado en la tabla (3.4).

	5.1*hct20	1.61*lnm28_36	year	sex
$c * \beta$	-0.5817	0.3541	0.0190	0.6927
$exp(c * \beta)$	0.5589	1.4249	1.0192	1.9990
z	-1.97	0.94	2.77	1.85
p -value	0.049	0.345	0.006	0.065
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	3.8627(3.7607)			
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.4731(0.6446)			
$cov(subno, lnlead)[\sigma_{cov}]$	-1.3518(1.5603)			

Tabla 3.13: Modelo $logit(E(crecio|t)) = \beta_{hct20} * hct20 + \beta_{lnm28.36} * lnm28_36 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

La razón de momios de la variable *crecio* al incrementarse en 5.1 unidades *hct20* y manteniendo constantes *lnm28_36*, *year* y *sex* es de 55.89% menos. La razón de momios de *crecio* cuando al incrementarse en 1.61 unidades *el logaritmo del promedio de la concentración del plomo en la mamá durante*

el tercer trimestre de embarazo ($lnm28_36$) es del 42.49% más, manteniendo $hct20$, $year$ y sex constantes. Mientras que la razón de momios de $crecio$ al pasar un año es de 1.92% más, al mantener constantes $hct20$, $lnm28_36$ y sex . Por otro lado, la razón de momios de $crecio$ al cambiar de niño a niña es de 99.90% más, al mantener constantes $hct20$, $lnm28_36$ y $year$; por lo que de nuevo se tiene que los niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses y que $hct20$ es un factor protector. En cuanto a las varianzas se ve que los valores son mayores respecto a los modelos de las tablas (3.7) y (3.12).

	$35*hgb36$	$1.61*lnm28_36$	$year$	sex
$c * \beta$	1.3340	0.3053	0.0102	0.7309
$exp(c * \beta)$	3.7962	1.3570	1.0103	2.0770
z	2.47	0.63	1.29	1.52
p -value	0.014	0.529	0.198	0.128
$\sigma_{subno}^2(\sigma_{\sigma_{subno}^2})$		4.0740(6.0199)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}^2})$		0.5330(0.9684)		
$cov(subno, lnlead)[\sigma_{cov}]$		-1.4736(2.4037)		

Tabla 3.14: Modelo $logit(E(crecio|t)) = \beta_{hgb36} * hgb36 + \beta_{lnm28_36} * lnm28_36 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

La razón de momios de la variable $crecio$ al incrementarse en 35 unidades $hgb36$ y manteniendo constantes $lnm28_36$, $year$ y sex es de 279.62% más. La razón de momios de $crecio$ cuando al incrementarse en 1.61 unidades $lnm28_36$ es del 35.70% más, manteniendo $hgb36$, $year$ y sex constantes. Mientras que la razón de momios de $crecio$ al pasar un año es de 1.03% más, al mantener constantes $hgb36$, $lnm28_36$ y sex . Por otro lado, la razón de momios de $crecio$ al cambiar de niño a niña es de 107.70%, al mantener constantes $hgb36$, $lnm28_36$ y $year$; por lo que de nuevo se tiene que los

niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses. En este modelo ocurre el mismo aumento que el modelo anterior respecto a sus efectos fijos, la dispersión es mayor que en los respectivos modelos individuales.

	$5.1 * hct20$	$35 * hgb36$	$1.61 * lnm28_36$	$year$	sex
$c * \beta$	-1.1081	1.8645	0.3296	0.0109	0.8321
$exp(c * \beta)$	0.3302	6.4529	1.3904	1.0109	2.2982
z	-2.51	2.68	0.58	1.19	1.50
p -value	0.012	0.007	0.560	0.235	0.135
$\sigma_{subno}^2 (\sigma_{\sigma_{subno}}^2)$	7.4917(6.9023)				
$\sigma_{lnlead}^2 (\sigma_{\sigma_{lnlead}}^2)$	1.1062(1.1999)				
$cov(subno, lnlead) [\sigma_{cov}]$	-2.8788(2.8541)				

