ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

FACULTAD DE CIENCIAS NATURALES Y MATEMATICAS
DEPARTAMENTO DE MATEMÁTICAS

EXAMEN COMPLEXIVO

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

"MAGÍSTER EN INVESTIGACIÓN MATEMÁTICA"

TEMA:

ESTIMACIÓN DE LA FUNCIÓN LOG-VEROSIMILITUD DEL MODELO LINEAL MIXTO

AUTOR:
MANUEL PABLO ÁLVAREZ ZAMORA

Guayaquil - Ecuador

2015

DEDICATORIA

A mis hijos por ser el aliento para culminar con éxito este Proyecto de Investigación, el mismo que es resultado del esfuerzo de varios años de trabajo.

Pablo Álvarez

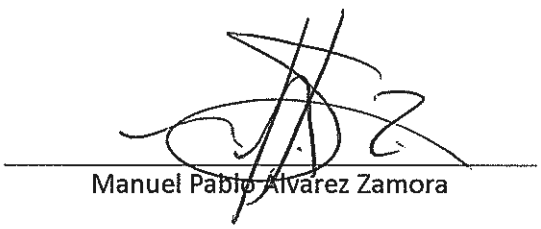
AGRADECIMIENTO

AL Ph.D. Francisco Vera Alcívar por ser mi guía en el desarrollo de este Proyecto de Investigación Científica.

Pablo Álvarez

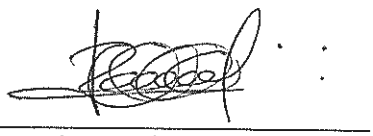
DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este proyecto de examen complexivo, me corresponde exclusivamente; el patrimonio intelectual del mismo, corresponde exclusivamente a la **Facultad de Ciencias Naturales y Matemáticas, Departamento de Matemáticas** de la Escuela Superior Politécnica del Litoral.

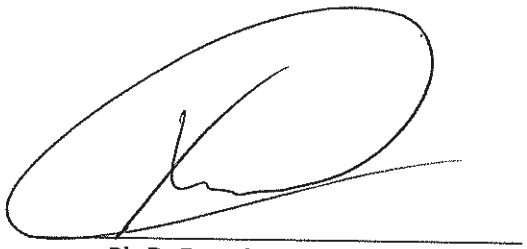


Manuel Pablo Álvarez Zamora

TRIBUNAL DE GRADUACIÓN



M.Sc. John Ramírez Figueroa
PRESIDENTE DEL TRIBUNAL



Ph.D. Francisco Vera Alcívar
DIRECTOR DEL EXAMEN COMPLEXIVO

ÍNDICE GENERAL

ABREVIATURAS

MLE : ……Máxima verosimilitud

LME : ……Modelo Lineal de efectos aleatorios

SS : ……Suma de cuadrados

GLS : ……Mínimos cuadrados generalizados

RML: ……Modelo Lineal Residual

OLS : ……Mínimos Cuadrados Ordinario

AIC : ……Criterio de información de AKAIKE=-2 l + k*m

L : ……..log(verosimilitud, k=2, m= número de parámetros

ANOVA : ……Análisis de Varianzas

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

OBJETIVOS GENERALES

1.- Difundir las técnicas para resolver los modelos lineales mixtos a la comunidad científica de la ESPOL y del país.

2.- Promover la cultura con el rigor científico de las matemáticas.

OBJETIVOS ESPECÍFICOS

1.- Utilizar un modelo de curva de crecimiento para determinar la validez de la dosis suministrada en los pollitos en un ejemplo de producción avícola.

2.- Realizar un documento científico de consulta en lenguaje español para que nuestros investigadores sean capaces de leer otros artículos científicos que usan estas técnicas y así poder replicar dichas investigaciones en nuestro país Ecuador.

INTRODUCCIÓN

Los modelos mixtos se están usando en muchas situaciones como en el área de ecología en donde se requiere ver el comportamiento de los individuos de una familia y sometido a situaciones predictivas.

Para comenzar revisaremos el modelo de coeficientes aleatorios para después tratar con otros modelos como lo son el modelo de curva de crecimiento y el modelo de curva de crecimiento rectangular y otros. Estos modelos pueden ser tratados por investigadores de otras áreas y la intención es que este trabajo sea una fuente de consulta para aplicar en temas dentro del campo de la agricultura, médico, educativo etc.

Este trabajo también contribuye para que los investigadores puedan entender las publicaciones de otros investigadores. Especialmente se incluye el apéndice que presenta conceptos y teoremas básicos de algebra matricial que permiten comprender el desarrollo

ESTIMACIÓN DE MÁXIMA VEROSIMILITUD DE LOS PARÁMETROS DE UN MODELO MIXTO.

I. MODELO LINEAL DE EFECTOS MIXTOS.

El modelo lineal de efectos mixtos (lme)

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i, \qquad \mathrm{i} = 1, \dots, \mathrm{N} \qquad (1)$$

$$\begin{bmatrix} y_{i,1} \\ \vdots \\ y_{i,n_i} \end{bmatrix} = \begin{bmatrix} X_{i,11} & \vdots & X_{i,1m} \\ & & \\ X_{i,n_i1} & & X_{i,n_im} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} Z_{i,11} & \vdots & Z_{i,1k} \\ & & \\ Z_{i,n_i1} & & Z_{i,n_ik} \end{bmatrix} \begin{bmatrix} b_{i,1} \\ \vdots \\ b_{i,k} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i,1} \\ \vdots \\ \varepsilon_{i,n_i} \end{bmatrix}$$

Donde

y_i es un vector n_ix1 (cluster)

X_i es una matriz $n_i$xm (matriz de diseño o efectos fijos)

β es un vector mx1 de parámetros poblacionales (coeficientes de efectos fijos)

Z_i es una matriz de diseño $n_i$xk de efectos aleatorios

ε_i es un vector n_ix1, término del error con componentes independientes

b_i vector kx1 de efectos aleatorios con media cero y matriz de covarianza $D_* = \sigma^2 D$

Se supone que los vectores $b_{i,}$y ε_i , $\mathrm{i} = 1, \dots, \mathrm{N}$ son mutuamente independientes

(1) Modelo lineal mixto: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

(2)

I.I. MODELO LINEAL CON INTERCEPTO ALEATORIO

Si el modelo lineal con intercepto aleatorio es escrito en la forma

$$W_i = a_i + \beta H_i + \varepsilon_i,$$

donde ε_i es un vector aleatorio con media 0 y componentes no correlacionados con varianza constante σ^2, esto es, $var(\varepsilon_i) = \sigma^2 I$ remplazando $a_i\ por\ \alpha + b_i$, llegamos al modelo lineal con efectos mixtos

$$W_i = \propto + \beta H_i + b_i Z_i + \varepsilon_i, \qquad \mathrm{i} = 1, \dots, \mathrm{N} \qquad (2.2)$$

Para este caso $Z_i = 1_i = (1,1, \dots, 1)'$ es un vector columna de dimensión n_i,

W_i es un vector $n_i x 1$

H_i es un vector $n_i x 1$

ε_i es un vector $n_i x 1$

b_i es un vector $kx1, para\ el\ caso\ presentado, k = 1$

$var(b_i) = \sigma_d^2 = \sigma^2 D$, $donde\ D\ es\ kxk$

$cov(b_i, \varepsilon_{ij}) = 0,\ porque\ se\ consideran\ independientes$

β es un vector $mx1, para\ este\ caso\ es\ escalar$

$$\begin{bmatrix} W_{i1} \\ \vdots \\ W_{in_i} \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \beta \begin{bmatrix} H_{i1} \\ \vdots \\ H_{in_i} \end{bmatrix} + b_i \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix}, \qquad (2.2) forma\ vectorial$$

La función log(verosimilitud) para el modelo lme es más simple si se usa la matriz de covarianza escalada,

$$D = \frac{1}{\sigma^2} D_* = \frac{1}{\sigma^2} cov(b_i).$$

Volviendo al ejemplo de los datos de la familia, el modelo para la relación entre pesos y alturas es un caso especial de LME (2.4) con un efecto aleatorio (k=1), donde $y_i = W_i$,

$$\beta = \begin{bmatrix} \alpha \\ \beta^* \end{bmatrix},\ X_i = [1_i\ H_i],\ Z_i = 1_i$$

Las N ecuaciones pueden ser arregladas en una ecuación así

$$y = X\beta + Zb + \varepsilon \qquad (2.7)$$

Después se apilan los datos en un vector simple y forma de matriz como sigue:

$$y: N_T\ x\ 1, \quad X: N_T\ x\ m, \quad Z: N_T\ x\ Nk, \quad b: Nk\ x\ 1, \quad \varepsilon: N_T\ x\ 1$$

Donde $N_T = \sum_{i=1}^{N} n_i$, y la $cov(b) = \sigma^2 (I \otimes D)$.

$$y = \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix}, \qquad X = \begin{bmatrix} X_1 \\ \dots \\ X_N \end{bmatrix}, \qquad Z = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & Z_N \end{bmatrix}$$

$$b = \begin{bmatrix} b_1 \\ \dots \\ b_N \end{bmatrix}, \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_N \end{bmatrix}$$

El modelo (2.7) [1] puede ser rescrito así

$$y = X\beta + \eta, \qquad (2.10)$$

Donde, $\eta = \begin{bmatrix} \eta_1 \\ \cdots \\ \eta_N \end{bmatrix} = \begin{bmatrix} Z_1 b_1 + \varepsilon_1 \\ \cdots \\ Z_N b_N + \varepsilon_N \end{bmatrix}$ (2.11)

El valor esperado de η E(η)=0, y la matriz de covarianzas de η tiene forma diagonal por bloques debido a que se supone que las variables aleatorias de diferentes familias no están correlacionados.

En esta sección un modelo LME muy especial con un efecto aleatorio es considerado en detalle, el modelo LME con interceptos aleatorios. Este modelo servirá como un punto de referencia para comparar estimadores y procedimientos numéricos para la maximización de log-verosimilitud.

El modelo LME con interceptos aleatorios (2) es escrito como

$$y_{ij} = a_i + \gamma' u_{ij} + \varepsilon_{ij}, \;\; j = 1, \dots, n_i, i = 1, \dots, N, \quad (2.63)$$

Donde y_{ij} es interpretado como la j_ésima observación del i_ésimo sujeto. El intercepto individual es la suma de un parámetro promediado α y un efecto aleatorio, $a_i = \alpha + b_i$. Si se supone que $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ y $b_i \sim \mathcal{N}(0, \sigma^2 d)$, son independientes, donde σ^2 es la varianza dentro del sujeto y d es la escala de la varianza del efecto aleatorio, El caso más simple del modelo (2.63) , cuando no existe covariación, se reduce a un modelo de componentes de varianza (1.8).

El modelo LME con interceptos aleatorios individuales puede emerger en la siguiente colocación longitudinal. Considerar un grupo de N pacientes, cada paciente tiene su propia covariación, tales como edad, género, peso, dieta, rendimiento físico, etc. Sea el tratamiento llegar a ser representado por una variable longitudinal, tal como la presión sanguínea y_{ij} del i_ésimo paciente en el tiempo t_{ij}. Luego, el vector de covarianzas es $u_{ij} = \left(Age_{ij},\ Gender_i, Weight_{ij}, Diet_{ij}\right)$. Si las observaciones cubren un largo periodo de tiempo, Edad, Peso y Dieta varían con el tiempo, así que ellas son suministradas con subíndice j. Estamos interesados en como la presión sanguínea es afectada por estas variables. La clave de estos supuestos es que la relación no varía de paciente a paciente, lo que significa es que γ es un vector fijo. Sin embargo, admitimos que en el tiempo cuando comienza el estudio, cada paciente puede tener una presión sanguínea con línea de base diferente, aun para pacientes de la misma edad, género, peso y dieta. Ciertamente, uno podría diseñar el experimento, lo cual involucra pacientes con las mismas condiciones iniciales, por ejemplo la misma edad, género, peso y dieta. Sin embargo el lector estará de acuerdo en cuan difícil sería obtener tales datos. Un modelo con interceptos individuales sería más realista porque aquello permite el análisis de pacientes con presión sanguínea con línea de base diferente.

El modelo LME con interceptos aleatorios también emerge en Econometría como modelo para datos de panel en un análisis seccional cruzado, ver Maddala (1987) para una revisión. En la literatura de componentes de varianzas este modelo es llamado modelo lineal con estructura de error nested, por ejemplo, Christensen (1996), Wan and Ma (2002).

(2) Modelo lineal conInterceptos aleatorios: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

En notación matricial, el modelo LME con interceptos aleatorios puede ser escrito como

$$y_i \sim \mathcal{N}\left(X_i\beta, \sigma^2(I_i + d1_i1_i')\right), \qquad \text{i} = 1, \dots, \text{N} \quad (2.64).$$

Donde X_i *es una matriz* $n_i x m$ *de rango completo* con la i_ésima fila $x_{ij}' = (1, u_{ij}')$; $\beta = (\alpha, \gamma')'$ es un vector mx1 de efectos fijos; 1_i es un vector unidad de orden $n_i x 1$; σ^2 es la varianza dentro del sujeto de los efectos aleatorios. En la notación de (2.13), k=1, $1_i = Z_i$, d=D. El modelo (2.63) tiene una estructura de correlación intercambiable (compuesto simétricamente) porque los coeficientes de correlación entre $y_{ij} = y_{ik}$ es constante, $\frac{d}{1+d}$ $para\ j \neq k$. En efecto, nuestro modelo LME con datos de familia de pesos y alturas de la sección (2.1) tiene la forma (2.63).

La función log-verosimilitud de varianza perfilada, escrita en la forma (2.38), es

$$l_p(\beta, d) = -\frac{1}{2}\left\{N_T \ln \sum_{i=1}^{N}[(y_i - X_i\beta)'(I_i + d1_i1_i')^{-1}(y_i - X_i\beta)] + \sum_{i=1}^{N}\ln|I_i + d1_i1_i'|\right\} \quad (2.65)$$

Donde $e_i = e_i(\beta) = y_i - X_i\beta$ es el vector residual $n_i x 1$. Las fórmulas de reducción de dimensión (2.19) y (2.20) simplifican

$$\ln|I_i + d1_i1_i'| = 1 + n_i d, (I_i + d1_i1_i')^{-1} = I - \frac{d}{1 + n_i d}1_i1_i'. \quad (2.66)$$

Así, el estimador GLS (2.26) para este modelo es

$$\hat{\beta}_{GLS} = \left(\sum_{i=1}^{N} X_i'X_i - \frac{n_i^2 d}{1 + n_i d}\bar{x}_i\bar{x}_i'\right)^{-1}\left(\sum_{i=1}^{N} X_i'y_i - \frac{n_i^2 d}{1 + n_i d}\bar{x}_i\bar{y}_i'\right) \quad (2.67)$$

donde

$$\bar{x}_i = \frac{1}{n_i}X_i'1_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}\,, \qquad \bar{y}_i = \frac{1}{n_i}y_i'1_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_j.$$

El estimador de efectos fijos (2.29) es el límite de (2.67) cuando $d \to \infty$. Notamos que la matriz inversa en (2.67) llega a ser singular cuando $d \to \infty$ porque la primera fila y columna son cero. Podríamos usar la inversa generalizada como en (2.29), pero es más fácil minimizar la forma cuadrática (2.31) como una función de N+m parámetros $a_1, \dots, a_N, \gamma$ directamente,

$$S = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - a_i - u'_{ij}\gamma)^2 \rightarrow min.$$

Derivando S con respecto a a_i y luego igualando a 0, se tiene,

$$\frac{\delta S}{\delta a_i} = \sum_{j=1}^{n_i}(y_{ij} - a_i - u'_{ij}\gamma)(-2) = 0$$

$$n_i\bar{y}_i - n_i a_i - n_i\bar{u}'_i\gamma = 0$$

El mínimo es obtenido en $\hat{a}_i = \bar{y}_i - \hat{\gamma}'_{\infty}\bar{u}_i$, lo cual llega al estimador de efectos fijos (FE),

$$S = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i + \bar{u}_i{}'\gamma - u'_{ij}\gamma)^2$$

$$S = \sum_{i=1}^{N}\sum_{j=1}^{n_i}((y_{ij} - \bar{y}_i) - (u'_{ij} - \bar{u}_i{}')\gamma)^2$$

$$S = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left((y_{ij} - \bar{y}_i)^2 - 2(y_{ij} - \bar{y}_i)(u'_{ij} - \bar{u}_i{}')\gamma + \left((u'_{ij} - \bar{u}_i{}')\gamma\right)^2\right)$$

$$S = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 - 2\sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)(u'_{ij} - \bar{u}_i{}')\gamma + \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left((u'_{ij} - \bar{u}_i{}')\gamma\right)^2$$

$$S = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 - 2\sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)(u'_{ij} - \bar{u}_i{}')\gamma + \sum_{i=1}^{N}\sum_{j=1}^{n_i}(u'_{ij} - \bar{u}_i{}')\gamma'(u'_{ij} - \bar{u}_i{}')\gamma$$

$$S = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 - 2\sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)(u'_{ij} - \bar{u}_i{}')\gamma + \sum_{i=1}^{N}\sum_{j=1}^{n_i}\gamma'(u_{ij} - \bar{u}_i)(u'_{ij} - \bar{u}_i{}')\gamma$$

Derivando con respecto a γ, e igualando a 0, se tiene

$$\frac{\delta S}{\delta \gamma} = -2\sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)(u_{ij} - \bar{u}_i) + \sum_{i=1}^{N}\sum_{j=1}^{n_i}2(u_{ij} - \bar{u}_i)(u'_{ij} - \bar{u}_i{}')\gamma = 0$$

$$-\sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij}u_{ij} - \bar{y}_i u_{ij} - y_{ij}\bar{u}_i + \bar{y}_i\bar{u}_i) + \left(\sum_{i=1}^{N}\sum_{j=1}^{n_i}(u_{ij}u'_{ij} - \bar{u}_i u'_{ij} - u_{ij}\bar{u}_i{}' - \bar{u}_i\bar{u}_i{}')\right)\gamma = 0$$

$$-\sum_{i=1}^{N}\left[\left(\sum_{j=1}^{n_i}(y_{ij}u_{ij})\right)-n_i(\bar{y}_i\bar{u}_i+\bar{y}_i\bar{u}_i-\bar{y}_i\bar{u}_i)\right]$$
$$+\left(\sum_{i=1}^{N}\left[\left(\sum_{j=1}^{n_i}(u_{ij}u'_{ij})\right)-n_i(\bar{u}_i\bar{u}_i{}'-\bar{u}_i\bar{u}_i{}'+\bar{u}_i\bar{u}_i{}')\right]\right)\gamma=0$$

$$-\sum_{i=1}^{N}\left[\left(\sum_{j=1}^{n_i}(y_{ij}u_{ij})\right)-n_i(\bar{y}_i\bar{u}_i)\right]+\left(\sum_{i=1}^{N}\left[\left(\sum_{j=1}^{n_i}(u_{ij}u'_{ij})\right)-n_i(\bar{u}_i\bar{u}_i{}')\right]\right)\gamma=0$$

$$\hat{\gamma}_{\infty}=\left(\sum_{i=1}^{N}\left(\sum_{j=1}^{n_i}(u_{ij}u'_{ij})-n_i\bar{u}_i\bar{u}'_i\right)\right)^{-1}\left(\sum_{i=1}^{N}\left(\sum_{j=1}^{n_i}(u_{ij}y_{ij})-n_i\bar{u}_i\bar{y}_i\right)\right). \quad (2.68)$$

Este estimador podría ser obtenido centrando las observaciones alrededor de las medias individuales para cada i y aplicando OLS. Interesantemente, el estimador $\hat{\gamma}_{\infty}$ es consistente a pesar del hecho que el número de parámetros ruidosos (interceptos de sujetos específicos) se incremente con el número de cluster $N\to\infty$. En efecto esto es verdad solo para modelos lineales y nosotros estableceremos similares propiedades asintóticas para el modelo de curva de crecimiento general, como un compromiso entre los modelos de efectos fijos y aleatorios. Contrario al modelo lineal, como aprenderemos en la sección 7.2.2, el LME para el modelo de regresión logística con intercepto de sujeto específico fijo no es consistente, pero la máxima verosimilitud condicional es –aunque, como será mostrado en la subsección 3.2.1, la presencia simultánea de efectos fijos y mixtos lleva a un modelo estadístico inválido.

