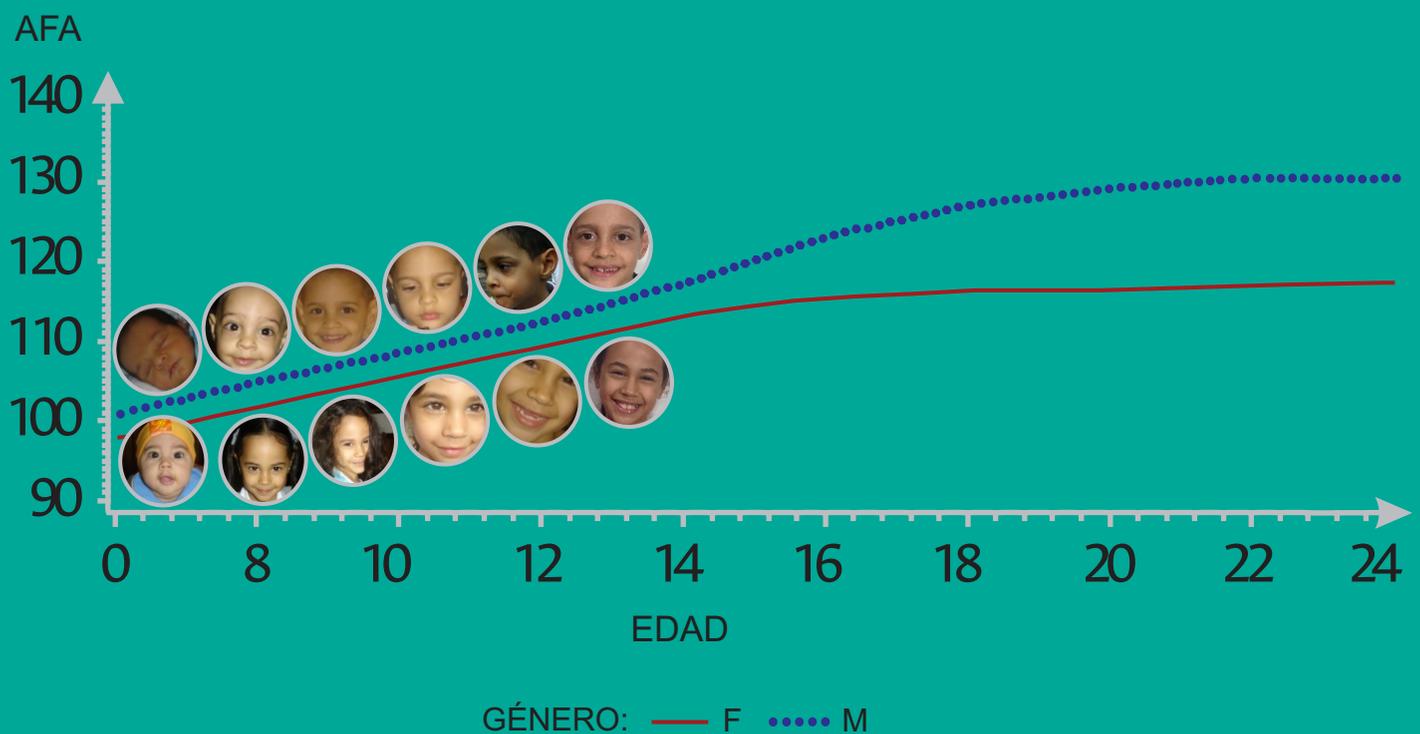
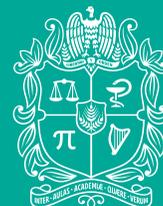


# INTRODUCCIÓN A LOS MODELOS MIXTOS

Juan Carlos Correa Morales  
Juan Carlos Salazar Uribe



Escuela de Estadística  
Facultad de Ciencias  
Sede Medellín



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA



# **Introducción a los modelos mixtos**

**Juan Carlos Correa Morales**

Profesor Asociado Escuela de Estadística, Facultad de Ciencias  
Universidad Nacional de Colombia, Sede Medellín

**Juan Carlos Salazar Uribe**

Profesor Asociado Escuela de Estadística, Facultad de Ciencias  
Universidad Nacional de Colombia, Sede Medellín

# Introducción a los modelos mixtos

**Juan Carlos Correa Morales**

Profesor Asociado Escuela de Estadística, Facultad de Ciencias  
Universidad Nacional de Colombia, Sede Medellín

**Juan Carlos Salazar Uribe**

Profesor Asociado Escuela de Estadística, Facultad de Ciencias  
Universidad Nacional de Colombia, Sede Medellín



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

Medellín, Colombia, 2016

519.53

C67 Correa Morales, Juan Carlos

Introducción a los modelos mixtos / Juan Carlos Corra Morales, Juan Carlos Salazar Uribe. – Medellín: Universidad Nacional de Colombia. Facultad de Ciencias. Escuela de Estadística, 2016.