Tabla 3.15: Modelo $logit(E(crecio|t)) = \beta_{hct20} * hct20 + \beta_{hgb36} * hgb36 + \beta_{lnm28_36} * lnm28_36 + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

La razón de momios de la variable *crecio* al incrementarse en 5.1 unidades *hct20* y manteniendo constantes *hgb36*, *lnm28_36*, *year* y *sex* es de 33.02% menos, lo cual indica que *hct20* sigue siendo un factor protector; la razón de momios de la variable *crecio* al incrementarse en 35 unidades *hgb36* y manteniendo constantes *hct20*, *lnm28_36*, *year* y *sex* es de 545.29% más. La razón de momios de *crecio* cuando al incrementarse en 1.61 unidades *lnm28_36* es del 39.04% más, manteniendo *hct20*, *hgb36*, *year* y *sex* constantes. Mientras que la razón de momios de *crecio* al pasar un año es de 1.09% más, al mantener constantes *hct20*, *hgb36*, *lnm28_36* y *sex*. Por otro lado, la razón de momios de *crecio* al cambiar de niño a niña es de 129.82%, al mantener constantes *hct20*, *hgb36*, *lnm28_36* y *year*; por lo que de nuevo se tiene que los niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses. Se ve un marcado aumento en las varianzas con

respecto a los modelos de las tablas (3.9), (3.13), (3.14).

	2.3026*lnlead	year	sex
$c * \beta$	1.3679	0.0174	0.5412
$\exp(c * \beta)$	3.9273	1.0176	1.7180
z	1.72	2.90	1.68
p -value	0.085	0.004	0.092
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$	5.1652(4.1724)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$	0.6469(0.6587)		
$cov(subno, lnlead)[\sigma_{cov}]$	-1.8280(1.6517)		

Tabla 3.16: Modelo $logit(E(crecio|t)) = \beta_{lnlead} * lnlead + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

La razón de momios de *crecio* cuando al incrementarse en 2.30 unidades *el logaritmo del promedio de la concentración del plomo en los últimos seis meses (lnlead)* es del 292.73% más, manteniendo *year* y *sex* constantes. Mientras que la razón de momios de *crecio* al pasar un *año* es de 1.76% más, al mantener constantes *lnlead* y *sex*. Por otro lado, la razón de momios de *crecio* al cambiar de niño a niña es de 71.80% más, al mantener constantes *lnlead* y *year*; aquí también se tiene que el plomo afecta más a niños que a niñas y que los niños tiene mayor riesgo de dejar de crecer sin importar el sexo cuando se incrementa la concentración de plomo. En este caso no hay un cambio grande en las varianzas, aunque se reportan mayores a las dadas en la tabla (3.4) para *lnlead*.

La razón de momios de la variable *crecio* al incrementarse en 5.1 unidades *hct20* y manteniendo constantes *lnlead*, *year* y *sex* es de 53.01% menos. La razón de momios de *crecio* cuando al incrementarse en 2.30 unidades *el logaritmo del promedio de la concentración del plomo en los últimos seis meses (lnlead)* es del 124.64%, manteniendo *hct20*, *year* y *sex* constantes.

	5.1*hct20	2.3026*lnlead	year	sex
$c * \beta$	-0.6347	0.8094	0.0179	0.6515
$\exp(c * \beta)$	0.5301	2.2464	1.0180	1.9184
z	-2.18	0.86	2.58	1.77
p -value	0.029	0.388	0.010	0.076
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$		5.2560(4.2763)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$		0.7039(0.7330)		
$cov(subno, lnlead)[\sigma_{cov}]$		-1.9234(1.7625)		

Tabla 3.17: Modelo $\text{logit}(E(\text{crecio}|t)) = \beta_{hct20} * hct20 + \beta_{lnlead} * lnlead + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

Mientras que la razón de momios de *crecio* al pasar un *año* es de 1.80 % más, al mantener constantes *hct20*, *lnlead* y *sex*. Por otro lado, la razón de momios de *crecio* al cambiar de niño a niña es de 91.84 % más, al mantener constantes *hct20*, *lnlead* y *year*; por lo que de nuevo se tiene que los niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses y que *hct20* es un factor protector. Pero en este caso *lnlead* indica el mayor grado de riesgo de dejar de crecer. En cuanto a los cambios de varianza se ve un aumento respecto al modelo de la tabla (3.7). Mientras que con respecto al modelo (3.16), las varianzas se mantienen estables.