En un caso especial del modelo VARCOMP (1.8), nosotros obtenemos

$$\hat{\alpha}_{GLS}=\frac{\sum_{i=1}^{N}\frac{n_i}{1+n_id}\bar{y}_i}{\sum_{i=1}^{N}\frac{n_i}{1+n_id}},\quad \hat{\alpha}_{\infty}=\frac{1}{N}\sum_{i=1}^{N}\bar{y}_i \quad (2.69)$$

Note que estas estimaciones coinciden cuando $n_i$es constante, esto es, cuando el diseño está balanceado.

Recordemos que la aproximación de los efectos fijos supone que los interceptos son fijos y diferentes para cada persona. Por el contrario, el modelo de efectos mixtos supone que a_i son independientes idénticamente distribuidos (iid) variables aleatorias con la misma media, α. Una versión económica de la función log-verosimilitud varianza perfilada, basada en las fórmulas de reducción de dimensión (2.66), tiene la forma

$$l_p(\beta,d)=-\frac{1}{2}\left\{N_T\ln\left(S-d\sum_{i=1}^{N}\frac{n_i^2h_i^2}{1+n_id}\right)+\sum_{i=1}^{N}\ln(1+n_id)\right\} \quad (2.70)$$

Y como un caso especial de (2.50) para el ML restringido

$$l_{Rp}(\beta,d) = -\frac{1}{2}\left\{(N_T - \mathrm{m})\ln\left(S - d\sum_{i=1}^{N}\frac{n_i^2 h_i^2}{1+n_i d}\right) + ln\left|\sum_{i=1}^{N} X_i' X_i - \frac{n_i^2 d}{1+n_i d}\bar{x}_i\bar{x}_i'\right| + \sum_{i=1}^{N}\ln(1+n_i d)\right\} \quad (2.71)$$

Donde el escalar S y $h_i$se definen como

$$S = S(\beta) = \sum_{i=1}^{N}\|y_i - X_i\beta\|^2,$$

$$h_i = h_i(\beta) = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} - X_{ij}'\beta = \bar{y}_i - \beta'\bar{x}_i. \quad (2.72)$$

La función log-verosimilitud lp puede ser simplificada para datos balanceados ($n_i = n$). En este caso la función admite una solución de forma cerrada para el máximo sobre d. En efecto, si $n_i = n$, entonces (2.70) toma la forma

$$-\frac{1}{2}\left\{Nn\ln\left(S - d\frac{n^2 A}{1+nd}\right) + N\ln(1+nd)\right\}$$

Donde $A = \sum_{i=1}^{N} h_i^2$. El MLE para d vuelve la derivada parcial a 0, lo cual da

$$d = \frac{n^2 A - S}{n(S - nA)} = — \quad (2.73)$$

Abajo consideraremos el caso cuando un modelo lineal con interceptos aleatorios admite una solución de forma cerrada. Una comprensiva cobertura del modelo intercepto aleatorio con igual número de observaciones por cluster ($n_i = n$) con aplicaciones econométricas puede ser encontrado en un reciente libro publicado por Hsiao (2003).

Modelo intercepto aleatorio balanceado (3)

En esta sección consideramos el caso cuando $n_i = n$ y $X_i = X$, los datos balanceados. Nuestro propósito es obtener expresiones cerradas para el ML estimado como hicimos en la sección 2.3. Primero se muestra que la estimación GLS (2.67) no depende de d y coincide con la estimación OLS.

Si $X_i = X$toma la forma

$$\hat{\beta}_{GLS} = \left(X'X - \frac{n^2 d}{1+nd}\bar{x}\bar{x}'\right)^{-1}\left(X'\bar{y} - \frac{n^2 d}{1+nd}\bar{x}\bar{\bar{y}}\right) \quad (2.74)$$

Donde $\bar{\bar{y}} = \bar{y}'1/n$. (VER ANEXO INTERCEPTO ALEATORIO BALANCEADO)

(3) Modelo lineal intercepto aleatorio balanceado: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

¿Cómo los efectos aleatorios afectan a la varianza de MLE?

Antes de avanzar al negocio de efectos mixtos es importante conocer como la presencia de un efecto aleatorio afectaría la varianza de la máxima verosimilitud estimada. ¿La introducción de los efectos aleatorios reduce o incrementa la varianza de los estimados de los mínimos cuadrados ordinarios? ¿Es la varianza del MLE una función creciente o decreciente de la varianza del efecto aleatorio? Responderemos estas preguntas para un modelo balanceado con intercepto aleatorio donde existe una solución en forma cerrada. Encontraremos como la varianza escalada del efecto aleatorio d afecta la matriz de covarianza de los mínimos cuadrados generalizados estimados (2.74) en el modelo intercepto aleatorio balanceado.

La matriz de covarianzas para el estimador GLS está dado por

$$cov(\hat{\beta}_{GLS}) = \sigma^2 \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1}.$$

Para el modelo intercepto aleatorio balanceado, la matriz de covarianza de (2.74) se reduce a

$$cov(\hat{\beta}_{GLS}) = \frac{\sigma^2}{N} (X'V^{-1}X)^{-1}$$

$$= \frac{\sigma^2}{N} \left(X'X - \frac{n^2 d}{1+nd} \bar{x}\bar{x}' \right)^{-1}$$

$$= \frac{\sigma^2}{N} ((X'X)^{-1} + n^2 d (X'X)^{-1} \bar{x}\bar{x}' (X'X)^{-1})$$

$$= \frac{\sigma^2}{N} \left((X'X)^{-1} + \frac{n^2}{n^2} d (X'X)^{-1} (X'1_n)(X'1_n)' (X'X)^{-1} \right)$$

$$= \frac{\sigma^2}{N} ((X'X)^{-1} + d (X'X)^{-1} X'1_n \, ((X'X)^{-1} X'1_n)')$$

$$= \frac{\sigma^2}{N} \left((X'X)^{-1} + d \begin{pmatrix} 1 \\ 0_{m-1} \end{pmatrix} \begin{pmatrix} 1 \\ 0_{m-1} \end{pmatrix}' \right)$$

$$= \frac{\sigma^2}{N} (X'X)^{-1} + \frac{d_*}{N} \begin{bmatrix} 1 & 0_{m-1}{}' \\ 0_{m-1} & 0_{(m-1)x(m-1)} \end{bmatrix}.$$

Debido a la sentencia a) del lema 3.

Así, se puede inferir lo siguiente:

La varianza del intercepto aleatorio en el modelo intercepto aleatorio con datos balanceados afecta solamente a la varianza del término intercepto.

La varianza de la pendiente no cambia con d_*.

Se puede interpretar este resultado diciendo que la correlación igual por datos balanceados no afecta la estimación de la pendiente: a) El estimado GLS/ML no cambia; b) la varianza del estimado no cambia (el mismo resultado se obtiene con regresión de Poison). Se aclara que esto es verdad solo para datos balanceados. Para ilustrar esto, consideremos el ejemplo previo con datos de familia desbalanceados.

Ejemplo (continuación). Consideremos los datos de las familias de la sección 2.1, donde el modelo LME (2.2) con igual correlación dentro de las familias tiene la forma del modelo intercepto aleatorio (2.63). Los datos no están balanceados y queremos conocer si el GLS estimado (2.67) o su varianza dependen de d. En la figura 2.1 ploteamos la pendiente de la altura y su SE como una función de d. Ya que los datos no están balanceados, $\hat{\beta}_{GLS}\ y\ SE(\hat{\beta}_{GLS})$ cambian con d. Dos extremos son el OLS (d=0) y el estimador FE (d=∞). Note que la diferencia entre los valores extremos de la pendiente es bastante pequeña.

Una forma atractiva del modelo intercepto aleatorio que es la prueba exacta están disponibles para una hipótesis lineal sobre los coeficientes β, ver la sección 3.8 [1] para detalles.

Hacemos unos comentarios para las varianzas de los parámetros en el ML restringido estimado. Generalmente, RML son cerrados para el ML estándar estimado. La diferencia está en los grados de libertad del denominador. Para σ^2, el denominador es ajustado por el número de pendientes de efectos fijos (m-1) y para d por 1. Como aprenderemos en la sección 3.14 el RMLE para el modelo intercepto aleatorio con datos balanceados son insesgados y coincide con otros estimadores cuadráticos insesgados: norma mínima, método de los momentos y cuadrados de mínima varianza. Interesantemente, para el modelo de coeficientes aleatorios balanceados de la sección precedente, ML=RML para σ^2, pero ellos son diferentes para el modelo interceptos aleatorios balanceado.

I.II. MODELO DE COEFICIENTE ALEATORIO BALANCEADO (4)

El modelo lineal de efectos mixtos LME se denomina balanceado si n_i= constante y las matrices de diseño de los efectos aleatorios Z_i es la misma para todos los cluster (sujetos), $Z_i = Z$.

$$y_i = X_i\beta + Zb_i + \varepsilon_i, \qquad \mathrm{i} = 1, \dots, \mathrm{N} \qquad \mathrm{ver}(2.4)$$

Llamamos un modelo de coeficientes aleatorio balanceado a:

$$y_i = X_i a_i + \varepsilon_i, \qquad \mathrm{i} = 1, \dots, \mathrm{N} \qquad (1.12)$$

Donde $a_i = \beta + b_i = A_i\beta^* + b_i$

Si además la matriz de diseño de efectos fijos es la misma para todos los $X_i = X$ individuales, el modelo de coeficientes aleatorios es balanceado.

$$y_i = X\beta + Xb_i + \varepsilon_i, \qquad \mathrm{i} = 1, \ldots, \mathrm{N}$$

(4)Modelo de coeficiente aleatorio balanceado: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

II. FUNCIÓN DE LOGARITMO DE LA VERSOSIMILITUD CON EL SUPUESTO DE NORMALIDAD.

Función log-verosimilitud

$$l(\theta) = \ln\big(L(\beta, \sigma^2, D)\big)$$

$$l(\theta) = \{-\frac{N_T}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln\big|\sigma^2\big(I_{n_i} + Z_i D Z_i'\big)\big| -$$

$\frac{1}{2}\sigma^{-2}\sum_{i=1}^{N}[(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta)]\}$

$$l(\theta) = \{-\frac{N_T}{2}\ln(2\pi) - \sum_{i=1}^{N} ln(\sigma^2)^{n_i} + \ln\big|\big(I_{n_i} + Z_i D Z_i'\big)\big| - \frac{1}{2}\sigma^{-2}\sum_{i=1}^{N}[(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta)]\}$$

Dejando a un lado el término constante $C = -\left(\frac{N_T}{2}\right)\ln(2\pi)$, debido a que no afecta al proceso de maximización, la función log-verosimilitud que consideraremos para el modelo LME está dada por

$$l(\theta) =$$

$-\frac{1}{2}\{\sum_{i=1}^{N} n_i \ln\sigma^2 + \sum_{i=1}^{N}[\ln\big|I_{n_i} + Z_i D Z_i'\big| + \sigma^{-2}(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta)]\}$

$l(\theta) = -\frac{1}{2}\{N_T \ln\sigma^2 + \sum_{i=1}^{N}[\ln\big|I_{n_i} + Z_i D Z_i'\big| + \sigma^{-2}(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta)]\}$ (2.14)

Donde $\theta = \big(\beta', \sigma^2, vech'(D)\big)'$ *es el vector combinado de parámetros desconocidos.*

En lo sucesivo vech(D) denota el vector k(k+1)/2 x 1 de elementos únicos de la matriz simétrica D de orden kxk, (Magnus 1988). Por lo tanto, la dimensión total del vector de parámetros es m+1+k(k+1)/2. El estimador de máxima verosimilitud mle maximiza la función l sobre el espacio de parámetros

$\Theta = \{\theta: \beta \in \mathbb{R}^m, \sigma^2 > 0, D\ es\ definida\ no\ negativa\}$.

Note que buscamos un estimado de la matriz D en el conjunto de todas las matrices definida no negativa, el cual será denotado D_+ con una frontera finita, donde la matriz D es singular. Existe una aproximación alternativa tomada por Rao y Kleffe (1988), quienes supusieron que la matriz D puede ser definida negativa siempre que la matriz (2.12) sea definida positiva. Esto es verdad si y solo si, todas las matrices $I_{n_i} + Z_i D Z_i'$ son definidas positivas. Sin embargo el supuesto de la definición de no negatividad de la matriz D es estadísticamente razonable debido que D es una matriz de covarianzas. Estrictamente hablando, nosotros tratamos con un problema de maximización restringido no lineal, lo cual complica tremendamente a los

algoritmos numéricos. En efecto la probabilidad de dar con la frontera de Θ es muy alta, especialmente para N relativamente pequeño; los detalles se ven en 2.15.

La omisión del termino C en (2.14) no afecta a la maximización de l y estimación de ML.

La matriz de covarianzas escalada del vector de la variable dependiente y_i está dado por:

$$V_i = V_i(D) = I_{n_i} + Z_i D Z_i' \qquad (2.16)$$

Esta notación es usada a través del documento. Los efectos aleatorios inducen correlación dentro del cluster (grupo) entre componentes del vector y, debido a que la matriz V_i es no diagonal.

Usando la notación V_i, la función log-verosimilitud (2.14) puede ser re-escrita como

$$l(\theta) = -\frac{1}{2}\left\{N_T \ln \sigma^2 + \sum_{i=1}^{N}\left[\ln|V_i| + \sigma^{-2} e_i{}' V_i{}^{-1} e_i\right]\right\}, \quad (2.15)$$

Donde $e_i = e_i(\beta) = y_i - X_i\beta, \qquad (2.18)$

es un vector residual $n_i x 1$, del i_ésimo cluster, i=1,…,N.

II.I. ESTIMACIÓN DE LA MÁXIMA VEROSIMILITUD

De forma similar puede llegar al otro determinante.

Usando las fórmulas de reducción de dimensión (ver anexos), la función log-verosimilitud (2.14) podemos re-escribirla en una forma equivalente y más económica como

$$l(\theta) = -\frac{1}{2}\{N_T \ln \sigma^2 + \sum_{i=1}^{N}[\ln|I_k + D Z_i' Z_i|$$

$$+\sigma^{-2}\sum_{i=1}^{N}[(y_i - X_i\beta)'(I_{n_i} - Z_i(I_k + D Z_i' Z_i)^{-1} D Z_i')(y_i - X_i\beta)]\}$$

Considerando $S_i = e_i(\beta) = e_i' e_i = \|y_i - X_i\beta\|^2$, la suma de cuadrados residuales del i_ésimo cluster. (2.25)

$$l(\theta) = -\frac{1}{2}\{N_T \ln \sigma^2 + \sum_{i=1}^{N}[\ln|I_k + D Z_i' Z_i|$$

$$+\sigma^{-2}\textstyle\sum_{i=1}^{N}[S_i - (Z_i' e_i)'(I_k + D Z_i' Z_i)^{-1} D(Z_i' e_i)]. \quad (2.24)$$

Una forma característica del modelo (2.13) es que manteniendo la matriz D constante, l es maximizado por el estimador de mínimos cuadrados generalizado (GLS).

$$\hat{\beta}_{GLS} = \left(\textstyle\sum_{i=1}^{N} X_i'(I_{n_i} + Z_i D Z_i')^{-1} X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'(I_{n_i} + Z_i D Z_i')^{-1} y_i\right). \quad (2.26)$$

Se puede mostrar que la matriz $\sum_i X_i'(I_{n_i} + Z_i D Z_i')^{-1} X_i$ es definida positiva si la matriz $\sum_i X_i' X_i$ es no singular como se supuso al comienzo de la sección. En un caso especial cuando D=0, el estimador GLS colapsa al estimador de mínimos cuadrados ordinario (OLS),

$$\hat{\beta}_{OLS} = \left(\sum_{i=1}^{N} X_i'(I_{n_i} + 0)^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i'(I_{n_i} + 0)^{-1} y_i\right). \text{ (2.27)}$$

Otro caso extremo es cuando la matriz D tiende a infinito $D = dI, d \to \infty$. Luego, usando la fórmula de reducción de dimensión (2.19) y denotando $\delta = 1/d \to 0$, obtenemos

$$\lim_{d\to\infty} V_i^{-1} = \lim_{d\to\infty} (I_{n_i} - Z_i(D^{-1} + Z_i' Z_i)^{-1} Z_i')$$

$$\lim_{d\to\infty} V_i^{-1} = \lim_{d\to\infty} (I_{n_i} - Z_i(d^{-1} I_k + Z_i' Z_i)^{-1} Z_i')$$

$$\lim_{d\to\infty} V_i^{-1} = I_{n_i} - Z_i \lim_{\delta\to 0} (\delta I_k + Z_i' Z_i)^{-1} Z_i'$$

Pero el último límite es la matriz inversa generalizada de de Moore-Penrose (Albert 1972), tal que

$$Z_i^{+} = \lim_{\delta\to 0} (\delta I_k + Z_i' Z_i)^{-1} Z_i', \quad \text{(2.28)}$$

Note que si la matriz tiene rango completo, la matriz $Z_i' Z_i$ es no singular y

$Z_i^{+} = (Z_i' Z_i)^{-1} Z_i'$. Así, el estimador GLS (OLS) para la matriz infinita D llega a ser

$$\hat{\beta}_{\infty} = \left(\sum_{i=1}^{N} X_i' (I_{n_i} - Z_i \lim_{\delta\to 0} (\delta I_k + Z_i' Z_i)^{-1} Z_i') X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' (I_{n_i} - Z_i \lim_{\delta\to 0} (\delta I_k + Z_i' Z_i)^{-1} Z_i') y_i\right)$$

$$\hat{\beta}_{\infty} = \left(\sum_{i=1}^{N} X_i' (I_{n_i} - Z_i Z_i^{+}) X_i\right)^{+} \left(\sum_{i=1}^{N} X_i' (I_{n_i} - Z_i Z_i^{+}) y_i\right), \quad \text{(2.29)}$$

Note que esta fórmula trabaja aun cuando la matriz inversa de $\sum_{i=1}^{N} X_i' (I_{n_i} - Z_i Z_i^{+}) X_i$ es cero, por ejemplo, sucede cuando $X_i = Z_i A_i$ en el modelo de curva creciente, capítulo 4. Ya que la inversa generalizada de la matriz nula es la matriz nula $\hat{\beta}_{\infty} = 0$ para este caso.