257 páginas: ilustraciones

ISBN: 978-958-775-953-2

1. MODELOS LINEALES (ESTADÍSTICA). 2. MODELOS LOG-LINEALES. 3. ANÁLISIS MULTIVARIANTE. 4. ANÁLISIS DE VARIANZA. I. Salazar Uribe, Juan Carlos. II. Título

Catalogación en la publicación Universidad Nacional de Colombia, Sede Medellín

Introducción a los modelos mixtos

© Universidad Nacional de Colombia Sede Medellín.

Centro Editorial

Facultad de Ciencias

© Juan Carlos Correa Morales, Juan Carlos Salazar Uribe

ISBN 978-958-775-953-2

Medellín, Colombia

Primera edición: noviembre 2016

Medellín, Colombia, 2016

Diseño de portada: María Aristizábal

Gráfico portada: cortesía Doctor Iván Darío Jiménez Vargas, Universidad CES y grupo CES-LPH. Los datos corresponden a 24 años de seguimiento de crecimiento facial. Único estudio en su género en latino América.

Prohibida la reproducción total o parcial por cualquier medio sin la autorización escrita del titular de los derechos patrimoniales.

Impreso y hecho en Medellín, Colombia.

# Prefacio

Este libro presenta una introducción a los modelos mixtos y puede usarse en semestres avanzados a nivel de pregrado en estadística o como parte de los cursos de posgrado que se relacionen con estadística. El campo de los modelos mixtos ha tenido un gran desarrollo a nivel mundial en los últimos años, en especial desde la creación e introducción de rutinas computacionales intensivas, tal y como se evidencia en la abundante literatura sobre el tema. En nuestro medio ya se percibe el uso de estos modelos no solo en algunos trabajos de investigación sino en el quehacer de empresas y centros de investigación del medio externo.

Muchas ideas han estado circulando desde hace mucho tiempo, pero la imposibilidad de llevarlas a la práctica hacían que se miraran con cierta desilusión, ya que eran muy atractivas pero inaplicables por los retos computacionales que representaban. Esto afortunadamente ha cambiado. En nuestro medio, ya se encuentran trabajos publicados relacionados con el tema y poco a poco los modelos lineales mixtos se han posicionado como una poderosa herramienta de análisis estadístico. Sin embargo, es lamentable que la mayoría de libros básicos en estadística no hagan una presentación de los elementos básicos de esta aproximación basada en modelos para la solución de problemas estadísticos.

El principal objetivo de este trabajo es familiarizar al lector con conceptos básicos del modelo lineal mixto y mixto generalizado, así como presentar algunos elementos relacionados con modelos mixtos no lineales. Con esto se pretende que el lector pueda implementar modelos de este tipo a situaciones prácticas propias de su quehacer cotidiano o como apoyo en la solución de problemas de investigación en diversas áreas del conocimiento. En este trabajo se va a presentar y desarrollar una aproximación teórico-práctica, esto es, el lector puede aplicar de forma casi inmediata los métodos a problemas reales. Sin embargo, se presentan algunos desarrollos teóricos de resultados estándar de la teoría de modelos mixtos lineales y no lineales. El software que se utilizará en la mayoría de los ejemplos es el  $R^1$ , que es de permiso libre para educación. También se usará software con licencia como por ejemplo el SAS©<sup>2</sup>. Se requiere familiaridad con estos programas al menos a un nivel

---

<sup>1</sup><http://www.r-project.org/>

<sup>2</sup>[www.sas.com](http://www.sas.com)

operativo básico.

Se asume que el lector también está familiarizado con métodos estadísticos, especialmente aquellos relacionados con modelos lineales y modelos lineales generalizados. Conocimiento de álgebra lineal e inferencia a un nivel de un texto básico de estadística matemática puede ser de gran ayuda.