La razón de momios de la variable *crecio* al incrementarse en 5.1 unidades *hct20* y manteniendo constantes *hgb36*, *lnlead*, *year* y *sex* es de 32.67 % menos, lo cual indica que *hct20* sigue siendo un factor protector; la razón de momios de la variable *crecio* al incrementarse en 35 unidades *hgb36* y manteniendo constantes *hct20*, *lnlead*, *year* y *sex* es de 530.01 % más. La razón de momios de *crecio* cuando al incrementarse en 2.30 unidades *lnlead* es del 138.59 %, manteniendo *hct20*, *hgb36*, *year* y *sex* constantes. Mientras que la razón de momios de *crecio* al pasar un *año* es de 1.35 % más, al mante-

	5.1*hct20	35*hgb36	2.3026*lnlead	year	sex
$c * \beta$	-1.1187	1.8406	0.8696	0.0134	0.9107
$exp(c * \beta)$	0.3267	6.3001	2.3859	1.0135	2.4860
z	-2.50	2.68	0.65	1.40	1.66
p -value	0.012	0.007	0.518	0.160	0.097
$\sigma_{subno}^2(\sigma_{\sigma_{subno}}^2)$			8.5282(7.9623)		
$\sigma_{lnlead}^2(\sigma_{\sigma_{lnlead}}^2)$			1.2360(1.3154)		
$cov(subno, lnlead)[\sigma_{cov}]$			-3.2467(3.2105)		

Tabla 3.18: Modelo $logit(E(crecio|t)) = \beta_{hct20} * hct20 + \beta_{hgb36} * hgb36 + \beta_{lnlead} * lnlead + \beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$.

ner constantes $hct20$, $hgb36$, $lnlead$ y sex . Por otro lado, la razón de momios de $crecio$ al cambiar de niño a niña es de 148.60%, al mantener constantes $hct20$, $hgb36$, $lnlead$ y $year$; por lo que de nuevo se tiene que los niños con mayores concentraciones de plomo tienen mayor riesgo de dejar de crecer por seis meses. Este modelo presenta un aumento muy drástico en los valores de sus varianzas con respecto a los modelos anteriores como (3.9), (3.16), (3.15). Por lo que el nivel de plomo en la sangre en los últimos seis meses es más disperso entre los niños respecto a las variables $hct20$ y $hgb36$.

Los criterios AIC y BIC se usan como métodos de bondad de ajuste para modelos lineales, lineales generalizados, mixtos y lineales generalizados mixtos aunque el número de variables entre los modelos no sea el mismo. La decisión del mejor modelo según estos criterios se toma por el mínimo valor entre los valores AIC o BIC de los modelos ajustados. La manera en que se calculan es la siguiente:

$$AIC = -2\ell(\hat{\theta}|y) + 2n_{par},$$

$$BIC = -2\ell(\hat{\theta}|y) + n_{par} \log(N)$$

donde $\ell(\hat{\theta}|y)$ es el logaritmo de la verosimilitud del vector de parámetros estimados $\hat{\theta}$, n_{par} es el número de parámetros en el modelo, N es el total de observaciones involucradas en el modelo.

Los valores de estos criterios para los modelos presentados aquí, se dan en la tabla (3.19). Así, el mejor modelo es el dado en la tabla (3.15), seguido por el modelo de la tabla (3.18).