El estimador (2.29) tiene una bonita interpretación como el estimador de efectos fijos. En efecto, consideremos el modelo de efectos fijos como una alternativa para el modelo LME, a saber,

$$y_i = X_i \beta + Z_i b_i + \varepsilon_i, \quad b_i \; es \; fijo$$

Y tan pronto $\{\varepsilon_i\}$ son vectores aleatorios independientes con media cero y varianza σ^2. En el enfoque de efectos fijos, $\{b_i$, i=1,…,N$\}$ son parámetros ruidosos desconocidos. Ya que los $\{b_i\}$ son fijos y $cov(y_i) = \sigma^2$I, el estimador OLS es el mejor estimador lineal insesgado (BLUE), el cual minimiza la suma total de cuadrados,

$$\lim_{\beta, b_{1,\dots} b_N} \sum_{i=1}^{N} \|y_i - X_i\beta - Z_i b_i\|^2, \quad \text{(2.31)}$$

Podemos rescribir este problema de optimización como un problema de mínimos cuadrados estándar introduciendo una matriz compuesta $N_T x(m + Nk)$,

$$W = [X, Z] = \begin{bmatrix} X_1 & Z_1 & 0 & \dots & 0 \\ X_2 & 0 & Z_2 & & \\ \dots & & & \ddots & \\ X_N & 0 & 0 & \dots & Z_N \end{bmatrix}, \quad (2.32)$$

Con $\nu = (\beta', b')'$. Luego (2.31) es equivalente a $\|y - W\nu\|^2$, donde y y X están definidas en (2.8). La solución OLS para la última suma de cuadrados tiene el mínimo

$S_{min} = y'(I - WW^+)y.$ (2.33)

Alternativamente, podemos obtener una fórmula económica para (2.33) manteniendo β constante y minimizando la suma de cuadrados sobre $\{b_i\}$. En efecto, si β es fijo, la forma cuadrática en (2.31) puede ser minimizada para b_i separadamente. Como sigue de la teoría de mínimos cuadrados con la matriz de diseño no necesariamente de rango completo (Rao 1973, Graybill 1983), los estimados son $\widehat{b_i} = Z_i^+(y - X\beta)$, con el mínimo

$$\min_{b_i} \|y_i - X_i\beta - Z_i b_i\|^2 = (y_i - X_i\beta)' (I_{n_i} - Z_i Z_i^+)(y_i - X_i\beta), \quad (2.34)$$

Después, minimizando $(y_i - X_i\beta)' (I_{n_i} - Z_i Z_i^+)(y_i - X_i\beta)$ para β finalmente conduce a (2.29). Podemos interpretar este resultado diciendo que el modelo de efectos fijos corresponde al modelo de efectos aleatorios con matriz de covarianzas D cuando tiende a infinito.

Cuál de los modelos, (2.4) o (2.30) es mejor: ¿efectos aleatorios o fijos? Uno no puede responder esta pregunta porque esta elección es en efecto, un supuesto; algunas discusiones sobre este tópico puede ser encontrado en la literatura de varianzas de componentes, por ejemplo, Searle (1971b), Lindman (1992). Claramente, un modelo de efectos fijos es menos restrictivo y es fácil manejar debido a que se reduce a un modelo lineal estándar, pero el precio es que el número de parámetros ruidosos se incrementa con el número de sujetos (cluster). En particular, el enfoque de efecto fijo será preferible si el número de cluster (N) es pequeño y el número de observaciones por cluster es grande. Un model general de curva de crecimiento, será considerado en el capítulo 4, es un compromiso razonable entre modelos de efectos fijos y aleatorios. Como sigue de (2.29), el estimador OLS de efectos fijos, $\hat{\beta}_\infty$, es inconsistente para β cuando n tiende a infinito a pesar de un número creciente de parámetros ruidosos, si la matriz inversa es no singular. Un caso especial de un modelo lineal con un cluster especifico con término intercepto es considerado en la sección 2.4. En la subsección 3.2.1 explicaremos porque la combinación de efectos aleatorios y fijos al mismo tiempo conduce a un modelo inválido.

Como aprenderemos más tarde, la cantidad S_{min} juega un rol importante en el modelo LME; particularmente, la desigualdad

$$S_{min} = \sum_{i=1}^{N} \left(y_i - X_i\hat{\beta}_\infty\right)' (I_{n_i} - Z_i Z_i^+)\left(y_i - X_i\hat{\beta}_\infty\right) > 0$$

Provee una condición necesaria y suficiente para la existencia de la máxima verosimilitud estimada, sección 2.5. También, S_{min} es la característica clave de la prueba estadística para la presencia del efecto aleatorio (sección 3.5) y MINQUE para el parámetro σ^2 (subsección 3.10.2)

Probaremos que es el límite inferior para la suma ponderada de cuadrados, o más precisamente,

$$\sum_{i=1}^{N}(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta) \geq S_{min}, \quad (2.36)$$

Para cualquier β y D.

Demostración. Primero mostraremos que $(I + Z_i D Z_i')^{-1} \geq (I - Z_i Z_i^+)$. Multiplicar ambos lados de la desigualdad por $I + Z_i D Z_i'$ siguiendo una sencilla Algebra matricial

$$(I + Z_i D Z_i')^{-1}(I + Z_i D Z_i') \geq (I - Z_i Z_i^+)(I + Z_i D Z_i')$$

$$I \geq I + Z_i D Z_i' - Z_i Z_i^+ + Z_i Z_i^+ Z_i D Z_i', \text{ pero } Z_i Z_i^+ Z_i = Z_i$$

$$I \geq I + Z_i D Z_i' - Z_i Z_i^+ + Z_i D Z_i'$$

lo que resulta

$$I \geq I - Z_i Z_i^+$$

De aquí, el lado izquierdo de (2.36) es mayor que o igual a S_{min}

$$\sum_{i=1}^{N}(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta) \geq \min_{\beta} \sum_{i=1}^{N}(y_i - X_i\beta)'\,(I_{n_i} - Z_i Z_i^+)(y_i - X_i\beta)$$

Lo cual completa la demostración.

II.II FUNCIÓN DE LOG-VEROSIMILITUD PERFILADA

Tomando la derivada de (2.14) con respecto a σ^2, es fácil ver que la función l se maximiza en

$$\hat{\sigma}^2 = \frac{1}{N_T}\sum_{i=1}^{N}(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta), \quad (2.37)$$

Lindstrom and Bates (1988) and Wolfinger et al. (1994), entre otros, tomaron ventaja de la fórmula (2.37) y sugirieron la función log-verosimilitud varianza perfilada sustituyendo (2.37) en (2.14). Esto conduce a un problema de maximización equivalente con σ^2 eliminada del modelo.

$$l_p(\beta, D) = -\frac{1}{2}\left\{N_T \ln \sum_{i=1}^{N}\left[(y_i - X_i\beta)' {V_i}^{-1}(y_i - X_i\beta)\right] + \sum_{i=1}^{N}\ln|V_i|\right\} \quad (2.38)$$

Donde la constante $c = \frac{1}{2}N_T(N_T - 1)$ es ignorada (el índice p indica que la función es perfilada). Aplicando las fórmulas (2.19) y (2.21) y suponiendo que D es no singular, obtenemos una parametrización de reducción de dimensión económica de la función perfilada,

$$l_p(\beta, D) = -\frac{1}{2}\left\{N_T \ln \sum_{i=1}^{N}[S_i - {r_i}'(D^{-1} + M_i)^{-1} r_i] + \sum_{i=1}^{N}\ln|I + D M_i|\right\} \quad (2.39)$$

Donde S_i está definida en (2.25), $r_i = Z_i' e_i$, $M_i = Z_i' Z_{i.}$ (2.40)

Además, en la parametrización (2.39) podemos considerar $D\_ = D^{-1}$ como el argumento; algunas veces esta matriz es llamada matriz de precisión. Esto conduce a la parametrización de la matriz de precisión

$$l_p(\beta, D\_) = -\frac{1}{2}\left\{N_T \ln \sum_{i=1}^{N} [S_i - r_i'(D\_ + M_i)^{-1} r_i] + \sum_{i=1}^{N} \ln|D\_ + M_i| - N \ln|D\_|\right\} \quad (2.41)$$

Una posibilidad adicional es excluir β usando la fórmula GLS (2.26). Necesitamos el siguiente resultado general para obtener la última función (σ^2, β)-perfil log-verosimilitud.

Usando la notación "larga" (2.8), la suma total de cuadrados puede ser re-escrita como

$$\sum_{i=1}^{N}\left[(y_i - X_i\beta)' {V_i}^{-1} (y_i - X_i\beta)\right] = (y - X\beta)' V^{-1} (y - X\beta)$$

Luego aplicando la proposición 1, obtenemos

$$q = \min_{\beta \in R^m} \sum (y_i - X_i\beta)' {V_i}^{-1} (y_i - X_i\beta) \quad (2.42)$$

$$= \left(\sum y_i' {V_i}^{-1} y_i\right) - \left(\sum X_i' {V_i}^{-1} y_i\right)' \left(\sum X_i' {V_i}^{-1} X_i\right)^{-1} \left(\sum X_i' {V_i}^{-1} y_i\right)$$

Consecuentemente, si V_i es conocida, el estimador GLS (2.26) conduce a la siguiente fórmula para (2.37)

$$\hat{\sigma}^2_{GLS} = \min_{\beta \in R^m} \frac{1}{N_T} \sum_{i=1}^{N} (y_i - X_i\beta)' {V_i}^{-1} (y_i - X_i\beta)$$

$$\hat{\sigma}^2_{GLS} = \frac{q}{N_T}$$

La función log-verosimilitud perfilada completamente, como una función de D_, está dado por

$$l_p(D\_) = -\frac{1}{2}\left\{N_T \ln\big(q(D\_)\big) + \sum_{i=1}^{N} \ln|D\_ + M_i| - N \ln|D\_|\right\}, \quad (2.43)$$

Esta forma de función log-verosimilitud es tal vez la más económica debido a las cantidades tales como $\sum y_i' y_i$, $Z_i' y_i$, $\sum X_i' y_i$, y $\sum X_i' X_i$ puede ser calculado previamente. La parametrización de la verosimilitud perfilada está bien adecuada para la construcción del intervalo de confianza de la verosimilitud perfilada., ver sección 3.4 [1].

III. MÁXIMA VEROSIMILITUD RESTRINGIDA.

Es conocido que la estimación de la varianza de la máxima verosimilitud es sesgada para muestras finitas. Por ejemplo, en un modelo de regresión lineal estándar $y \sim \mathcal{N}(X\beta, \sigma^2 I)$, el MLE de la varianza $\hat{\sigma}^2_{ML} = \frac{SS}{n}$, subestima σ^2, donde SS es la suma residual de cuadrados, n es el número de observaciones, y m es el número de coeficientes de regresión. El estimador insesgado de σ^2 es $\frac{SS}{n-m}$, lo cual toma los grados de libertad en cuenta. Para reducir el sesgo en los componentes de varianza del modelo, Patterson y Thompson (1971), y después Harville (1974), sugirieron la modificación de la función de verosimilitud estándar usando mínimos cuadrados residuales generalizados. El método resultante se conoce como restringido (o tal vez más precisamente, Residual) Estimación de máxima verosimilitud (RMLE). Laird and Ware (1982) aplicaron este método al modelo LME (2.13).

Primero obtenemos el RML para el modelo lineal general y luego aplicarlo al modelo LME.

Sea el modelo lineal general definido como $y \sim \mathcal{N}(X\beta, V)$ donde X es la matriz nxm de rango completo y V es la matriz de covarianzas nxn, dependiente de algunos parámetros θ. En la estimación RML maximizamos la función log-verosimilitud para el vector residual $\hat{e} = y - X\hat{\beta}$, donde $\hat{\beta}$ es el estimador GLS, $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$. Note que y tiene distribución normal también. Además, $\hat{\beta}$ y $\hat{e}$ son independientes porque

$$cov(X'V^{-1}y, \hat{e}) = cov\big(X'V^{-1}y, y - X\hat{\beta}\big) = cov(X'V^{-1}y, y - X(X'V^{-1}X)^{-1}X'V^{-1}y)$$

Esto implica que la función de verosimilitud para y es el producto de la función de verosimilitud para $\hat{e}$ y para $\hat{\beta}$. Pero $\hat{\beta} \sim \mathcal{N}(\beta, (X'V^{-1}X)^{-1})$, y por lo tanto la función log-verosimilitud para el vector residual $\hat{e}$, sin considerar la constante, es

$$L(y, \theta) = L(\hat{\beta}, \hat{e}, \theta)$$

$$L(y, \theta) = L(\hat{\beta}, \theta)L(\hat{e}, \theta)$$

$$l(y, \theta) = l\big(\hat{\beta}, \theta\big) + l(\hat{e}, \theta)$$

$$l(y, \theta) - l\big(\hat{\beta}, \theta\big) = l(\hat{e}, \theta)$$

$$l(\hat{e}, \theta)$$

$$= -\frac{1}{2}\Big\{\ln|X'V^{-1}X| + \ln|V| + (y - X\beta)'V^{-1}(y - X\beta) - \big(\hat{\beta} - \beta\big)'X'V^{-1}X\big(\hat{\beta} - \beta\big)\Big\} \quad (2.45)$$

Pero ya que

$$(y - X\beta)'V^{-1}(y - X\beta) = \big(y - X\hat{\beta}\big)'V^{-1}\big(y - X\hat{\beta}\big) + \big(\hat{\beta} - \beta\big)'X'V^{-1}X\big(\hat{\beta} - \beta\big)$$

Podemos re-escribir la función log-verosimilitud para $\hat{e}$ como

$$l(\hat{e}, \theta) = -\frac{1}{2}\{\ln|X'V^{-1}X| + \ln|V| + \hat{e}'V^{-1}\hat{e}\}$$

Claramente, la maximización de esta función es equivalente a

$$l_R(\hat{e}, \theta) = -\frac{1}{2}\{\ln|X'V^{-1}X| + \ln|V| + (y - X\beta)'V^{-1}(y - X\beta)\} \quad (2.46)$$

Porque la maximización de l_R para β da $\hat{e} = y - X\hat{\beta}$. La función l_R es llamada función log-verosimilitud residual. Note que l_R difiere de la función log-verosimilitud estándar por el término $-\frac{1}{2}\ln|X'V^{-1}X|$. Como una palabra de precaución, l_R no es una función log-verosimilitud real y consecuentemente la matriz de covarianza para θ no puede ser obtenida como la inversa de la segunda derivada esperada. Aunque asintóticamente, ML y RML y las respectivas matrices de covarianza coinciden, ver subsección 3.6.3 [1].

Ya que las versiones regular y restringida difieren por el término $-0{,}5ln|X'V^{-1}X|$ que se traduce a $\ln\left|\sum X_i'V_i^{-1}X_i\right|$ para el modelo LME, llegamos a la función RML

$$L_R(\theta) = -\frac{1}{2}\Bigg\{(N_T - m)\ln\sigma^2 + \ln\left|\sum X_i'V_i^{-1}X_i\right| + \sum_{i=1}^{N}\left[\ln|V_i| + \sigma^{-2}(y_i - X_i\beta)'V_i^{-1}(y_i - X_i\beta)\right]\Bigg\}. \quad (2.47)$$

Note que esta es la función log-verosimilitud estándar (2.14) aumentado por el término

$$-\frac{1}{2}\left\{-m\ln\sigma^2 + \ln\left|\sum X_i'V_i^{-1}X_i\right|\right\}. \quad (2.48)$$

La función log-verosimilitud para RML puede ser reparametrizada siguiendo la línea de la reparametrización previa. Por ejemplo, llegamos a la función σ^2–perfilada RML sustituyendo

$$\hat{\sigma}_R^2 = \frac{1}{N_T - m}\sum_{i=1}^{N}(y_i - X_i\beta)'(I_{n_i} + Z_iDZ_i')^{-1}(y_i - X_i\beta), \quad (2.49)$$

dentro de (2.47), lo cual conduce a la función log-verosimilitud de varianza perfilada (hasta un término constante)

$$L_{R_P}(\beta, D) =$$

$$-\frac{1}{2}\left\{(N_T - m)\ln\left[\sum_{i=1}^{N}(y_i - X_i\beta)'V_i^{-1}(y_i - X_i\beta)\right] + \ln\left|\sum X_i'V_i^{-1}X_i\right| + \sum_{i=1}^{N}\ln|V_i|\right\}, \quad (2.50)$$

Note que los grados de libertad en la versión restringida del estimador σ^2 (2.49) son ajustados por el número de efectos fijos, m. Sin embargo, si el número de observaciones es mucho mayor que m, el ajuste será despreciable.

Podemos llegar además a la parametrización de reducción de dimensión usando las fórmulas (2.19)-(2.23) [1]. En efecto, podemos usar todas las parametrizaciones log-verosimilitud previamente obtenido aumentado por el término (2.48). Claramente, uno puede usar las fórmulas (2.19) para reducir la dimensión de la matriz inversa en (2.48).

IV. MODELO LINEAL DE EFECTOS MIXTOS BALANCEADO (5)

En esta sección consideramos un caso muy especial del modelo LME (2.4), el modelo de coeficientes aleatorios balanceados. En este modelo todos los cluster tienen el mismo tamaño ($n_i = n$) y $Z = X_i = Z_i,\ i = 1, \dots, N,\ \ (2.51)$

Donde la matriz Z es de rango completo. En la presentación de (1.12), puede ser llamado el modelo de coeficientes aleatorios balanceados. Este modelo es atractivo porque nos permite obtener estimaciones de máxima verosimilitud en forma cerrada. Una situación más general cuando MLEs admite formas cerradas será considerada en la subsección (4.1.5).