# Contenido

<b>1. Introducción</b>	<b>10</b>
1.1. Datos longitudinales . . . . .	11
1.2. Efectos fijos y efectos aleatorios . . . . .	12
1.3. Tareas para el analista . . . . .	13
1.4. Ejemplos de motivación . . . . .	19
1.4.1. Electroplateado . . . . .	19
1.4.2. Hueso ramus . . . . .	22
1.4.3. Secado de madera . . . . .	25
1.4.4. Toros . . . . .	26
1.4.5. Conteo de pasos en humanos . . . . .	26
1.4.6. Duración de pilas alcalinas . . . . .	26
1.4.7. 100 metros planos Olímpicos de Londres 2012 . . . . .	27
<b>2. ¿Qué es un modelo mixto?</b>	<b>29</b>
2.1. Análisis estadístico: estadísticos de resumen . . . . .	31
2.2. Modelo de dos etapas . . . . .	32
2.2.1. Etapa 1 (formulación general) . . . . .	32
2.2.2. Etapa 2 . . . . .	32
2.2.3. Ejemplo: Abetos . . . . .	33
2.3. Forma matricial del MLM . . . . .	35
2.4. Ejemplo: Abetos (Continuación) . . . . .	36

2.5. Modelo de Laird-Ware . . . . .	37
2.5.1. Modelo de interceptos aleatorios . . . . .	38
2.5.2. Formulaciones especiales del MLM . . . . .	39
2.5.3. Modelo de interceptos y tendencias aleatorias . . . . .	47
2.6. Modelo de regresión con errores autocorrelacionados . . . . .	48
<b>3. Inferencias en el Modelo Lineal Mixto</b>	<b>51</b>
3.1. MLM: Modelos jerárquicos versus marginales . . . . .	51
3.2. Estimación . . . . .	52
3.2.1. Modelo clásico . . . . .	52
3.2.2. Estimación basada en la verosimilitud perfilada . . . . .	56
3.2.3. Estimación por máxima verosimilitud (ML) . . . . .	57
3.2.4. Máxima verosimilitud restringida (REML) . . . . .	58
3.3. Métodos para ajustar MLM a los datos . . . . .	63
3.3.1. Algoritmo EM . . . . .	63
3.3.2. Algoritmo de Newton-Raphson . . . . .	65
3.3.3. Estimación por mínimos cuadrados generalizados . . . . .	66
3.3.4. ¿Y si la estructura de $\mathbf{R}_i$ es más compleja que $\sigma^2\mathbf{I}$ ? . . . . .	68
3.4. Criterios de información . . . . .	69
3.4.1. El criterio de información de Akaike (AIC) . . . . .	69
3.4.2. El criterio de información corregido de Akaike (AICC) . . . . .	70
3.4.3. El criterio de información bayesiano de Schwarz (BIC) . . . . .	70
3.5. Ejemplos . . . . .	70
3.6. Aproximación bayesiana . . . . .	84
3.6.1. Método Bayes empírico . . . . .	84
3.6.2. Estimadores shrinkage . . . . .	86
3.6.3. Chequeo de normalidad de los efectos aleatorios . . . . .	86
3.6.4. Consideraciones de inferencia . . . . .	87
3.7. Pruebas de hipótesis . . . . .	100
3.7.1. $\mathbf{V}$ conocida . . . . .	101
3.7.2. $\mathbf{V}$ desconocida . . . . .	101
3.7.3. $VarCov(\mathbf{Y}_i)$ desconocida . . . . .	104

3.8. Inferencia para la media . . . . .	104
3.8.1. Otra forma de obtener estos resultados . . . . .	106
3.9. Pruebas de hipótesis e intervalos de confianza . . . . .	106
3.9.1. Método 1: Test aproximado de Wald . . . . .	106
3.9.2. Método 2: Test aproximados a la distribución F . . . . .	107
3.10. Inferencia para las componentes de varianza . . . . .	111
3.10.1. Test de Wald aproximados . . . . .	112
3.10.2. Tests de razón de verosimilitud (LRT) . . . . .	112
3.11. Pruebas marginales para la necesidad de efectos aleatorios . . . . .	113
3.12. Modelo autorregresivo de primer orden . . . . .	114
3.13. Pruebas de hipótesis sobre las varianzas de los efectos aleatorios	116
3.13.1. Prueba de la razón de la verosimilitud restringida . . . . .	117
3.14. Recomendación para la construcción del modelo lineal mixto . . . . .	119
3.15. Medidas de bondad de ajuste . . . . .	120
3.15.1. Análisis de residuales . . . . .	120
3.15.2. Estructuras de covarianza en el MLM . . . . .	122
3.16. Predicción . . . . .	128
3.16.1. Mejor predictor lineal insesgado . . . . .	128
<b>4. Ejemplos</b>	<b>132</b>
4.1. Electroplateado . . . . .	132
4.2. Toros . . . . .	140
4.3. Hueso ramus . . . . .	146
4.4. Puntajes . . . . .	159
<b>5. Diagnósticos en el modelo mixto</b>	<b>162</b>
5.1. Diagnósticos vía eliminación de observaciones . . . . .	162
5.1.1. Influencia en los estimadores de los efectos fijos . . . . .	164
5.1.2. Influencia en los efectos aleatorios . . . . .	166
5.1.3. Influencia sobre la función de verosimilitud . . . . .	167
5.2. Verificación de la normalidad de los efectos aleatorios . . . . .	168