Modelo $\text{logit}(E(\text{crecio} t))$	$\ell(\hat{\theta}, y)$	AIC	BIC
$\beta_{sex} * sex + b_0 + b_{lnlead} * t$	-271.7195	549.4390	565.9948
$\beta_{year} * year + b_0 + b_{lnlead} * t$	-269.7670	545.5339	562.0897
$\beta_{lnlead} * lnlead + b_0 + b_{lnlead} * t$	-273.1500	552.3000	568.7345
$\beta_{lnlead18_36} * lnlead18_36 + b_0 + b_{lnlead} * t$	-267.8117	541.6234	558.1283
$\beta_{lead6_30} * lead6_30 + b_0 + b_{lnlead} * t$	-269.2134	544.4267	560.9465
$\beta_{hgb36} * hgb36 + b_0 + b_{lnlead} * t$	-149.7268	305.4536	320.2156
$\beta_{hct36} * hct36 + b_0 + b_{lnlead} * t$	-164.4212	334.8425	349.8950
$\beta_{hct20} * hct20 + b_0 + b_{lnlead} * t$	-204.9140	415.8281	431.3614
$\beta_{lnm28_36} * lnm28_36 + b_0 + b_{lnlead} * t$	-244.3385	494.6771	511.0098
$\beta_{year} * year + \beta_{sex} * sex + b_0 + b_{lnlead} * t$	-268.4163	544.8326	566.9070
$\beta_{lnm28_36} * lnm28_36 + \beta_{year} * year +$ $\beta_{sex} * sexo + b_0 + b_{lnlead} * t$	-237.5957	485.1914	512.4126
$\beta_{lnm28_36} * lnm28_36 + \beta_{hct20} * hct20 +$ $\beta_{year} * year + \beta_{sex} * sexo + b_0 + b_{lnlead} * t$	-187.1595	386.3190	417.3857
$\beta_{lnm28_36} * lnm28_36 + \beta_{hgb36} * hgb36 +$ $\beta_{year} * year + \beta_{sex} * sexo + b_0 + b_{lnlead} * t$	-145.9787	303.9573	333.4813
$\beta_{lnm28_36} * lnm28_36 + \beta_{hgb36} * hgb36 +$ $\beta_{hct20} * hct20 + \beta_{year} * year +$ $\beta_{sex} * sexo + b_0 + b_{lnlead} * t$	-109.2046	232.4092	266.8539
$\beta_{lnlead} * lnlead + \beta_{year} * year +$ $\beta_{sex} * sexo + b_0 + b_{lnlead} * t$	-266.8773	543.7545	571.1453
$\beta_{lnlead} * lnlead + \beta_{hct20} * hct20 +$ $\beta_{year} * year + \beta_{sex} * sexo + b_0 + b_{lnlead} * t$	-199.3554	410.7107	441.7774
$\beta_{lnlead} * lnlead + \beta_{hct20} * hct20 +$ $\beta_{hgb36} * hgb36 + \beta_{year} * year +$ $\beta_{sex} * sexo + b_0 + b_{lnlead} * t$	-109.6856	233.3712	267.8159
$lnlead18_36 + \beta_{year} * year +$ $\beta_{sex} * sexo + b_0 + b_{lnlead} * t$	-263.6332	537.2664	564.7746

Tabla 3.19: Criterio de Información de Akaike (AIC) y Bayesiano (BIC).

3.2. Conclusión.

Del análisis que se realizó se concluye que el plomo en efecto daña al crecimiento de los niños, en particular aumenta el riesgo de no crecer durante los últimos seis meses de tomada la medida de sangre por analizar, pues la mayor influencia que hay es el nivel de plomo en los últimos seis meses (*lnlead*), el cual se reporta con valor de 238.59% en el modelo dado en la tabla (3.18) (el segundo mejor modelo ajustado) y la concentración de plomo en la sangre de la madre durante el tercer trimestre de embarazo con un valor de 139.04% obtenido en el modelo de la tabla (3.15), el mejor modelos según los criterios de bondad de ajuste. Ambos modelos consideran a las variables *year* y *sex*.