Primero, notemos que para el modelo (2.4) con datos (2.51), el estimado de los efectos fijos (2.29) es cero porque la matriz a invertir $Z'(I - Z(Z'Z)^{-1}Z')Z$, es nula y por lo tanto la inversa generalizada también es nula. Segundo, el modelo con datos balanceados puede ser representado en formato rectangular como $Y = Z\beta 1' + E$, donde Y es uma matriz nxN y 1 es un vector de unos Nx1; E es la matriz del término del error nxN con media cero, filas mutuamente independientes, con la matriz de covarianzas $\sigma^2 V = \sigma^2(I + ZDZ')$. Este modelo es un caso especial del modelo clásico curva de crecimiento $Y = Z\beta X + E$, donde los vectores columnas de E son iid con matriz de covarianza no estructurada, estudiada por muchos autores: Potthoff and Roy (1964), Rao (1965), Khatri (1966) y Grizzle and Allen (1969). Pan y Fang (2002), donde fue denominado simple el modelo de curva de crecimiento. En particular, los últimos autores obtuvieron formas cerradas para los parámetros de la varianza, Como lo hacemos abajo en el teorema 2.

(5) Modelo lineal de efectos mixtos: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

Iniciamos probando que el modelo (2.4) con los datos (2.51), el estimador GLS (2.26) colapsa al estimador OLS y por tanto aquello no depende de la matriz D. Más precisamente,

$$\hat{\beta}_{GLS} = \hat{\beta}_{OLS} = (Z'Z)^{-1}Z'\bar{y}, \qquad (2.52)$$

Donde

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}.$$

Y por lo tanto la minimización de la suma ponderada es equivalente a la minimización de $(\bar{y} - Z\beta)'V^{-1}(\bar{y} - Z\beta)$, porque el primer término es constante, Además, denotando $\hat{\beta} = (Z'Z)^{-1}Z'\bar{y}$ y aplicando la fórmla de reducción de dimensión (2.19) para el último término de (2.53), obtenemos

$$(\bar{y} - Z\beta)'V^{-1}(\bar{y} - Z\beta) = \big(\hat{e} - Z(\beta - \hat{\beta})\big)'[I - Z(D^{-1} + Z'Z)^{-1}Z']\big(\hat{e} - Z(\beta - \hat{\beta})\big), (2.54)$$

Donde $\hat{e} = \bar{y} - Z\hat{\beta}$ es un vector residual. Pero como en el modelo lineal estándar, los regresores y los residuos son ortogonales, $Z'\hat{e} = 0$, porque

$$Z'\hat{e} = Z'[I - Z(Z'Z)^{-1}Z']\bar{y} = 0.$$

De aquí, la suma (2.54) se simplifica a

$$\hat{e}'\hat{e} + \left(Z(\beta - \hat{\beta})\right)' V^{-1}\left(Z(\beta - \hat{\beta})\right)$$

Finalmente, ya que la matriz $Z'V^{-1}Z$ es definida positiva, el mínimo de (2.54) ocurre en $\beta = \hat{\beta}$; así, (2.52) está probada.

Así, de (2.52) vemos que para un modelo de regresión con coeficientes aleatorios balanceado, OLS=GLS=MLE.

V. MODELO DE CURVA DE CRECIMIENTO LINEAL (LGC).(6)

MODELO DE CURVA DE CRECIMIENTO LINEAL (LGC)

(4.1) $y_i = Z_i a_i + \varepsilon_i \quad , \quad \in (\varepsilon_i) = 0 \quad , \quad cov(\varepsilon_i) = \sigma^2 I$

(4.1) $a_i = A_i \beta + b_i \quad , \quad \in (b_i) = 0 \quad , \quad cov(b_i) = \sigma^2 D$

$\sum_{i=1}^{N} A_i' A_i$ es no singular

Z_i es de rango completo, $rank(Z_i) = k \leq n_i$

$y_i \quad n_i x1 \quad respuesta, \ \varepsilon_i \ n_i x1$

$Z_i \quad n_i xk \quad matriz\ de\ diseño$

$a_i \quad kx1 \quad vector\ de\ coeficiente\ individuales$

$a_i \ y \ \varepsilon_i \quad son\ mutuamente\ independientes$

$A_i \ kxm \quad matriz\ de\ diseño$

$\beta \ mx1 \quad vector\ de\ parámetros$

$\sigma^2 \quad escalar \ varianza\ dentro\ del\ sujeto$

$D \ kxk \quad parámetros\ de\ varianza\ escalada$

(6) Modelo de curva de crecimiento lineal: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

Estimación del modelo de la curva de crecimiento lineal

4.1.1 Matriz D conocida, de 2.28

$$\hat{\beta}_{GLS=}(\sum_{i=1}^{N} A_i' Z_i' V_i^{-1} Z_i A_i)^{-1}(\sum_{i=1}^{N} A_i' Z_i' V_i^{-1} y_i)$$

Con $cov\left(\hat{\beta}_{GLS}\right) = \sigma^2\left(\sum_{i=1}^{N} A_i' Z_i' V_i^{-1} Z_i A_i\right)^{-1}$

$$V_i = I_{n_i} + Z_i D Z_i^{'}$$

$$W_i = (Z_i' Z_i)^{-1} + D\ ,\ i = 1, \dots, N$$

Aplicando la fórmula (2.25) de la reducción de dimensión, obtenemos:

$$Z_i' V_i^{-1} y_i = Z_i' \left[I - Z_i (D^{-1} + Z_i^{'} Z_i)^{-1} Z_i' \right] y_i$$

$$= \left(Z_i' - Z_i^{'} Z_i (D^{-1} + Z_i^{'} Z_i)^{-1} Z_i' \right) y_i$$

$$= \left(I - Z_i^{'} Z_i (D^{-1} + Z_i^{'} Z_i)^{-1} \right) Z_i' y_i$$

$$= \left(\left(D^{-1} + Z_i^{'} Z_i \right) \left(D^{-1} + Z_i^{'} Z_i \right)^{-1} - Z_i^{'} Z_i (D^{-1} + Z_i^{'} Z_i)^{-1} \right) Z_i' y_i$$

$$= \left(\left(D^{-1} + Z_i^{'} Z_i \right) - Z_i^{'} Z_i \right) \left((D^{-1} + Z_i^{'} Z_i)^{-1} Z_i' y_i \right)$$

$$= D^{-1} \left(D^{-1} + Z_i^{'} Z_i \right)^{-1} Z_i' y_i$$

$$= \left((D^{-1} + Z_i' Z_i) D \right)^{-1} Z_i^{'} y_i$$

$$= \left(I + Z_i^{'} Z_i D \right)^{-1} Z_i' y_i$$

$$= \left(I + Z_i^{'} Z_i D \right)^{-1} \left(Z_i^{'} Z_i \right) \left(Z_i^{'} Z_i \right)^{-1} (Z_i' y_i)$$

$$= [(Z_i' Z_i)^{-1} (I + Z_i' Z_i D)]^{-1} \left[(Z_i' Z_i)^{-1} \left(Z_i^{'} y_i \right) \right]$$

$$= [(Z_i' Z_i)^{-1} + D]^{-1} \left[(Z_i' Z_i)^{-1} Z_i^{'} y_i \right]$$

$$= W_i^{-1} \cdot a_i^o$$

Donde $a_i^o = (Z_i' Z_i)^{-1} Z_i^{'} y_i$ (4.11)

El estimador se convierte en una versión económica

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{N} \left(A_i' W_i^{-1} A_i \right) \right]^{-1} \left(\sum_{i=1}^{N} A_i' W_i^{-1} a_i^o \right) \qquad (4.10)$$

$$cov\left(\hat{\beta}_{GLS} \right) = \sigma^2 \left(\sum A_i^{'} W_i^{-1} A_i \right)^{-1}$$

Modelo de curva de crecimiento lineal (LGC) en forma matricial

$$y_i = Z_i a_i + \varepsilon_i$$

$$a_i = A_i \beta + b_i$$

$$y_i = Z_i A_i \beta + Z_i b_i + \varepsilon_i$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} Z_1 A_1 \\ \vdots \\ Z_N A_N \end{bmatrix} \beta + \begin{bmatrix} Z_1 b_1 + \varepsilon_1 \\ \vdots \\ Z_N b_N + \varepsilon_N \end{bmatrix}$$

$$Y = X\beta + \eta \qquad \eta = \begin{bmatrix} Z_1 b_1 + \varepsilon_1 \\ \vdots \\ Z_N b_N + \varepsilon_N \end{bmatrix}$$

$$\sum = var(\eta) = \sigma^2 \begin{bmatrix} Z_1 D Z_1' + I & & 0 \\ & \ddots & \\ 0 & & Z_N D Z_N' + I \end{bmatrix} \qquad v(\eta_1) = \sigma^2 \, Z_1 D Z_1' + \sigma^2 \, I$$

$$\sum = var(\eta) = \sigma^2 V \qquad \sum = v(\ \eta)$$

$$\hat{\beta} = \left(X' \Sigma^{-1} X\right)^{-1} X' \Sigma^{-1} Y$$

$$\hat{\beta} = \left([(Z_1 A_1)' \quad \cdots \quad (Z_N A_N)'](\sigma^2 V)^{-1} \begin{bmatrix} Z_1 A_1 \\ \vdots \\ Z_N A_N \end{bmatrix} \right)^{-1} [(Z_1 A_1)' \quad \cdots \quad (Z_N A_N)'](\sigma^2 V)^{-1} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

$$\left([(Z_1 A_1)' \quad \cdots \quad (Z_N A_N)'] \frac{1}{\sigma^2} \begin{bmatrix} Z_1 D Z_1' + I & & 0 \\ & \ddots & \\ 0 & & Z_N D Z_N' + I \end{bmatrix}^{-1} \begin{bmatrix} Z_1 A_1 \\ \vdots \\ Z_N A_N \end{bmatrix} \right)^{-1}$$

$$= \sigma^2 [(Z_1 A_1)' \quad \cdots \quad (Z_N A_N)'] \begin{bmatrix} (Z_1 D Z_1' + I)^{-1} Z_1 A_1 \\ \vdots \\ (Z_N D Z_N' + I)^{-1} Z_N A_N \end{bmatrix} = \sigma^2 \sum (Z_i A_i)' (Z_i D Z_i' + I)^{-1} (Z_i A_i)$$

$$\hat{\beta} = \frac{\sigma^2}{\sigma^2} \left(\sum (Z_i A_i)' (Z_i D Z_i' + I)^{-1} (Z_i A_i) \right)^{-1} \sum (Z_i A_i)' (Z_i D Z_i' + I)^{-1} y_i$$

$$\hat{\beta} = \left(\sum_{i=1}^{N} A_i' Z_i' (I + Z_i D Z_i')^{-1} Z_i A_i \right)^{-1} \sum_{i=1}^{N} A_i' Z_i' (I + Z_i D Z_i')^{-1} y_i$$

ALGUNOS CASOS DE MODELO DE CURVA DE CRECIMIENTO LINEAL (7)

Llamamos un modelo LME un modelo de curva de crecimiento balanceado si $X_i = X$ y $A_i = I \otimes q_i'$, donde $q_i$es un vector columna px1 y $\otimes$ es el producto de Kroneker.

$$A_i = I \otimes q_i{}' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{kxk} \otimes \begin{bmatrix} q_1 \\ \vdots \\ q_p \end{bmatrix}'_{1xp}$$

$$= \begin{bmatrix} [q_1 & \cdots & q_p] & [0 & \ldots & 0] & [0 & \ldots & 0] \\ [0 & \ldots & 0] & [q_1 & \cdots & q_p] & [0 & \ldots & 0] \\ [0 & \ldots & 0] & [0 & \ldots & 0] & [q_1 & \cdots & q_p] \end{bmatrix}_{kxm}$$

Donde m=pk.

De manera específica

$$y_i = X_i a_i + \varepsilon_i, \ \text{con} \ a_i = A_i \beta^* + b_i, \ \ \text{i} = 1, \ldots, \text{N}$$

Remplazando en el modelo, queda

$$y_i = X_i(A_i \beta^* + b_i) + \varepsilon_i, \qquad \text{i} = 1, \ldots, \text{N}$$

$$y_i = X_i((I \otimes q_i{}')\beta^* + b_i) + \varepsilon_i$$

con $\beta = A_i \beta^*$ remplazamos y queda,

$$y_i = X_i(\beta + b_i) + \varepsilon_i$$

Ejemplo para ilustrar el modelo de crecimiento

En un estudio longitudinal se analiza el crecimiento de un grupo de N niños en función del tiempo y de la influencia que tiene el consumo de calcio. $y_i \ es \ un \ vector \ n_i x 1$,

$y_{ij} \ es \ la \ estatura \ del \ niño \ i \ en \ el \ tiempo \ j$,

$t_{ij} \ es \ el \ tiempo \ j \ en \ que \ el \ niño \ i \ consume \ la \ dosis \ de \ calcio \ y \ se \ mide \ la \ estatura,$

$\varepsilon_{ij} \ es \ una \ variable \ aleatoria \ N(0, \sigma^2)$.

$$y_{ij} = a_{i1} + a_{i2} t_{ij} + \varepsilon_{ij}$$

La misma ecuación en forma vectorial es:

$$y_i = \begin{bmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in} \end{bmatrix} \begin{bmatrix} a_{i1} \\ a_{i2} \end{bmatrix} + \varepsilon_i$$

Donde los coeficientes aleatorios a_{i1}, a_{i2} estan en función del consumo de calcio

$$a_{i1} = \beta_1 + \beta_2 c_i + b_{i1}$$

$$a_{i2} = \beta_3 + \beta_4 c_i + b_{i2}$$

Donde
$b_{i1}, b_{i2} \ son \ variables \ aleatorias \ normales \ con \ media \ 0 \ y \ varianza \ {\sigma^2}_{i1}, {\sigma^2}_{i2} \ no \ correlacianados$

En forma matricial es

$$\begin{bmatrix} a_{i1} \\ a_{i2} \end{bmatrix} = \begin{bmatrix} [1 & c_i] & [0 & 0] \\ [0 & 0] & [1 & c_i] \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix}$$

Esto es

$$a_i = (I \otimes q_i{}')\beta + b_i$$

Entonces

$$y_i = X_i a_i + \varepsilon_i,$$

Donde $X_i = \begin{bmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in} \end{bmatrix}$

(7) Casos de curva de crecimiento lineal: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

V.I.ESTIMADOR GLS (GLS *MÍNIMOS CUADRADOS GENERALIZADOS*)

Implementación del modelo LGC

El estimador GLS (mínimos cuadrados generalizados) tiene la siguiente implementación de dos pasos:

1.- Separadamente estimar (4.1) $y_i = Z_i a_i + \varepsilon_i$ por el OLS para obtener (4.11) para cada $i' = 1, \dots, N$ $a_i^o = (Z_i' Z_i)^{-1} Z_i^{'} y_i$

2.- Reemplace a_i con a_i^o en (4.2) $a_i = A_i \beta + b_i$ y aplique mínimo cuadrado ponderado (4.9)

$a_i = A_i \beta + b_i$

$a_i^o = (Z_i' Z_i)^{-1} Z_i^{'} y_i \qquad y_i = Z_i A_i \beta + Z_i b_i + \varepsilon_i$

$a_i^o = (Z_i' Z_i)^{-1} Z_i^{'} (Z_i A_i \beta + Z_i b_i + \varepsilon_i)$

$a_i^o = (Z_i' Z_i)^{-1} Z_i^{'} Z_i A_i \beta + (Z_i' Z_i)^{-1} Z_i^{'} Z_i b_i + (Z_i' Z_i)^{-1} Z_i^{'} \varepsilon_i$

$a_i^o = A_i \beta + b_i + (Z_i' Z_i)^{-1} Z_i^{'} \varepsilon_i$

Se observa que:

$V(a_i^o) = \sigma^2 D + V((Z_i' Z_i)^{-1} Z_i^{'} \varepsilon_i)$

$$= \sigma^2 D + (Z_i' Z_i)^{-1} Z_i^{'} \sigma^2$$

$$= \sigma^2 D + \sigma^2 (Z_i' Z_i)^{-1} Z_i Z_i^{'} (Z_i' Z_i)^{-1}$$

$$= \sigma^2 (D + (Z_i' Z_i)^{-1})$$

De aquí

$$\in (a_i^o) = A_i \beta \quad , \quad V(a_i^o) = cov(a_i^o)$$

$$\sigma^2 W_i = \sigma^2 (D + (Z_i' Z_i)^{-1})$$

Y por lo tanto 4.10 como una implementación de los mínimos cuadrados ponderados con peso W_i^{-1}

Dos casos para D

D = 0, entonces

$$\hat{\beta}_{OLS} = \left(\sum A_i' [(Z_i' Z_i)^{-1} + D]^{-1} \right)^{-1} \left(\sum A_i' W_i^{-1} a_i^o \right)$$

$$\hat{\beta}_{OLS} = \left(\sum A_i' \, Z_i' Z_i \right)^{-1} \left(\sum A_i' Z_i' Z_i a_i^o \right), donde \; a_i^o = (Z_i' Z_i)^{-1} Z_i' y_i$$

$$\hat{\beta}_{OLS} = \left(\sum_{i=1}^{N} A_i' \, Z_i' Z_i A_i \right) \left(\sum A_i' \, Z_i' y_i \right) \qquad (4.12)$$

El otro caso $D \to \infty \quad , \quad D = dI \quad y \quad d \to \infty$

$$\left(\sum A_i' ((Z_i' Z_i)^{-1} + D)^{-1} \, A_i \right)^{-1} \left(\sum A_i' ((Z_i' Z_i)^{-1} + D)^{-1} a_i^o \right)$$

$$= \lim_{d \to \infty} \left(\sum A_i' ((Z_i' Z_i)^{-1} + dI)^{-1} A_i \right)^{-1} \left(\sum A_i' ((Z_i' Z_i)^{-1} + dI)^{-1} a_i^o \right)$$

$$= \lim_{d \to 0} \left(\sum A_i' \left((Z_i' Z_i)^{-1} + \frac{1}{d} I \right)^{-1} A_i \right)^{-1} \left(\sum A_i' \left((Z_i' Z_i)^{-1} + \frac{1}{d} I \right)^{-1} a_i^o \right)$$

$$= \lim_{d \to 0} \left(\sum A_i' (d(Z_i' Z_i)^{-1} + I)^{-1} A_i \right)^{-1} \left(\sum A_i' (d(Z_i' Z_i)^{-1} + I)^{-1} a_i^o \right)$$

$$= \left(\sum A_i' A_i \right)^{-1} \left(\sum A_i' \, a_i^o \right)$$

$$\hat{\beta}_{OLS} = \left(\sum_{i=1}^{N} A_i' A_i\right)^{-1} \left(\sum_{i=1}^{N} A_i' a_i^o\right)$$

Estos estimados pueden ser interpretados como mínimos cuadrados ordinarios aplicados a la segunda etapa del modelo ($W_i = I,\ \ no\ ponderado$)

V.VI. MODELO DE CURVA DE CRECIMIENTO ESPECIAL (RECTANGULAR) (RLGC)(8)

*Def*inición.- *El RLGC con igual matriz de diseño es llamado el modelo de curva de crecimiento balanceado BLGC, en adición a (4.38)* $\mathrm{A_i} = (\mathrm{I} \otimes \mathrm{q_i})$ $\mathrm{q_i}$ es $\mathrm{px1}$ *tenemos* $n_\mathrm{i} = n,\ Z = Z_i$ con $k.p < n$ cuando $p = 1$ y $q_i = 1$, $BLGC = BRC$ o más precisamente el modelo de la curva de crecimiento colapsa al modelo del coeficiente aleatorio balanceado como en la sección (2.3), $GLS = OLS$ para el modelo BRC.