<b>6. Modelo lineal generalizado mixto</b>	<b>170</b>
6.1. Estimación	170
6.1.1. Algoritmo para la estimación	172
6.2. Otras formas de estimación usando métodos Monte Carlo	174
6.2.1. Algoritmo EM Monte Carlo (MCEM)	174
6.2.2. Algoritmo Monte Carlo Newton-Raphson (MCNR)	177
6.2.3. Máxima verosimilitud simulada (SML)	179
6.3. Ejemplos	179
6.3.1. Genética porcina (respuesta dicótoma)	179
6.3.2. Modelo lineal mixto tipo Poisson	187
<b>7. Modelo no lineal mixto</b>	<b>190</b>
7.1. El modelo no lineal tradicional	190
7.2. Modelos mixtos no lineales	191
7.2.1. Modelos no lineales que son linealizables	191
7.2.2. Parámetros condicionalmente linealizables	193
7.3. Extensión al caso de efectos aleatorios	194
7.4. Estimación de $\beta$ y $\mathbf{b}$	196
7.4.1. Vía máxima verosimilitud	198
7.4.2. Máxima verosimilitud restringida	199
7.5. Algoritmo de dos pasos	200
7.6. Ejemplo	202
<b>8. Ejercicios Propuestos</b>	<b>207</b>
8.1. Tratamiento con Enalapril	207
8.2. Resultados de tratamientos en clínicas	208
8.3. Duración de pilas alcalinas	209
<b>A. Algunas bases de datos</b>	<b>210</b>
A.1. Electroplateado	210
A.2. Hueso ramus	211
A.3. Toros	211
A.4. Secado de madera	212
A.5. Conteo de pasos en humanos	214

<b>B. Resultados de álgebra Lineal</b>	<b>217</b>
<b>C. Repaso de modelos lineales generalizados</b>	<b>229</b>
C.1. Familia exponencial . . . . .	229
C.2. Modelos lineal general y lineal generalizado . . . . .	231
3.3. Funciones de vínculo Logit . . . . .	233
3.4. Funciones de vínculo Log . . . . .	234
3.5. Ajuste del modelo lineal generalizado . . . . .	234
<b>4. Software para modelos mixtos</b>	<b>237</b>
4.1. Software R . . . . .	237
4.2. Función <code>lme()</code> . . . . .	237
4.3. Función <code>glmmML</code> . . . . .	238
4.4. Función <code>glmm()</code> . . . . .	238
4.5. Función <code>lmer</code> . . . . .	239
4.6. Otros paquetes para modelos lineales mixtos . . . . .	241
<b>5. Referencias</b>	<b>242</b>

# Capítulo 1

## Introducción

El modelo lineal ha sido una de las herramientas estadísticas más importantes para el usuario a nivel aplicado. Su interpretabilidad y la disponibilidad de resultados computacionales, junto con los resultados teóricos, han hecho de este modelo una técnica muy popular. Para obtener estimaciones con este modelo, uno de los requerimientos fundamentales es el de la independencia de las observaciones muestrales.

Datos en forma de “cluster” (agrupados y relacionados) aparecen con mucha frecuencia en diversas áreas y el estadístico aplicado debe disponer de métodos para analizarlos de manera adecuada. En investigaciones es corriente tomar varias mediciones (de la misma variable o de una respuesta específica) sobre el mismo sujeto. Puede ser que el interés sea estudiar la evolución de un fenómeno en el tiempo, lo que se conoce como datos longitudinales, o que en el sujeto se realicen varias mediciones.