Por otro lado, se ve que las niñas tienen menos riesgo de presentar periodos de tiempo sin crecer que los niños, con un valor de 64.94% los niños presentar crecer menos que las niñas, según el modelo de la tabla (3.5); al ser niños entre 3.5 y 10 años esta situación podría ser natural. También al paso de los meses los niños reportan crecer menos, en el mismo modelo se reporta un valor de 1.49% de dejar de crecer al paso de un año, por otro lado en la gráfica (3.1), se ve que hay más riesgo de periodos sin crecer en niños entre 7 y 10 años que en niños entre 3.5 y 6.5 años, de nuevo esto es un comportamiento natural. La gráfica también muestra que las niñas crecen durante periodos más largos que los niños.

Con respecto a la variable *hct20*, es decir, un alto porcentaje de eritrocitos en la sangre a las 20 semanas de gestación en las madres, implica que los niños manifestaban un beneficio en el riesgo de presentar su periodo sin crecimiento. Mientras que el porcentaje de eritrocitos en la sangre a las 36 semanas de gestación (*hct36*) sí es una característica que afecta al crecimiento.

La hemoglobina en la sangre a las 36 semanas de gestación reporta una influencia al no crecimiento, siendo significativa.

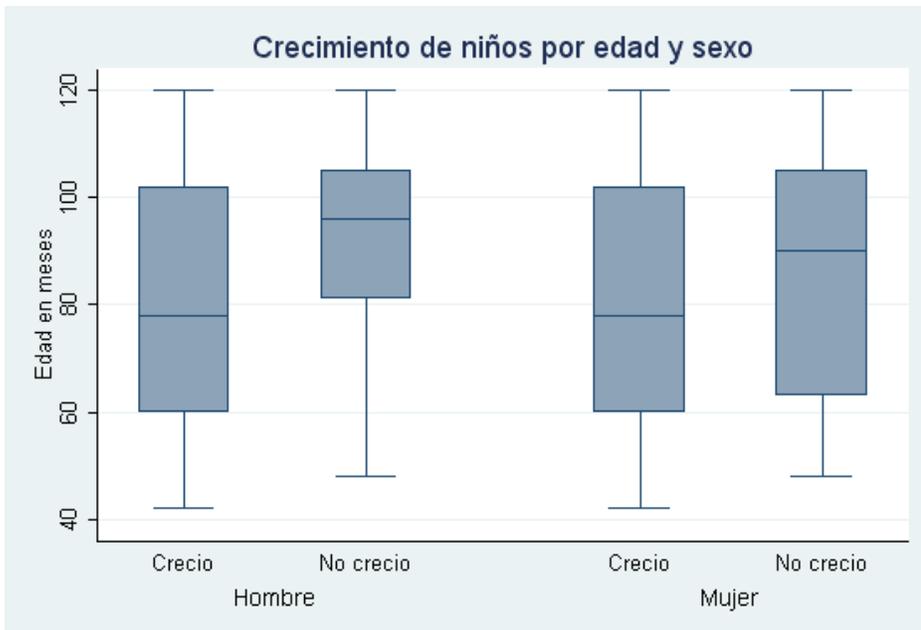


Figura 3.1: Crecimiento de niños por edad y sexo

Capítulo 4

CONCLUSIÓN.

Los modelos GLMM proporcionan una mejor descripción de los datos, pues caracterizan a cada individuo según los factores o efectos que pueden comportarse de manera distinta para cada individuo, así ya es más preciso particularizarlos.

Esto se manifiesta en la aplicación del capítulo tres, en donde cada niño se asocia con un nivel de plomo particular de la medida dada cada seis meses, esto al ser aleatorio se ve afectado por un valor de varianza, el cual es muy fluctuante de modelo a modelo, lo cual corrobora el hecho de ser distinto cada niño.

En cuanto a los resultados del análisis se ve que los niños si tienen una repercusión en el crecimiento debido a la presencia del plomo, pero también tienen un beneficio al presentar un porcentaje alto de eritrocitos en la sangre de las madres a las 20 semanas de gestación.

Apéndice A

RESULTADOS DE STATA.