Teorema 25. Para el BLGC model, $\hat{\beta}_{GLS}$ no depende de la matriz D y coincide con el estimador OLS $\hat{\beta}_j = (\sum q_i\, q_i')^{-1}\left(\sum q_i a_{ij}^o\right),\ j = 1, \cdots, k$ y $cov\left(\hat{\beta}_j\right) = \sigma^2\left((Z'Z)_{ij}^{-1} + D_{jj}\right)(\sum q_i\, q_i')^{-1}$

(8) Modelo lineal mixto: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

V.VII. ESTUDIO LONGITUDINAL POR MEDIO DE MODELO DE CURVA DE CRECIMIENTO (9)

ESTUDIO LONGITUDINAL DE UNA PRODUCCIÓN AVÍCOLA-(EN R).

(9) Crecimiento de cada pollito a través del tiempo: *Crowder, M. and Hand, D. (1990), Analysis of Repeated Measures, Chapman and Hall (example 5.3)...*

Descripción:

Para ilustrar el modelo de efectos mixtos consideramos un estudio longitudinal que nos provee el programa R en la cual presenta los datos de peso, tiempo pollito y dieta suministrada.

La base de datos ChickWeight tiene 578 filas y 4 columnas de un experimento para observar el efecto de dieta en el crecimiento de los pollitos.

Weight: Es un vector numérico que nos indica el peso del pollito a través del tiempo.

Time: Número de día que la dieta es suministrada al pollito y que la medida del peso fue tomada.

Chick: Es un factor ordenado de 1 a 50 que identifica el número de pollitos.

Diet: Es un factor de niveles de 1-4 que indica la dieta que recibe el pollito en cada medición de su peso durante todo el estudio (Cada pollito recibe uno y solo un tipo de dieta).

Detalles:

Los pesos de los pollitos son medidos al nacer y cada dos días, hasta el día 20. También fueron medidos el día 21.Hubieron 4 grupos de pollitos con diferentes dietas.

Para observar el crecimiento de los pollitos de manera individual se utilizó `plot(ChickWeight)`

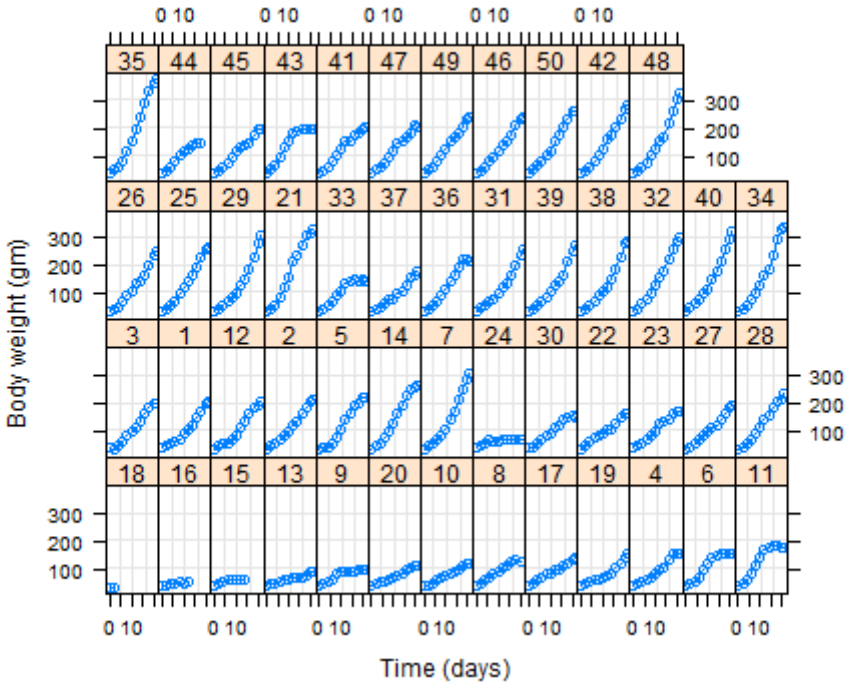


Figura 1: Crecimiento de cada pollito a través del tiempo.
Crowder, M. and Hand, D. (1990), Analysis of Repeated Measures, Chapman and Hall (example 5.3)...

El crecimiento de los pollitos en función del tiempo en un mismo plano

plot(weight~Time,ChickWeight,col=Chick,pch=as.numeric(Chick))

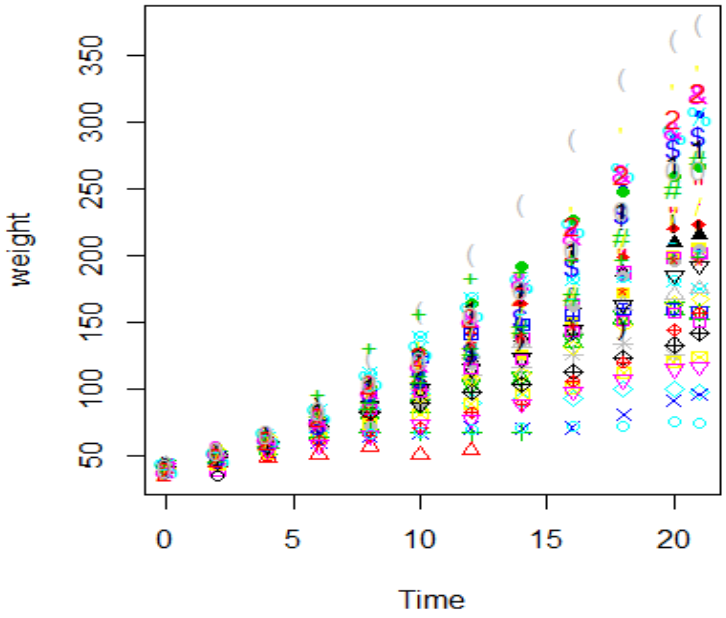


Figura 2: Crecimiento de los pollitos a través del tiempo en un mismo plano. *Crowder, M. and Hand, D. (1990), Analysis of Repeated Measures, Chapman and Hall (example 5.3)...*

Modelo de intercepto aleatorio

$$weight_{ij} = \beta_0 + \beta_1 Time_{ij} + b_i + \varepsilon_{ij},\ \ j = 1, \dots, n_i, i = 1, \dots, N$$

```
m1=lme(weight~Time,ChickWeight,~1|Chick)
summary(m1)
```

```
Linear mixed-effects model fit by REML
 Data: ChickWeight
       AIC      BIC    logLik
  5627.398 5644.822 -2809.699

Random effects:
 Formula: ~1 | Chick
        (Intercept) Residual
StdDev:    26.79274 28.27404

Fixed effects: weight ~ Time
                Value Std.Error  DF  t-value p-value
(Intercept) 27.845104  4.387674 527  6.34621       0
Time         8.726062  0.175518 527 49.71592       0
 Correlation:
     (Intr)
Time -0.422
```

```
abline(27.845104,8.726062)
```

```
Los coeficientes explicativos del modelo son:
```

$$weight_{ij} = 27.845 + 8.726062 Time_{ij} + b_i + \varepsilon_{ij},\ \ j = 1, \dots, n_i, i = 1, \dots, N$$

Los coeficientes del modelo lineal de cada pollito se encuentran en el anexo.

Según el modelo 1 existe fuerte evidencia estadística que los pollitos tienen su propio intercepto considerando que tienen la misma pendiente.

El modelo de coeficientes aleatorios con intercepto de tiempo por pollito es:

$$weight_{ij} = (\beta_0 + b_{i1}) + (\beta_1 + b_{i2}) Time_{ij} + \varepsilon_{ij},\ \ j = 1, \dots, n_i, i = 1, \dots, N$$

```
m3=lme(weight~Time,ChickWeight,~Time|Chick)
summary(m3)
```

```
Linear mixed-effects model fit by REML
 Data: ChickWeight
       AIC      BIC    logLik
  4839.499 4865.636 -2413.75

Random effects:
 Formula: ~Time | Chick
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 11.854721 (Intr)
Time         3.760789 -0.951
Residual    12.786927

Fixed effects: weight ~ Time
               Value Std.Error  DF  t-value p-value
(Intercept) 29.177998 1.9572600 527 14.90757       0
Time         8.453052 0.5408264 527 15.62988       0
 Correlation:
     (Intr)
Time -0.871

Number of Observations: 578
Number of Groups: 50
```

El valor p del intercepto y el tiempo son adecuados esto significa que cada pollito tiene su propio peso de inicio y su propio crecimiento. Es decir que cada pollito tiene una forma diferente de crecimiento.

Modelo 4:

Modelo de Coeficientes aleatorios con intercepto tiempo y pollito añadiendo el efecto por dieta es:

$$weight_{ij} = (\beta_0 + b_{i1}) + (\beta_1 + b_{i2})Time_{ij} + \beta_2 Diet + \varepsilon_{ij},\ \ j = 1, \dots, n_i, i = 1, \dots, N$$

```
Linear mixed-effects model fit by REML
 Data: ChickWeight
       AIC      BIC    logLik
  4821.754 4860.912 -2401.877

Random effects:
 Formula: ~Time | Chick
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 12.404362 (Intr)
Time         3.759579 -0.981
Residual    12.784862

Fixed effects: weight ~ Time + Diet
               Value Std.Error  DF   t-value p-value
(Intercept) 26.356156 2.2907133 527 11.505655  0.0000
Time         8.443779 0.5403076 527 15.627725  0.0000
Diet2        2.838620 2.3626972  46  1.201432  0.2357
Diet3        2.004432 2.3626972  46  0.848366  0.4006
Diet4        9.254785 2.3657267  46  3.912026  0.0003
 Correlation:
      (Intr) Time   Diet2  Diet3
Time  -0.796
Diet2 -0.351 -0.004
Diet3 -0.351 -0.004  0.344
Diet4 -0.350 -0.005  0.343  0.343

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-2.76169054 -0.57577683 -0.03526591  0.47887989  3.50246956

Number of Observations: 578
Number of Groups: 50
```

Los coeficientes tienen un valor p< 0.05 lo que indica la diferencia de la forma del crecimiento de los pollos dependiendo de su dieta.

CRITERIO DE INFORMACIÓN DE AKAIKE entre eL modelo 3 y el modelo 4 nos indica lo siguiente

	df	AIC
m3	6	4839.499
m4	9	4821.754

AIC = -2 l -

Y la tabla ANOVA entre el modelo 3 y el modelo 4 nos da los siguientes resultados

Model	df	AIC	BIC	logLik
1	6	4839.499	4865.636	-2413.75
2	9	4821.754	4860.912	-2401.877

El valor p<0.0001

Para culminar y apreciar de manera detallada tomaremos un pollo y observaremos su crecimiento con los efectos de peso y tiempo del pollito. Detalles del modelo logístico al final de los anexos.

```
plot(weight~Time,ChickWeight,subset = Chick=="21")

abline(26.356156+2.838620-22.66139027,8.443779+7.5100731)
```

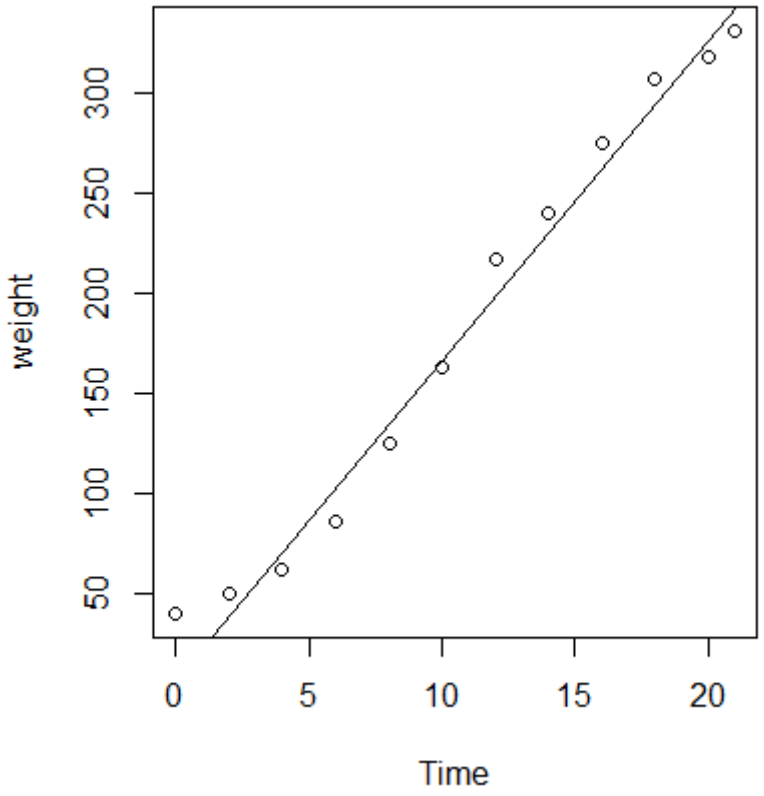


Figura 3: Crecimiento de un pollitos con los efectos de peso y tiempo. *Crowder, M. and Hand, D. (1990), Analysis of Repeated Measures, Chapman and Hall (example 5.3)...*

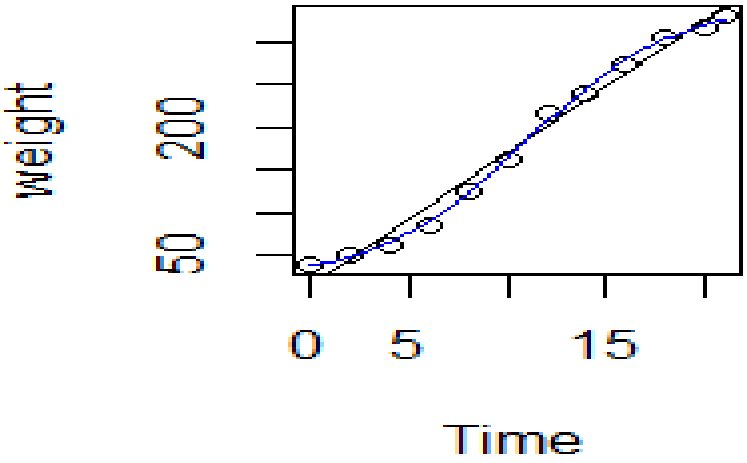


Figura 4: Crecimiento de un pollitos con los efectos de peso y tiempo. Modelo Lineal y Modelo Logístico de cuatro parámetros. *Crowder, M. and Hand, D. (1990), Analysis of Repeated Measures, Chapman and Hall (example 5.3..*

Los coeficientes de los modelos lineales quedan así:

>	coef(m3)	
	(Intercept)	Time
18	31.58	7.53
16	48.96	0.93
15	47.36	2.04
13	46.63	2.09
9	46.85	3.02
20	42.10	3.50
10	41.42	3.94
8	40.15	5.09
17	40.90	4.69
19	37.81	4.72
4	35.57	5.95
6	36.49	6.85

11	34.23	8.33
3	28.22	8.19
1	29.65	7.68
12	28.16	8.07
2	27.87	8.53
5	23.43	9.66
14	19.18	12.04
7	14.02	12.69
24	50.59	1.41
30	36.95	6.05
22	37.14	6.07
23	34.90	6.91
27	31.94	7.26
28	25.33	9.61
26	23.86	9.90
25	20.73	11.22
29	15.89	11.83
21	9.81	15.77
33	38.03	6.34
37	33.64	6.45
36	24.89	10.04
31	23.82	9.74
39	21.76	10.44
38	17.56	11.63
32	15.22	13.06
40	14.16	13.21

34	9.46	14.69
35	3.82	17.25
44	36.19	6.98
45	32.02	7.91
43	32.85	9.47
41	31.38	8.63
47	30.43	8.74
49	26.41	10.02
46	25.79	9.85
50	21.26	11.46
42	19.45	11.84
48	13.06	13.37

Tabla1: Coeficientes del Crecimiento de los pollitos

En donde se observa solo el tiempo que se toma cada pollo en crecer dependiendo de la dieta que tome lo cual se puede observar su comportamiento en el siguiente grafico

plot(ranef(m4))

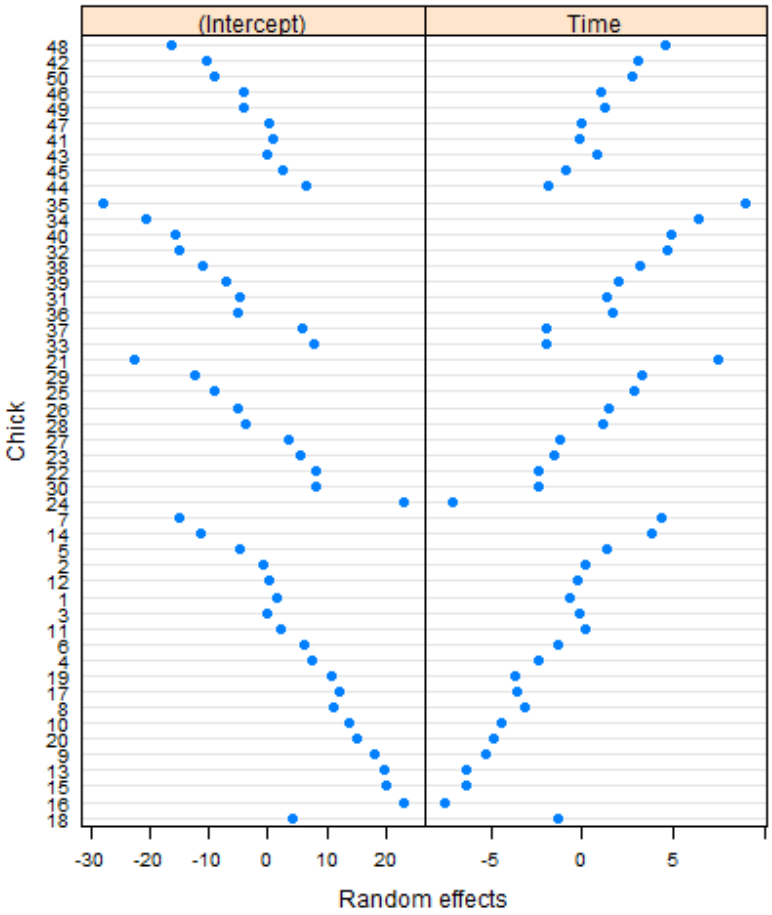


Figura 4: Crecimiento de los pollitos dependiendo de la dieta que tomen. *Crowder, M. and Hand, D. (1990), Analysis of Repeated Measures, Chapman and Hall (example 5.3)...*

VI. MODELO LME MULTIDIMENSIONAL (10)

Para fijar la idea tomamos un ejemplo con peso, altura y tamaño de pie de los miembros de las familias, entonces es razonable suponer que tenemos dos modelos de interceptos aleatorios; un modelo mixto para el peso en función de la altura y un modelo mixto para el tamaño del pie así mismo en función de la altura, de los miembros de las familias.

$$W_i = \propto_1 + \beta_1 H_i + b_{i1} + \varepsilon_{i1}$$

$$S_i = \propto_2 + \beta_2 H_i + b_{i2} + \varepsilon_{i2}$$

$$b_{i1} \sim \mathcal{N}(0, \sigma^2 d_{1)} \quad , \quad b_{i2} \sim \mathcal{N}(0, \sigma^2 d_{2)}$$

$$\varepsilon_{i1} \sim \mathcal{N}(0, \sigma^2 I) \quad , \quad \varepsilon_{i2} \sim \mathcal{N}(0, \sigma^2 I)$$

Algunos supuestos pueden hacerse:

1.- El supuesto más simple es que los dos modelos LME son independientes.