Esto puede deberse básicamente a dos causas distintas. Por ejemplo, en un experimento, una capa de pintura es aplicada a un pedazo de lámina y el grosor de la pintura puede medirse en varios puntos tomados al azar, esto se conoce como submuestreo, o en un caso típico de investigación social la unidad experimental está formada por un grupo de subunidades, por ejemplo una familia con los diferentes hijos; la boca de una persona con sus diferentes dientes, etc. Cada sujeto o unidad experimental va a generar varios registros en la base de datos, uno por cada subindividuo. Esto puede generalizarse más, por ejemplo, una familia, sujetos dentro de la familia, dientes dentro de cada sujeto de la familia.

Este problema lleva al surgimiento de una clase de modelos conocidos como

modelos lineales mixtos, que cubija los modelos lineales tradicionales, pero permite la inclusión de estas subunidades dentro del análisis. El estudio e implementación de estos modelos es el tema principal de este trabajo y en algunas ocasiones se expondrá el material a un nivel elemental y en otras a un nivel más complejo. Se usará principalmente el programa estadístico *R* y las librerías que dispone para ajustar estos modelos, ya que es un programa de uso gratuito y ampliamente reconocido. Existen muchos paquetes estadísticos que trabajan estos modelos, por ejemplo SAS, SPSS, STATA entre otros. El SAS también se usará para ilustrar algunas de las técnicas estudiadas.

Por datos “agrupados” (clustered data) se entiende una colección de observaciones que tienen una asociación estocástica, temporal, espacial u otra que sugiere que deben ser tratados como un grupo.

Las dos situaciones básicas que producen estructuras de datos agrupados son:

1. Procesos aleatorios jerárquicos
  - a) Submuestreo (unidades dentro del grupo no son correlacionadas)
  - b) Splitting (unidades dentro del grupo no son correlacionadas)
  - c) Selección aleatorias de tratamientos o bloques
2. Medidas repetidas o datos longitudinales (las unidades dentro del grupo son correlacionadas)
  - a) en el tiempo
  - b) en el espacio
  - c) en el tiempo y en el espacio.

## 1.1. Datos longitudinales

En los datos longitudinales la respuesta es observada varias veces en diferentes tiempos para cada sujeto o unidad involucrada en el estudio. Los sujetos no se observan necesariamente en los mismos tiempos ni se observan necesariamente el mismo número de veces (datos desbalanceados). Las covariables pueden ser diferentes para cada sujeto en los distintos tiempos. Se puede pensar de este tipo de datos que:

- cada individuo tiene una tendencia propia o “trayectoria”. No son necesariamente líneas rectas.

- Si se promedian todas las observaciones se obtiene un perfil promedio poblacional.

El modelo lineal mixto es flexible ya que permite obtener un perfil individual en adición al perfil promedio como se verá más adelante en algunos capítulos del presente trabajo. Para este tipo de datos se habla de “modelos específicos de sujetos” o “modelos de población promedios” y dependiendo de los objetivos de investigación uno puede estar interesado en uno o en el otro.

## 1.2. Efectos fijos y efectos aleatorios

En diseño experimental ha sido clara la diferencia entre “efectos fijos” y “efectos aleatorios”. En muchas situaciones es relativamente simple decidir si un efecto es fijo o no. En algunas situaciones no es claro y los investigadores se deciden por uno o por otro con muy pocas o ninguna explicación. Por ejemplo, suponiendo un estudio epidemiológico la variable *Fumar* puede ser considerada efecto fijo, pero algunos investigadores argumentan que esta es una variable continua dicotomizada. Una respuesta “Sí fuma” puede reflejar un gran número de situaciones.

Decidir entre si un factor es fijo o aleatorio puede llegar a ser difícil y controversial (Littell et al., 2006). Robinson (2008, pp. 99) comenta “Los modeladores pueden estar en desacuerdo si los efectos deben ser fijos o aleatorios, y el mismo efecto puede cambiar dependiendo de las circunstancias. Ciertamente, los estadísticos no están de acuerdo en una estrategia. Algunos dicen que depende completamente de la inferencia, y algunos que depende completamente del diseño.”