Una salida típica de STATA 9.0 para estimar los modelos presentados en el capítulo tres es la que se presenta a continuación.

En esta salida se ve la varianza de la intercepción aleatoria (*subno*) y la razón de momios de la pendiente aleatoria (*lnlead*), con sus respectivos errores estándar.

```
. eq modelo: constante lnlead

. gllamm  crecio  lnlead, link(logit) family(binomial) i(subno1)
eq(modelo)

number of level 1 units = 1769 number of level 2 units = 174

Condition Number = 30.303454

gllamm model
```

log likelihood = -273.15002

```
-----
crecio |          Coef.   Std. Err.   z    P>|z|   [95% Conf. Interval]
-----+-----
lnlead |   .3520125   .3447349   1.02   0.307   - .3236554   1.02768
_cons  |  -4.257337   .7982813  -5.33   0.000   -5.82194   -2.692734
-----
```

Variances and covariances of random effects

***level 2 (subno1)

var(1): 5.104073 (4.1717875)

loadings for random effect 1

lnlead: -.34214646 (.06614347)

Las instrucciones siguientes despliegan los datos anteriores, solo cambia a la salida anterior en que da las varianzas-covarianzas y correlación entre los

efectos aleatorios, en este caso *subno* y *lnlead*.

```
. eq modelo:lnlead
```

```
. eq intercep: constante
```

```
gllamm crecio lnlead, link(logit) family(binomial) i(subno1) nrf(2)
eq(intercep modelo)
```

number of level 1 units = 1769 number of level 2 units = 174

Condition Number = 21.571489

```
gllamm model
```

log likelihood = -273.15002

crecio	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnlead	.3520107	.3447327	1.02	0.307	-.323653	1.027674
_cons	-4.257334	.7982744	-5.33	0.000	-5.821923	-2.692744

Variances and covariances of random effects

***level 2 (subno1)

var(1): 5.1040397 (4.1716836)

cov(1,2): -1.7463236 (1.6448042) cor(1,2): -1

var(2): .59749657 (.64930675)

Bibliografía

- [1] McCullach P. y Nelder J. A. *Generalized Linear Models*. Chapman and Hall 1983. Second edition.
- [2] Dobson, Annette J. *An introduction to generalized linear models*. Chapman and Hall/CRC 2002. Second edition.
- [3] McCullach Charles E., Searle, Shayle R. *Generalized, Linear and Mixed Models*. Willey Intercience publication 2001.
- [4] Pinheiro, José C. y Bates, Douglas M. *Mixed—Effects models in S and S-Plus*. Springer 2000.
- [5] Diggle Peter J., Heagerty, Patrick, Liang, Kung-Yee, y Zeger, Scott L. *Analysis of Longitudinal Data*. Oxford 2002. Second edition.
- [6] Breslow, N. E.; Clayton D. G. Aproximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistics Association*, Vol.88, No. 421(Mar.,1993), 9-25.
- [7] Zhiying Pan and D. Y. Lin. (2005). Goodness-of-Fit Methods for Generalized Linear Mixed Models. *Biometrics*, 61, 1000-1009.
- [8] Cox, D. R. y Reid, N. (1987). Parameter Orthogonality and Approximate Condicional Infrence. *Journal of the Royal Statistical Society, Ser. B*, 49, 1-39.

- [9] Lindstrom, M. J., y Bates, D. M. (1990) Nolinear Mixed Effects Models for Repeated Measures Data. *Biometrics*, 46,673-687.
- [10] Zeger, S. L. (1988). A Regresssion Model for Time Series of Counts, *Biometrika*, 75, 621-629.
- [11] Ignacio Méndez y Hotensia Moreno. Razones de momoios y cocientes de probabilidades. Conceptos y estimación con modelos logísticos. *Mono-grafías*. Vol.13, No.29, noviembre de 2005.
- [12] <http://www.bepress.com/ucbbiostat/paper160>
GLLAMM Manual. Sophia Rabe-Hesketh, Anders Skrondal, Andrew Pickles.