2.- Los errores de los modelos en una misma familia están correlacionados. Pero b_{i1} y b_{i2} no. Podemos suponer que para cada familia la correlación entre la desviación del peso y tamaño del pie es la misma.

3.- e_{i1} y e_{i2} están correlacionados, pero b_{i1} y b_{i2} también están correlacionados.

Luego además de la distribución para e_i el par $(b_{i1}, b_{i2})'$ tiene distribución normal en dos dimensiones.

(10) Modelo lineal mixto: Tema traducido y verificado del libro [1] Mixed Models Theory and Applications de Eugene Demidenko, Dartmouth College, Wiley Interscience, 2004.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES.-

1.- El modelo de efectos mixtos permite estudiar el comportamiento de una variable de respuesta en función de variables explicativas y variables aleatorias.

2.- Las variables explicativas permiten identificar el nivel de la variable de respuesta pero las variables aleatorias permiten considerar la correlación existente en la misma familia o individuo.

3.- Las fórmulas de reducción en muchos casos permiten simplificar los cálculos para encontrar las inversas de las matrices.

4.- En el modelo de la curva de crecimiento lineal (LGC) para el caso de que D=dI para d=0, el estimador β se reduce a β_{OLS} (mínimo cuadrado ordinario aplicado a la segunda etapa del modelo 4.1, 4.2 $W_i = I$.

RECOMENDACIONES.-

1.- Continuar estudiando los modelos mixtos (LME) considerando estructuras de covarianza lineal.

2.- Plantear y estimar modelos lineales, con efectos aleatorios multidimensionales(MLME) para modelar problemas con respuestas vectoriales.

3.- Aplicar estos modelos mixtos en problemas relacionados con tratamientos a pacientes, producción, temas educativos, etc.

BIBLIOGRAFÍA

(1) Eugene Demidenko, Mixed Models Theory and Applications, Dartmouth College, Wiley Interscience, 2004 PAG 63

(2) Pinheiro, J.C., Bates, D.M. 2000. Mixed-Effects Models in S and S-PLUS. Statistics and Computing. Springer USA. 528 pag.

(3) Casanoves, F. (2004). Análisis de ensayos comparativos de rendimiento en mejoramiento vegetal en el marco de los modelos lineales mixtos. Tesis Doctoral. Escuela para Graduados, Facultad de Ciencias Agropecuarias, Univ. Nacional de Córdoba.

(4) Curso de Modelos Mixtos Utilizando R. Llorenç Badiella y Josep Antón Sánchez. http://eio.usc.es/pub/gridecmb/index.php/informacion-general.

ANEXOS

Matrices Definidas no negativas y definidas positivas

La relación de orden entre matrices $A \leq B$ $(A < B)$ significa que la matriz B-A es definida no negativa (positiva).

Definición. Una matriz C cuadrada es llamada definida no negativa si $a'Ca \geq 0$ para todo vector a. (C se dice definida positiva si $a'Ca > 0$).

Para que el modelo lme sea identificable para β, suponemos que la matriz $\sum X_i' X_i$ es no singular y que $\sum n_i > m$. Para hacer el modelo LME identificable para σ^2 y D, suponemos que al menos una matriz $Z_i' Z_i$ es definida positiva y

$$\sum_{i=1}^{N} (n_i - k) > 0$$

Identificable.- es una propiedad necesaria para que la inferencia sea posible. El modelo es identificable si teóricamente se puede obtener el verdadero valor del parámetro fundamental de este modelo después de obtener un número infinito de observaciones. Esto equivale a decir que diferentes valores del parámetro deben generar diferentes distribuciones de las variables observadas.

2 I.I. MODELO LINEAL CON INTERCEPTO ALEATORIO

Para cada individuo j de la familia i se tiene que:

$$W_{ij} = \propto + \beta H_{ij} + b_i + \varepsilon_{ij},$$

$$E(W_{ij}) = E(\propto + \beta H_{ij} + b_i + \varepsilon_{ij}) = \propto + \beta H_{ij},$$

Recordemos que $E(b_i) = 0$ y $E(\varepsilon_{ij}) = 0$

$$V(W_{ij}) = V(\propto + \beta H_{ij} + b_i + \varepsilon_{ij})$$

$$V(W_{ij}) = V(b_i + \varepsilon_{ij}) = V(b_i) + V(\varepsilon_{ij}) + 2cov(b_i, \varepsilon_{ij})$$

Pero $cov(b_i, \varepsilon_{ij}) = 0, porque\ son\ independientes$

De modo que $V(W_{ij}) = V(b_i + \varepsilon_{ij}) = V(b_i) + V(\varepsilon_{ij}) = \sigma_d^2 + \sigma^2$

$$cov(W_{ij}, W_{ik}) = E\left[(W_{ik} - E(W_{ik}))\left(W_{ij} - E(W_{ij})\right)\right]$$

$$cov(W_{ij}, W_{ik}) = E[(b_i + \varepsilon_{ik})(b_i + \varepsilon_{ij})]$$

$$cov(W_{ij}, W_{ik}) = E(\ b_i{}^2\ ) + E(\ b_i\ \varepsilon_{ik}\ ) + E(\ b_i\ \varepsilon_{ij}\ ) + E(\ \varepsilon_{ik}\ \varepsilon_{ij})$$

Pero $E(\ b_i\ \varepsilon_{ik}\ ) = 0,\ E(\ b_i\ \varepsilon_{ij}\ ) = 0\ y\ E(\ \varepsilon_{ik}\ \varepsilon_{ij}) = 0$

$$tal\ que\ \ cov(W_{ij}, W_{ik}) = \sigma_d^2$$

El coeficiente de correlación queda así

$$corr(W_{ij}, W_{ik}) = \frac{cov(W_{ij}, W_{ik})}{\sqrt{V(W_{ij})}\sqrt{V(W_{ik})}} = \frac{\sigma_d^2}{\sqrt{\sigma_d^2 + \sigma^2}\sqrt{\sigma_d^2 + \sigma^2}} = \frac{\sigma_d^2}{\sigma_d^2 + \sigma^2}$$

Y la matriz de covarianzas de W_i es

$$cov(W_i) = \begin{bmatrix} \sigma_d^2 + \sigma^2 & \dots & \sigma_d^2 \\ \dots & \dots & \dots \\ \sigma_d^2 & \dots & \sigma_d^2 + \sigma^2 \end{bmatrix}$$

La matriz de covarianzas de $\boldsymbol{W_i\ es\ de\ la\ forma\ \sigma^2 V_i}$, donde la matriz simétrica $n_i x n_i\ \ V_i$es de la forma

$$V_i = \begin{bmatrix} 1 + \sigma_d^2/\sigma^2 & \dots & \sigma_d^2/\sigma^2 \\ \dots & \dots & \dots \\ \sigma_d^2/\sigma^2 & \dots & 1 + \sigma_d^2/\sigma^2 \end{bmatrix}$$

Claramente, esta estructura de covarianzas implica equi-correlación (simetría compuesta)

$$corr(W_{ij}, W_{ik}) = \rho = \sigma_d^2/(\sigma^2 + \sigma_d^2)$$

Equi-correlación: Igual correlación entre cada dos elementos.

El punto clave del modelo de efectos mixtos es que podemos obtener estimaciones más eficientes para $\propto y\ \beta$, minimizando la suma de los cuadrados ponderado.

$$\sum_{i=1}^{19} (W_i - \propto - \beta H_i)' V_i^{-1} (W_i - \propto - \beta H_i)$$

lo cual produce la estimación mediante mínimos cuadrados generalizados.

En este capítulo se supone que $b_i\ y\ \varepsilon_i$ son normalmente distribuidas como:

$$b_i \sim \mathcal{N}(0, \sigma^2 D), \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2 I_{n_i})$$

El modelo de efectos mixtos

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i, \qquad \mathrm{i} = 1, \dots, \mathrm{N}$$

Es tal que

$$E(y_i) = E(X_i\beta) + Z_i E(b_i) + E(\varepsilon_i)$$

Pero $E(b_i) = 0, E(\varepsilon_i) = 0$, entonces

$$E(y_i) = E(X_i\beta) = X_i\beta$$

$$var(y_i) = var(X_i\beta) + var(Z_i b_i + \varepsilon_i)$$

$$var(y_i) = 0 + var(Z_i b_i + \varepsilon_i)$$

$$var(y_i) = var(Z_i b_i) + var(\varepsilon_i) + 2cov(Z_i b_i, \varepsilon_i)$$

Pero $cov(Z_i b_i, \varepsilon_i) = 0$, entonces

$$var(y_i) = var(Z_i b_i) + var(\varepsilon_i)$$

Esto es,

$$var(y_i) = var(Z_i b_i + \varepsilon_i) = Z_i var(b_i) Z_i{}' + var(\varepsilon_i)$$

Pero $var(b_i) = \sigma^2 D$, entonces remplazando, tenemos,

$$= Z_i \sigma^2 D Z_i{}' + \sigma^2 I$$

$$= \sigma^2 (Z_i D Z_i{}' + I) = \sigma^2 V_i$$

$$var(y_i) = \sigma^2 V_i$$

Entonces la matriz V queda así,

$$V = var(\eta) = \sigma^2 \begin{bmatrix} I_{n_1} + Z_1 D Z_1' & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & I_{n_N} + Z_N D Z_N' \end{bmatrix}, \qquad (2.12)$$

De aquí en adelante, el modelo lme escrito como una ecuación matricial (2.7) o (2.10) será referido como notación “larga”.

Por lo tanto, el modelo LME especificado por (2.4)[1] con variables aleatorias distribuidas normalmente puede ser escrita en forma marginal como

$$y_i \sim \mathcal{N}\left(X_i\beta, \sigma^2\left(I_{n_i} + Z_i D Z_i'\right)\right), \qquad \text{i} = 1, \dots \text{N} \quad (2.13)$$

Usamos la fórmula

$$(A - cbb')^{-1} = A^{-1} + c\frac{A^{-1}bb'A^{-1}}{1 - cb'A^{-1}b}, c\ es\ escalar$$

Para la matriz inversa en (2.74). Usando esta fórmula obtenemos

$$\left(X'X - \frac{n^2 d}{1 + nd}\bar{x}\bar{x}'\right)^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}\bar{x}\bar{x}'(X'X)^{-1}}{1/c - \bar{x}'(X'X)^{-1}\bar{x}} \quad (2.75)$$

MODELO INTERCEPTO ALEATORIO BALANCEADO

Donde $c = \frac{n^2 d}{1+nd}$. El siguiente hecho simple es muy útil en el estudio de los modelos con interceptos aleatorios.

Lema 3. Sea $X = [1_n, U]$ una matriz nxm de rango completo, donde $1_n = (1, 1, \dots 1)'$ es un vector columna nx1, $0_{(m-1)}$ es un vector cero, y U es una matriz nx(m-1) de rango completo, y $\bar{x} = X'1_n/n$ es un vector de promedios mx1. Entonces:

a) $(X'X)^{-1}X'1_n = (1, 0'_{(m-1)})'$
b) $X(X'X)^{-1}X'1_n = 1_n$
c) $1_n{}'X(X'X)^{-1}X'1_n = \mathrm{n}$
d) $\bar{x}'(X'X)^{-1}\bar{x} = \frac{1}{\mathrm{n}}$

Demostración de a)

Suponer que $(X'X)^{-1}$ existe,

Multiplicar ambos miembros por $X'X$,

$$X'X(X'X)^{-1}X'1_n = X'X(1, 0'_{(m-1)})', esto\ es,$$

$$X'1_n = X'X(1, 0'_{(m-1)})'$$

Ahora,

$$X'X = [1_n, U]'[1_n, U] = \begin{bmatrix} 1_n{}'1_n & 1_n{}'U \\ \mathrm{U}'1_n & \mathrm{U}'U \end{bmatrix}, y$$

$$X'1_n = [1_n, U]'1_n = \begin{bmatrix} 1_n{}' \\ \mathrm{U}' \end{bmatrix} 1_n = \begin{bmatrix} 1_n{}'1_n \\ \mathrm{U}'1_n \end{bmatrix} \begin{bmatrix} n \\ \mathrm{U}'1_n \end{bmatrix}$$

Entonces, $X'X(1, 0'_{(m-1)})' = \begin{bmatrix} n & 1_n{}'U \\ \mathrm{U}'1_n & \mathrm{U}'U \end{bmatrix} \begin{bmatrix} 1 \\ 0_{(m-1)} \end{bmatrix} \begin{bmatrix} n+0 \\ \mathrm{U}'1_n + 0_{(m-1)} \end{bmatrix} = \begin{bmatrix} n \\ \mathrm{U}'1_n \end{bmatrix} = X'1_n.$

Lqqd.

La igualdad en b) sigue directamente de a)

$$X(X'X)^{-1}X'1_n = [1_n, U] \begin{bmatrix} 1 \\ 0_{(m-1)} \end{bmatrix} = 1_n$$

Las sentencias c) y d) siguen de b).

Debido a la sentencia b) del lema 3, el denominador en (2.75) se simplifica. Así, sea $M = (X'X)^{-1}$, obtenemos

$$\hat{\beta}_{GLS} = \left((X'X)^{-1} + \frac{(X'X)^{-1}\bar{x}\bar{x}'(X'X)^{-1}}{1/c - \bar{x}'(X'X)^{-1}\bar{x}} \right) \left(X'\bar{y} - \frac{n^2 d}{1+nd}\bar{x}\bar{\bar{y}} \right)$$

$$\hat{\beta}_{GLS} = \left(M + \frac{M\bar{x}\bar{x}'M}{1/c - 1/n}\right)\left(X'\bar{y} - \frac{n^2 d}{1+nd}\bar{x}\bar{\bar{y}}\right)$$

$$\hat{\beta}_{GLS} = MX'\bar{y} - M\frac{n^2 d}{1+nd}\bar{x}\bar{\bar{y}} + \frac{M\bar{x}\bar{x}'M}{1/(n^2 d)}X'\bar{y} - \frac{M\bar{x}\bar{x}'M}{1/(n^2 d)}\frac{n^2 d}{1+nd}\bar{x}\bar{\bar{y}}$$

$$\hat{\beta}_{GLS} = MX'\bar{y} - \frac{n^2 d}{1+nd}M\bar{x}\bar{\bar{y}} + n^2 dM\bar{x}\bar{x}'MX'\bar{y} - \frac{n^4 d^2}{1+nd}M\bar{x}\bar{x}'M\bar{x}\bar{\bar{y}}$$

$$\hat{\beta}_{GLS} = MX'\bar{y} - \frac{n^2 d}{1+nd}M\bar{x}\bar{\bar{y}} + n^2 dM\bar{x}\bar{x}'MX'\bar{y} - \frac{n^4 d^2}{1+nd}M\bar{x}(1/n)\bar{\bar{y}}$$

$$\hat{\beta}_{GLS} = MX'\bar{y} - \frac{n^2 d}{1+nd}M\bar{x}\bar{\bar{y}} + n^2 dM\bar{x}\bar{x}'MX'\bar{y} - \frac{n^3 d^2}{1+nd}M\bar{x}\bar{\bar{y}}$$

$$\hat{\beta}_{GLS} = MX'\bar{y} - \left(\frac{n^2 d}{1+nd} + \frac{n^3 d^2}{1+nd}\right)M\bar{x}\bar{\bar{y}} + n^2 dM\bar{x}\bar{x}'MX'\bar{y}$$

Pero $\bar{x}'MX'\bar{y} = \bar{\bar{y}}$ es un número

$$\hat{\beta}_{GLS} = MX'\bar{y} - \left(\frac{n^2 d}{1+nd}\bar{\bar{y}} + \frac{n^3 d^2}{1+nd}\bar{\bar{y}} - n^2 d\bar{x}'MX'\bar{y}\right)M\bar{x}$$

$$\hat{\beta}_{GLS} = MX'\bar{y} - \left(\frac{n^2 d}{1+nd}\bar{\bar{y}} + \frac{n^3 d^2}{1+nd}\bar{\bar{y}} - \frac{n^2 d\bar{x}'MX'\bar{y}}{1+nd} - \frac{n^3 d^2\bar{x}'MX'\bar{y}}{1+nd}\right)M\bar{x}$$

$$\hat{\beta}_{GLS} = MX'\bar{y} - \left(\frac{n^2 d}{1+nd}\bar{\bar{y}} - \frac{n^2 d\bar{x}'MX'\bar{y}}{1+nd}\right)M\bar{x}$$

$$\hat{\beta}_{GLS} = MX'\bar{y} - \frac{n^2 d}{1+nd}(\bar{\bar{y}} - \bar{x}'MX'\bar{y})M\bar{x} \quad (2.77)$$

Donde $\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$. Observamos que $\bar{\bar{y}} - \bar{x}'MX'\bar{y} = 0$ sigue directamente de b) del lema 3. De hecho,

$$\bar{\bar{y}} - \bar{x}'MX'\bar{y} = \left(\frac{1}{n} - \bar{x}'MX'\right)\bar{y} = n^{-1}(1_n - 1_n{}'X(X'X)^{-1}X')\bar{y} = 0.$$

Así, el segundo término de (2.77) desaparece, y obtenemos:

$$\hat{\beta}_{GLS} = \hat{\beta}_{OLS} = MX'\bar{y}$$

De la teoría de la regresión estándar con el término intercepto, se tiene que el estimado OLS para las pendientes e interceptos pueden ser expresado más adelante como

$$\hat{\gamma}_{OLS} = \left(\tilde{U}'\tilde{U}\right)^{-1}\tilde{U}'\bar{y}, \;\; \hat{\alpha}_{OLS} = \bar{\bar{y}} - \bar{u}'\hat{\gamma}_{OLS}$$

Donde $\tilde{U}$ es la matriz centrada, $\tilde{U} = U - 1_n\bar{u}'$, donde $\bar{u} = \frac{U'1_n}{n}$.

Ahora trabajamos sobre σ^2 y d. Ya que $\hat{\beta}_{ML} = \hat{\beta}_{OLS}$ no depende de d, de (2.37)

$$\hat{\sigma}^2 = \frac{1}{N_T}\sum_{i=1}^{N}(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta),$$

obtenemos

$$\hat{\sigma}^2{}_{ML} = \frac{1}{Nn}\sum_{i=1}^{N}\left(y_i - X\hat{\beta}_{OLS}\right)'\left(I - \frac{\hat{d}_{ML}}{1 + n\hat{d}_{ML}}\right)(y_i - X\hat{\beta}_{\mathrm{OLS}})$$

$$\hat{\sigma}^2{}_{ML} = \frac{1}{Nn}\left[\sum_{i=1}^{N}\left\|y_i - X\hat{\beta}_{OLS}\right\|^2 - \frac{\hat{d}_{ML}}{1 + n\hat{d}_{ML}}\sum_{i=1}^{N}\left[\left(y_i - X\hat{\beta}_{OLS}\right)'1_n\right]^2\right]$$

$$\hat{\sigma}^2{}_{ML} = \frac{1}{Nn}\left[S - \frac{\hat{d}_{ML}n^2A}{1 + n\hat{d}_{ML}}\right], \qquad (2.78)$$

Donde $\sum_{i=1}^{N}\|\hat{e}_i\|^2$, $\hat{e}_i = y_i - X\hat{\beta}_{OLS}$ es un vector nx1 de los residuos de OLS y $A = \sum_{i=1}^{N}(\bar{y}_i - \bar{\bar{y}})^2$ como en la sección precedente. De (2.73) tenemos

$$\hat{d}_{ML} = \frac{n^2A - S}{n(S - nA)} = \frac{1}{N\hat{\sigma}^2{}_{ML}}\sum_{i=1}^{N}(\bar{y}_i - \bar{\bar{y}})^2 - \frac{1}{n}\,, \qquad (2.79)$$

donde

$$\hat{\sigma}^2{}_{ML} = \frac{S - nA}{N(n-1)} = \frac{1}{N(n-1)}\sum_{i=1}^{N}(\|\hat{e}_i\|^2 - n(\bar{y}_i - \bar{\bar{y}})^2)\,, \qquad (2.80)$$

Ahora encontramos el RML estimado para un modelo intercepto aleatorio balanceado. Usando la fórmula para el determinante

$$|A - abb'|(1 - ab'A^{-1}b),$$

donde a es un escalar, A es una matriz y b es un vector.

a) Y el hecho de que $\bar{x}'(X'X)^{-1}\bar{x} = \frac{1}{\mathrm{n}}$, podemos simplificar la función log(verosimilitud (2.71) a una función equivalente

$$-\frac{1}{2}\left\{N\ln(1 + nd) + (Nn - m)\ln\left(S - d\frac{n^2A}{1 + nd}\right) - \ln(1 + nd)\right\}.$$

Permitiendo $v = 1 + nd$ y tomándola derivada con respecto a v , llegamos a la solución forma-cerrada

$$\hat{d}_{RML} = \frac{[N(n-1) - m + 1]An - (N-1)(S - An)}{n(N-1)(S - An)}. \quad (2.81)$$

Análogo a (2.78), *obtenmos la relación entre los estimados RML de* σ^2 *y* d,

$$\hat{\sigma}^2{}_{RML} = \frac{1}{Nn-m}\left(S - \hat{d}_{RML}\frac{n^2 A}{1+n\hat{d}_{RML}}\right). \quad (2.82)$$

Sustituyendo (2.81) en (2.82), finalmente obtenemos el estimador RML de forma cerrada para $\sigma^2\ y\ d, y\ \ d_* = \sigma^2 d$,

$$\hat{\sigma}^2{}_{RML} = \frac{1}{N(n-1)-m+1}\sum_{i=1}^{N}(\|\hat{e}_i\|^2 - n(\bar{y}_i - \bar{\bar{y}})^2). \quad (2.83)$$

$$\hat{d}_{RML} = \frac{1}{(N-1)\hat{\sigma}^2{}_{ML}}\sum_{i=1}^{N}(\bar{y}_i - \bar{\bar{y}})^2 - \frac{1}{\mathrm{n}}, \quad (2.84)$$

$$\hat{d}_{*RML} = \frac{1}{(N-1)}\sum_{i=1}^{N}(\bar{y}_i - \bar{\bar{y}})^2 - \frac{\hat{\sigma}^2{}_{RML}}{\mathrm{n}}. \quad (2.85)$$

Hacemos unos comentarios para las varianzas de los parámetros en el ML restringido estimado. Generalmente, RML son cerrados para el ML estándar estimado. La diferencia está en los grados de libertad del denominador. Para σ^2, el denominador es ajustado por el número de pendientes de efectos fijos (m-1) y para d por 1. Como aprenderemos en la sección 3.14 el RMLE para el modelo intercepto aleatorio con datos balanceados son insesgados y coincide con otros estimadores cuadráticos insesgados: norma mínima, método de los momentos y cuadrados de mínima varianza. Interesantemente, para el modelo de coeficientes aleatorios balanceados de la sección precedente, ML=RML para σ^2, pero ellos son diferentes para el modelo interceptos aleatorios balanceado.

FUNCIÓN LOGARITMO DE LA VERSOSIMILITUD CON EL SUPUESTO DE NORMALIDAD.

La función de densidad de las componentes del vector y_i es

$$f_i(y_i) = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_i}\frac{1}{\sqrt{\left|\sigma^2(I_{n_i}+Z_i D Z_i')\right|}}e^{-\frac{1}{2}\sigma^{-2}(y_i - X_i\beta)'(I_{n_i}+Z_i D Z_i')^{-1}(y_i - X_i\beta)}$$

La función de verosimilitud del vector Y en la notación larga es

$$L(\beta, \sigma^2, D) = \prod_{i=1}^{N} f_i(y_i)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{N_T}\frac{1}{\sqrt{\prod_{i=1}^{N}\left|\sigma^2(I_{n_i}+Z_i D Z_i')\right|}}e^{-\frac{1}{2}\sigma^{-2}\sum_{i=1}^{N}(y_i - X_i\beta)'(I_{n_i}+Z_i D Z_i')^{-1}(y_i - X_i\beta)}$$

Función log-verosimilitud

$$l(\theta) = \ln\big(L(\beta, \sigma^2, D)\big)$$

$$l(\theta) = \{-\frac{N_T}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln\big|\sigma^2\big(I_{n_i} + Z_i D Z_i'\big)\big| -$$

$$\tfrac{1}{2}\sigma^{-2}\textstyle\sum_{i=1}^{N}[(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta)]\}$$

$$l(\theta) = \{-\frac{N_T}{2}\ln(2\pi) - \sum_{i=1}^{N} ln(\sigma^2)^{n_i} + \ln\big|\big(I_{n_i} + Z_i D Z_i'\big)\big| - \frac{1}{2}\sigma^{-2}\sum_{i=1}^{N}[(y_i - X_i\beta)'(I_{n_i} + Z_i D Z_i')^{-1}(y_i - X_i\beta)]\}$$

Es fácil mostrar que el espacio de parámetros Θ especificado por (2.15) es convexo. Más precisamente, Θ es un cono convexo:

$si\ \theta_1, \theta_2 \in \Theta\ entonces\ \lambda_1\theta_1 + \lambda_2\theta_2 \in \Theta, para\ cualquier\ escalar\ \lambda_1,$
$\lambda_2\ no\ negativos\ tal\ que\ \lambda_1, +\lambda_2 = 1.$

Esto sigue del hecho de que una combinación lineal no negativa de dos matrices definidas no negativas es una matriz definida no negativa.

Demostración de la convexidad de Θ

$$si\ \theta_1, \theta_2 \in \Theta\ entonces\ \lambda_1\theta_1 + \lambda_2\theta_2 \in \Theta$$

Lo que hay que demostrar es

$$\lambda_1\begin{pmatrix}\beta_1\\ \sigma^2{}_1\\ vech(D_1)\end{pmatrix} + \lambda_2\begin{pmatrix}\beta_2\\ \sigma^2{}_2\\ vech(D_2)\end{pmatrix} \in \Theta$$

$$\begin{pmatrix}\lambda_1\beta_1 + \lambda_2\beta_2\\ \lambda_1\sigma^2{}_1 + \lambda_2\sigma^2{}_2\\ \lambda_1 vech(D_1) + \lambda_2 vech(D_2)\end{pmatrix}$$

$$\lambda_1\beta_1 + \lambda_2\beta_2 \in \mathbb{R}^m$$

para $\lambda_1, \lambda_2 \geq 0, se\ tiene\ que\ \lambda_1\sigma^2{}_1 + \lambda_2\sigma^2{}_2 > 0$

En lo que respecta a la combinación de vectores $\lambda_1 vech(D_1) + \lambda_2 vech(D_2)$

Tenemos que D_1, D_2 son definidas no negativas

$$\lambda_1 D_1 + \lambda_2 D_2$$

$$x'(\lambda_1 D_1 + \lambda_2 D_2)x \geq 0$$

$$x'(\lambda_1 D_1)x + x'(\lambda_2 D_2)x \geq 0$$

$$\lambda_1(x' D_1 x) + \lambda_2(x' D_2 x) \geq 0$$

Pero

$$(x'D_1x) \geq 0, \qquad (x'D_2x) \geq 0$$

Por lo que queda demostrado que $\lambda_1 D_1 + \lambda_2 D_2$ es definida no negativa

Cabe señalar que el MLE para D puede estar en el límite del espacio de parámetros, lo que significa que la matriz D puede ser singular. Esta circunstancia complica al algoritmo de maximización debido, estrictamente hablando, la maximización de l trata con optimización restringida. Algunos remedios son discutidos en la sección 2.15.

La matriz de covarianzas escalada del vector de la variable dependiente y_i está dado por:

$$V_i = V_i(D) = I_{n_i} + Z_i D Z_i' \qquad (2.16)$$

Esta notación es usada a través del capítulo. Los efectos aleatorios inducen correlación dentro del cluster (grupo) entre componentes del vector y, debido a que la matriz V_i es no diagonal.

Usando la notación V_i, la función log-verosimilitud (2.14) puede ser re-escrita como

$$l(\theta) = -\frac{1}{2}\left\{N_T \ln \sigma^2 + \sum_{i=1}^{N}\left[\ln|V_i| + \sigma^{-2} e_i' V_i^{-1} e_i\right]\right\}, (2.15)$$

Donde $e_i = e_i(\beta) = y_i - X_i\beta, \qquad (2.18)$

es un vector residual $n_i x 1$, del i_ésimo cluster, i=1,…,N.

REDUCCIÓN DE DIMENSIÓN

Existen tres tipos equivalentes de parametrización para la función log-verosimilitud:

1. Parametrización de reducción de dimensión. La función log- verosimiltud estándar involucra el cálculo de la matriz inversa y el determinante de matrices $n_i x n_i$. Se puede reducir la dimensión a k usando las fórmulas de abajo.
2. Verosimilitud perfilada. Se puede tomar ventaja de que el vector óptimo de los efectos fijos β y la varianza dentro del sujeto σ^2 puede ser expresada por medio de la matriz D, así que ellos pueden ser eliminados de la función log-verosimilitud. Serán consideradas dos funciones perfiladas: perfil de varianza y perfil completo.
3. Parametrización de la Inversa de D. En la parametrización de reducción de dimensión, se nota que la función log-verosimilitud puede ser expresada por medio de D^{-1}, lo cual excluye la necesidad de invertir la matriz. **Desde un punto de vista computacional es la forma más económica de parametrización de la función log-verosimilitud de perfil completo vía la inversa de D.**

2.2.3 Fórmulas de reducción de dimensión

Nosotros usaremos las siguientes fórmulas de reducción de dimensión:

$$V_i^{-1} = (I_{n_i} + Z_i D Z_i')^{-1} = I_{n_i} - Z_i(I_k + D Z_i' Z_i)^{-1} D Z_i' = I_{n_i} - Z_i D (I_k + Z_i' Z_i D)^{-1} Z_i'$$

$$= I_{n_i} - Z_i(D^{-1} + Z_i'Z_i)^{-1}Z_i' . \qquad (2.19)$$

Estas fórmulas pueden ser verificadas por multiplicaciones directas (la última identidad se cumple si D es no singular).

Teorema. Sean A, B, C, D matrices tales que están definidas las operaciones que aparecen. Entonces,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(I + CDBA^{-1}B)^{-1}CDA^{-1}. \qquad (2.19.1)$$

Demostración:

$$(A + BCD)^{-1} = E$$

$$(A + BCD)^{-1}(A + BCD) = E(A + BCD)$$

$$I = EA + EBCD$$

$$A^{-1} = E + EBCDA^{-1}, entonces\ A^{-1} - EBCDA^{-1} = E$$

Pero multiplicando por B y despejando EB, se obtiene

$$A^{-1}B = EB + EBCDA^{-1}B$$

$$A^{-1}B = EB(I + CDA^{-1}B)$$

$$A^{-1}B(I + CDA^{-1}B)^{-1} = EB$$

Luego remplazamos en la ecuación de E para lograr la demostración

$$E = A^{-1} - EBCDA^{-1}$$

$$E = A^{-1} - A^{-1}B(I + CDA^{-1}B)^{-1}CDA^{-1}$$

Propiedades del producto de Kronecker

$(\mathrm{A} \otimes \mathrm{B})(\mathrm{C} \otimes \mathrm{D}) = (\mathrm{AC}) \otimes (\mathrm{BD})$ $\qquad$ $\mathrm{A} \otimes (\mathrm{B} + \mathrm{C}) = (\mathrm{A} \otimes \mathrm{B}) + (\mathrm{A} \otimes \mathrm{C})$

$(\mathrm{A} \otimes \mathrm{B})^1 = \mathrm{A}^1 \otimes \mathrm{B}^1$ $\qquad$ $(\mathrm{A} \otimes \mathrm{B})^{-1} = \mathrm{A}^{-1} \otimes \mathrm{B}^{-1}$

Una fórmula similar de reducción de dimensión se cumple para el determinante:

$$|V_i| = \left|I_{n_i} + Z_i D Z_i'\right| = |I_k + DZ_i'Z_i|. \quad (2.20)$$

Como vemos, el lado izquierdo involucra una matriz $n_i x n_i$, pero el lado derecho involucra a matrices kxk.

Si la matriz D es no singular, se puede expresar el log del determinante en función de D^{-1}, a saber,

$$ln|V_i| = \ln(|I_k + DZ_i'Z_i|) = \ln(|DD^{-1} + DZ_i'Z_i|)$$

$$= \ln(|D||D^{-1} + Z_i'Z_i|) = \ln|D^{-1} + Z_i'Z_i| - ln|D^{-1}|, \quad (2.21)$$

lo cual constituye la base para la parametrización de D^{-1}.

Otra fórmula matricial, que puede ser obtenida de (2.19), se usará:

$$Z_i'V_i^{-1}Z_i = (I_k + Z_i'Z_iD)^{-1}Z_i'Z_i$$

$$= Z_i'Z_i(I_k + DZ_i'Z_i)^{-1}, \quad (2.22)$$

Comprobación:

$$Z_i'V_i^{-1}Z_i = Z_i'\big(I_{n_i} - Z_iD(I_k + Z_i'Z_iD)^{-1}Z_i'\big)Z_i$$

$$= (Z_i'Z_i - Z_i'Z_iD(I_k + Z_i'Z_iD)^{-1}Z_i'Z_i)$$

$$= (I_k - Z_i'Z_iD(I_k + Z_i'Z_iD)^{-1})Z_i'Z_i$$

$$= \big((I_k + Z_i'Z_iD)(I_k + Z_i'Z_iD)^{-1} - Z_i'Z_iD(I_k + Z_i'Z_iD)^{-1}\big)Z_i'Z_i$$

$$= \big((I_k + Z_i'Z_iD) - Z_i'Z_iD\big)(I_k + Z_i'Z_iD)^{-1}Z_i'Z_i$$

$$= (I_k + Z_i'Z_iD)^{-1}Z_i'Z_i$$

Si la matriz $(Z_i'Z_i)^{-1}$ existe, las fórmulas (2.22) son más simples,

$$Z_i'V_i^{-1}Z_i = Z_i'(I_k + Z_iDZ_i')^{-1}Z_i = ((Z_i'Z_i)^{-1} + D)^{-1}. \quad (2.23)$$

Para verificar las fórmulas 2.19 se sugiere:

$$V_iV_i^{-1} = (I_{n_i} + Z_iDZ_i')\big(I_{n_i} + Z_iDZ_i'\big)^{-1} = (I_{n_i} + Z_iDZ_i')(I_{n_i} - Z_i(I_k + DZ_i'Z_i)^{-1}DZ_i')$$

Luego aplique propiedades distributivas para obtener:

$$= I_{n_i} + Z_iDZ_i' - Z_i(I_k + DZ_i'Z_i)^{-1}DZ_i' - Z_iDZ_i'Z_i(I_k + DZ_i'Z_i)^{-1}DZ_i'$$

$$= I_{n_i} + Z_iDZ_i' - Z_i[(I_k + DZ_i'Z_i)^{-1} + DZ_i'Z_i(I_k + DZ_i'Z_i)^{-1}]DZ_i'$$

$$= I_{n_i} + Z_iDZ_i' - Z_i[(I_k + DZ_i'Z_i)(I_k + DZ_i'Z_i)^{-1}]DZ_i'$$

$$= I_{n_i} + Z_iDZ_i' - Z_iDZ_i'$$

$$= I_{n_i}$$

L,q.q.d.

De forma similar se puede proceder con el resto de fórmulas

Demostración de 2.20

$$|V_i| = \left|I_{n_i} + Z_i D Z_i'\right| = |I_k + D Z_i' Z_i|$$

Consideremos el siguiente determinante de una matriz

$$|V_i| = \left|I_{n_i} + Z_i D Z_i'\right| = \left|I_{n_i} + Z_i I_k D Z_i'\right|$$

Haciendo el artificio de agregar la matriz identidad, el determinante de la matriz por bloque es

$$|I_k|\left|I_{n_i} + Z_i I_k D Z_i'\right| = |I_k|\left|I_{n_i} - (-Z_i)(I_k)^{-1} D Z_i'\right| = \begin{vmatrix} I_k & DZ_i' \\ -Z_i & I_{n_i} \end{vmatrix}$$

$$\begin{vmatrix} I_k & DZ_i' \\ -Z_i & I_{n_i} \end{vmatrix} = \left|I_{n_i}\right|\left|I_k + D Z_i' \left(I_{n_i}\right)^{-1} Z_i\right|$$

$$= |I_k + D Z_i' Z_i|$$

l.q.q.d.

FUNCIONES LOG-VEROSIMILITUD PERFILADA

Proposición 1. Sea y y X un vector nx1 y una matriz Nxm de rango completo (m<N), respectivamente, y V una matriz definida positiva NxN. Entonces

$$\min_{\beta \in R^m} (y - X\beta)' V^{-1} (y - X\beta)$$

$$= (y' V^{-1} X) - (X' V^{-1} y)' (X' V^{-1} X)^{-1} (X' V^{-1} y).$$

Demostración. El mínimo de la forma cuadrática de la izquierda se obtiene en $\hat{\beta} = (X' V^{-1} X)^{-1} (X' V^{-1} y)$, tal que el mínimo valor es $\left(y - X\hat{\beta}\right)' V^{-1} \left(y - X\hat{\beta}\right)$

$$= \left(y - X(X' V^{-1} X)^{-1} (X' V^{-1} y)\right)' V^{-1} \left(y - X(X' V^{-1} X)^{-1} (X' V^{-1} y)\right)$$

$$= y' \left[\left(I - X(X' V^{-1} X)^{-1} (X' V^{-1})\right)' V^{-1} \left(I - X(X' V^{-1} X)^{-1} (X' V^{-1})\right)\right] y$$

$$= y' [V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} (X' V^{-1})] y$$

$$= (y' V^{-1} y) - (X' V^{-1} y)' (X' V^{-1} X)^{-1} (X' V^{-1} y).$$

Lo cual prueba la proposición.

Donde q está definida en (2.42). Usando la fórmula de reducción de dimensión (2.19), las cantidades en (2.42) pueden ser encontradas como

$$\sum y_i' V_i^{-1} y_i = \sum y_i' (I_{n_i} - Z_i (D^{-1} + Z_i' Z_i)^{-1} Z_i') y_i$$

$$= \sum y_i' y_i - \sum (Z_i' y_i)' (Z_i (D^{-1} + Z_i' Z_i)^{-1} (Z_i' y_i)$$

$$\sum X_i' V_i^{-1} y_i = \sum X_i' (I_{n_i} - Z_i (D^{-1} + Z_i' Z_i)^{-1} Z_i') y_i$$

$$= \sum X_i' y_i - \sum (Z_i' X_i)' (Z_i (D^{-1} + Z_i' Z_i)^{-1} (Z_i' y_i)$$

$$\sum X_i' V_i^{-1} X_i = \sum X_i' (I_{n_i} - Z_i (D^{-1} + Z_i' Z_i)^{-1} Z_i') X_i$$

$$= \sum X_i' X_i - \sum (Z_i' X_i)' (Z_i (D^{-1} + Z_i' Z_i)^{-1} (Z_i' X_i). \quad (2.44)$$

MÁXIMA VEROSIMILITUD RESTRINGIDA

Sea el modelo lineal general definido como $y \sim \mathcal{N}(X\beta, V)$ donde X es la matriz nxm de rango completo y V es la matriz de covarianzas nxn, dependiente de algunos parámetros θ. En la estimación RML maximizamos la función log-verosimilitud para el vector residual $\hat{e} = y - X\hat{\beta}$, donde $\hat{\beta}$ es el estimador GLS, $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$. Note que ya que y tiene distribución normal también. Además, $\hat{\beta}$ y $\hat{e}$ son independientes porque

$$cov(X'V^{-1}y, \hat{e}) = cov\left(X'V^{-1}y, y - X\hat{\beta}\right) = cov(X'V^{-1}y, y - X(X'V^{-1}X)^{-1}X'V^{-1}y)$$

$$cov(X'V^{-1}y, \hat{e}) = X'V^{-1}[I - (X'V^{-1}X)^{-1}X'V^{-1}]\, V$$

$$cov(X'V^{-1}y, \hat{e}) = X'V^{-1}[V - (X'V^{-1}X)^{-1}X']$$

$$cov(X'V^{-1}y, \hat{e}) = X'V^{-1}V[I - V^{-1}X(X'V^{-1}X)^{-1}X']$$

$$cov(X'V^{-1}y, \hat{e}) = X' - X'V^{-1}X(X'V^{-1}X)^{-1}X'$$

$$cov(X'V^{-1}y, \hat{e}) = X' - X' = 0.$$

Esto implica que la función de verosimilitud para y es el producto de la función de verosimilitud para $\hat{e}$ y para $\hat{\beta}$. Pero $\hat{\beta} \sim \mathcal{N}(\beta, (X'V^{-1}X)^{-1})$, y por lo tanto la función log-verosimilitud para el vector residual $\hat{e}$, sin considerar la constante, es

$$L(y, \theta) = L(\hat{\beta}, \hat{e}, \theta)$$

$$L(y, \theta) = L(\hat{\beta}, \theta) L(\hat{e}, \theta)$$

$$l(y, \theta) = l\left(\hat{\beta}, \theta\right) + l(\hat{e}, \theta)$$

$$l(y, \theta) - l\left(\hat{\beta}, \theta\right) = l(\hat{e}, \theta)$$

2.3 Modelo de coeficientes aleatorios balanceados

Demostración

Iniciamos probando que el modelo (2.4) con los datos (2.51), el estimador GLS (2.26) colapsa al estimador OLS y por tanto aquello no depende de la matriz D. Más precisamente,

$$\hat{\beta}_{GLS} = \hat{\beta}_{OLS} = (Z'Z)^{-1}Z'\bar{y}, \quad (2.52)$$

Donde

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}.$$

Por definición, el estimador GLS minimiza la suma ponderada de cuadrados,

$$\sum_{i=1}^{N}(y_i - Z\beta)'V^{-1}(y_i - Z\beta) = \sum_{i=1}^{N}(y_i - \bar{y})'V^{-1}(y_i - \bar{y})$$

$$+2\sum_{i=1}^{N}(y_i - \bar{y})'V^{-1}(\bar{y} - Z\beta) + N(\bar{y} - Z\beta)'V^{-1}(\bar{y} - Z\beta) \quad (2.53)$$

Pero el término se elimina porque

$$\sum_{i=1}^{N}(y_i - \bar{y})'V^{-1}(\bar{y} - Z\beta) = \left[\sum_{i=1}^{N}(y_i - \bar{y})'\right]V^{-1}(\bar{y} - Z\beta) = 0.$$

Y por lo tanto la minimización de la suma ponderada es equivalente a la minimización de $(\bar{y} - Z\beta)'V^{-1}(\bar{y} - Z\beta)$, porque el primer término es constante, Además, denotando $\hat{\beta} = (Z'Z)^{-1}Z'\bar{y}$ y aplicando la fórmla de reducción de dimensión (2.19) para el último término de (2.53), obtenemos

$$(\bar{y} - Z\beta)'V^{-1}(\bar{y} - Z\beta) = \left(\hat{e} - Z(\beta - \hat{\beta})\right)'[I - Z(D^{-1} + Z'Z)^{-1}Z']\left(\hat{e} - Z(\beta - \hat{\beta})\right), \quad (2.54)$$

Donde $\hat{e} = \bar{y} - Z\hat{\beta}$ es un vector residual. Pero como en el modelo lineal estándar, los regresores y los residuos son ortogonales, $Z'\hat{e} = 0$, porque

$$Z'\hat{e} = Z'[I - Z(Z'Z)^{-1}Z']\bar{y} = 0.$$

De aquí, la suma (2.54) se simplifica a

$$\hat{e}'\hat{e} + \left(Z(\beta - \hat{\beta})\right)'V^{-1}\left(Z(\beta - \hat{\beta})\right)$$

Ahora encontramos los MLEs y RMLEs para σ^2 y D, Lair et al, (1987). Como aprenderemos más tarde en al capítulo 4, $\hat{\sigma}^2_{ML}$ y $\hat{\sigma}^2_{ML}\widehat{D}_{RML}$ son insesgados, pero $\hat{\sigma}^2_{ML}\widehat{D}_{ML}$ no lo es. Como una palabra de precaución, el sesgo del MLE restringido para los parámetros de la varianza no es una propiedad general –así es verdad solo para modelos balanceados.

Teorema 2. Los MLEs y RMLEs para modelos de coeficientes aleatorios balanceado ($Z = X_i = Z_i$) están dados por

$$\hat{\sigma}^2_{ML} = \hat{\sigma}^2_{RML} = \frac{1}{N(n-m)} \sum_{i=1}^{N} y_i'(I - Z(Z'Z)^{-1}Z')y_i, \quad (2.56)$$

$$\widehat{D}_{ML} = \frac{1}{N\hat{\sigma}^2_{ML}} (Z'Z)^{-1}Z'\hat{E}\hat{E}'Z(Z'Z)^{-1} - (Z'Z)^{-1}, \quad (2.57)$$

$$\widehat{D}_{RML} = \frac{1}{(N-1)\hat{\sigma}^2_{ML}} (Z'Z)^{-1}Z'\hat{E}\hat{E}'Z(Z'Z)^{-1} - (Z'Z)^{-1}, \quad (2.58)$$

Donde

$$\hat{E}\hat{E}' = \sum_{i=1}^{N} (y_i - Z\hat{\beta}_{GLS})(y_i - Z\hat{\beta}_{GLS})',$$

es la matriz de las sumas de productos cruzados, y$\hat{\beta}_{GLS} = \hat{\beta}_{OLS}$ como está definido en (2.52)

Demostración. Denote $e_i = y_i - Z\hat{\beta}_{GLS}$, el vector residual nx1 del i-ésimo cluster, y $\hat{E} = [e_1, \dots, e_N]$, la matriz nxN, tal que $\hat{E}\hat{E}' = \sum_{i=1}^{N} e_i e_i'$. Note que e_i no depende de D. Primero, expresamos σ^2, dado por (2.37), a través de la matriz D usando la fórmula de reducción de dimensión (2.19). Obtenemos

$$\sigma^2 = \frac{1}{Nn} \sum_{i=1}^{N} e_i'V^{-1}e_i = \frac{1}{Nn} tr\left(\sum_{i=1}^{N} e_i e_i'V^{-1}\right) = \frac{1}{Nn} tr\left(\hat{E}\hat{E}'V^{-1}\right)$$

$$= \frac{1}{Nn} tr\left(\hat{E}\hat{E}'[I - Z(D^{-1} + Z'Z)^{-1}Z']\right)$$

$$= \frac{1}{Nn} tr\left(\hat{E}\hat{E}'\right) - \frac{1}{Nn} tr\left(\hat{E}\hat{E}'Z(D^{-1} + Z'Z)^{-1}Z'\right)$$

$$= \frac{1}{Nn} tr\left(\hat{E}\hat{E}'\right) - \frac{1}{Nn} tr\left(\hat{E}\hat{E}'ZD(D + Z'Z)^{-1}T\right), \quad (2.59)$$

Donde $T = (Z'Z)^{-1}Z'$. Segundo, trabajamos en la matriz D. El MLE para D satisface la ecuación de resultados, la derivada parcial de la función log-verosimilitud con respecto a la matriz D. En un caso general la derivada es provista por la fórmula (2.101). En el caso balanceado la ecuación de resultados para D se simplifica a

$$NZ'V^{-1}Z - \sigma^{-2}Z'V^{-1}\hat{E}\hat{E}'V^{-1}Z = 0. \quad (2.60)$$

Nosotros tenemos

$$Z'V^{-1} = Z'[I - Z(D^{-1} + Z'Z)^{-1}Z']$$

$$= Z' - Z'Z(D^{-1} + Z'Z)^{-1}Z'$$

$$= [I - Z'Z(D^{-1} + Z'Z)^{-1}]Z'$$

$$= [(D^{-1} + Z'Z)(D^{-1} + Z'Z)^{-1} - Z'Z(D^{-1} + Z'Z)^{-1}]Z'$$

$$= [(D^{-1} + Z'Z) - Z'Z](D^{-1} + Z'Z)^{-1}Z'$$

$$= D^{-1}(D^{-1} + Z'Z)^{-1}Z'$$

$$= (D^{-1}D + Z'ZD)^{-1}Z'$$

$$= (I + Z'ZD)^{-1}Z'$$

$$= (Z'Z(Z'Z)^{-1} + Z'ZD)^{-1}Z'$$

$$= ((Z'Z)^{-1} + D)^{-1}(Z'Z)^{-1}Z'$$

$$= ((Z'Z)^{-1} + D)^{-1}(Z'Z)^{-1}Z' = M^{-1}T,$$

Donde $M = (Z'Z)^{-1} + D$. Luego, usando la fórmula (2.23) re-escribimos (2.60) como $NM^{-1} - \sigma^{-2}M^{-1}TRT'M^{-1} = 0$, lo cual da la ecuación para D,

$$D = \frac{1}{N\sigma^2}TRT' - (Z'Z)^{-1}, \quad (2.61)$$

Donde $R = \hat{E}\hat{E}'$. Sustituyendo esta fórmula en (2.59), uno obtiene

$$\sigma^2 = \frac{1}{Nn}tr(\hat{E}\hat{E}') - \frac{1}{Nn}tr\left(\hat{E}\hat{E}'Z\left(\frac{1}{N\sigma^2}TRT' - (Z'Z)^{-1}\right)xN\sigma^2(TRT')^{-1}T\right)$$

$$= \frac{1}{Nn}tr(\hat{E}\hat{E}'(I - P_z)) - \frac{\sigma^2}{n}tr\left(\hat{E}\hat{E}'T'\left(T\hat{E}T'\right)^{-1}T\right),$$

Donde $P_z = Z(Z'Z)^{-1}Z'$ es una matriz de proyección. Pero para el tercer término tenemos

$tr\left(\hat{E}\hat{E}'T'\left(T\hat{E}T'\right)^{-1}T\right) = tr\left(\left(T\hat{E}\hat{E}'T'\right)^{-1}T\hat{E}\hat{E}'T'\right) = m$, y de aquí resolvemos para σ^2, finalmente obtenemos el MLE

$$\hat{\sigma}^2_{ML} = \frac{1}{N(n-m)}tr(\hat{E}(I - P_z))$$

$$= \frac{1}{N(n-m)}\sum_{i=1}^{N}(y_i - Z\hat{\beta}_{GLS})'(I - P_z)(y_i - Z\hat{\beta}_{GLS})$$

Ya que $(I - P_z)Z = 0$, la fórmula precedente puede ser re-escrita como (2.56). Remplazando $\sigma^2\ con\ \hat{\sigma}^2_{ML}$ en (2.61), llegamos al MLE

$$\widehat{D}_{ML} = \frac{1}{N\hat{\sigma}^2_{ML}}(Z'Z)^{-1}Z'\hat{E}\hat{E}'Z(Z'Z)^{-1} - (Z'Z)^{-1}.$$

De nuevo, usando el hecho que $(I - P_z)Z = 0$, la fórmula presente es equivalente a (2.57).

Para la máxima verosimilitud restringida la expresión para σ^2 y la ecuación de resultado para D llega a ser (2.49) y (2.129) respectivamente. De aquí, la ecuación (2.60) transforma a

$(N-1)Z'V^{-1}Z - \sigma^{-2}Z'V^{-1}\hat{E}\hat{E}'V^{-1}Z = 0$, y consecuentemente (2.61) cambia a

$$D = \frac{1}{(N-1)\sigma^2} T\hat{E}\hat{E}'T' - (Z'Z)^{-1}. \quad (2.62)$$

Luego, análogo a (2.59), para el estimador RML,

$$\sigma^2 = \frac{1}{Nn-m} tr\left(\hat{E}\hat{E}'\right) - \frac{1}{Nn-m} tr\left(\hat{E}\hat{E}'ZT + (N-1)\sigma^2 \hat{E}\hat{E}'T'\left(T\hat{E}\hat{E}'T'\right)^{-1}T\right)$$

$$= \frac{1}{Nn-m} tr\left(\hat{E}\hat{E}'(I-P_z)\right) - \frac{(N-1)\sigma^2}{Nn-m} tr\left(\hat{E}\hat{E}'T'\left(T\hat{E}\hat{E}'T'\right)^{-1}T\right)$$

$$= \frac{1}{Nn-m} tr\left(\hat{E}\hat{E}'(I-P_z)\right) - \frac{(N-1)\sigma^2}{Nn-m} \text{m.}$$

Resolviendo para σ^2, llegamos al mismo MLE, (2.56). El estimador (2.58) sigue de (2.62).

Desde el acuerdo de nuestra convención, D denota la matriz de covarianza escalada, el estimador para la matriz de covarianza de efectos aleatorios es $\hat{\sigma}^2_{ML}\widehat{D}$ donde $\widehat{D}$ es el estimador ML o RML (preferimos RML porque es insesgado).

$\hat{\sigma}^2_{ML}$ es positivo si al menos un y_i no es una combinación lineal de vectores columnas de la matriz Z (aprenderemos más tarde en la sección 2.5 que esto es cuna condición suficiente y necesaria para la existencia del MLE). Sin embargo, no se garantiza que $\widehat{D}_{ML}$ o $\widehat{D}_{RML}$ sean matrices definidas positivas (hay un ligero mejor chance para $\widehat{D}_{RML}$ que sea definida positiva). Si las matrices no son definidas no-negativas, uno podría proyectar la matriz en el espacio de las matrices definidas no-negativas, como se ha definido en la sección 2.15. Note que para datos balanceados, $\widehat{D}_{ML} < \widehat{D}_{RML}$. Esto tiene perfecto sentido porque $\widehat{D}_{ML}$ es negativamente sesgada y $\widehat{D}_{RML}$ es insesgada.

Modelos lineal, logístico de 3 parámetros, logístico de 4 parámetros y gompertz.

```
P=ChickWeight

pollito=subset(P,Chick=="21")

plot(weight~Time,pollito)

m1=lm(weight~Time,pollito)

abline(m1,col=1)

m3=nls(weight~SSlogis(Time,a,b,c),pollito)

curve(predict(m3,data.frame(Time=x)),add=TRUE,col=3)

m4=nls(weight~SSfpl(Time,a,b,c,d),pollito)

curve(predict(m4,data.frame(Time=x)),add=TRUE,col=4)

m5=nls(weight~SSgompertz(Time,a,b,c),pollito)

curve(predict(m5,data.frame(Time=x)),add=TRUE,col=5)
```

```
AIC(m1,m3,m4,m5)

summary(m1)

summary(m4)

plot(resid(m1))

plot(resid(m4))
```

de las fórmulas existente e este y otros documentos relacionados.