**Efectos aleatorios** Efectos aleatorios son aquellos que supuestamente son tomados al azar de una población de posibles niveles. Para ilustrar esto, considere el caso de un experimento para comparar los tiempos para obtener el secado de una pintura sobre una superficie de madera. La pintura puede provenir de varios productores, por ejemplo productor A o productor B. Si se tiene información sobre la temperatura ambiente, sobre el porcentaje de humedad, estas variables entrarían al modelo como efectos aleatorios y las marcas serían efectos fijos. Pero si el interés es modelar el tiempo de secado dada la temperatura, el productor y la humedad relativa serían los efectos aleatorios y la temperatura el efecto fijo.

Otra manera de ver un efecto aleatorio es sencillamente si no estamos interesados en obtener una estimación de su localización como objetivo principal. Una forma diferente de pensarlos es que un efecto es aleatorio si uno lo desea marginalizar, por cualquier razón.

**Efectos fijos** Estos efectos se asumen que son determinados a propósito, y que no representan a otros sino a ellos mismos.

Es importante entender que los papeles de los efectos fijos y los efectos aleatorios son diferentes: los efectos fijos *explican* variabilidad y los efectos aleatorios *organizan* variabilidad (Robinson, 2008).

Una lista para decidir si usar efectos fijos o aleatorios:

1. ¿Cuál es el número de niveles?

Pequeño	Fijo
Grande o infinito	Posiblemente aleatorio

2. ¿Son los niveles repetibles?

Sí	Fijo
No	Aleatorio

3. ¿Hay, conceptualmente, un número infinito de niveles?

No	Posiblemente fijo
Sí	Posiblemente aleatorio

4. ¿Se necesitan realizar inferencias para niveles no incluidos en el muestreo?

No	Posiblemente fijo
Sí	Posiblemente aleatorio

### 1.3. Tareas para el analista

Watts (1981) presenta una lista de tareas que debe realizar un analista de datos que esté tratando de ajustar un modelo de regresión que pueden considerarse como obligatorias para la buena práctica estadística.

Muchas de estas tareas deben realizarse también cuando se ajustan modelos mixtos (lineales o no lineales).

**1. Entender el problema y la situación experimental u observacional.**

- a) Lea reportes y estudie la situación. Luego escriba una descripción del problema, enfatizando porqué el problema es importante.
- b) Discuta esta descripción con el cliente para asegurarse que ud. ha entendido apropiadamente la situación.
- c) Distinga cuidadosamente las variables involucradas como respuestas y predictoras.
- d) Anote cualquier característica especial de los datos o de la situación que pueda ser útil en la selección de las transformaciones de cualquier variable o en aclarar la naturaleza de las relaciones.
- e) Visite el lugar de experimentación o examine cómo fueron recolectados los datos de forma cuidadosa para aclarar cualquier punto y entender mejor la situación. Haga preguntas (preferiblemente informativas y no cerradas a menos que sea necesario formularlas).
- f) Determine cuáles son los puntos importantes. El principal objetivo es
  - 1) ¿estimar parámetros?
  - 2) ¿probar si un subconjunto de parámetros es cero?
  - 3) ¿desarrollar una relación descriptiva?
  - 4) ¿predecir observaciones futuras?
  - 5) ¿tomar decisiones?

**2. Decida si el análisis de regresión es apropiado.**

- a) Identifique la respuesta legítima y las variables predictoras.
- b) Decida si el análisis de regresión es apropiado o si hay otra técnica multivariable que lo sea.
- c) Si hay más de una variable dependiente decida si la regresión multirespuesta es apropiada.

**3. Lleve a cabo un análisis preliminar.**

- a) Recuerde los supuestos involucrados en la regresión de mínimos cuadrados y relaciónelos con la situación práctica con miras a determinar si ellos son apropiados para realizar el análisis preliminar.

- b) Determine si hay errores en las variables.
- c) Si hay réplicas use gráficos normales o boxplots para verificar visualmente la normalidad de los errores y para ver si la varianza es constante.
- d) Si hay réplicas, calcule  $\bar{y}$ ,  $s^2$  y  $s$  para cada conjunto de réplicas y gráfíquelas para ver si la varianza es constante.
- e) Si no hay réplicas, un ajuste simple preliminar puede hacerse y calcular los residuales y chequear si la varianza es constante gráfícalos contra  $\hat{y}$  o contra  $x$ .

4. **Transforme para estabilizar varianza o para simplificar el modelo.**

- a) Examinando la naturaleza de la respuesta y “recordando” el comportamiento de la varianza en algunas situaciones comunes, sugieren una transformación lógica para estabilizar la varianza.
- b) ¿Existe alguna restricción en  $Y$ , y por lo tanto ésta sugiere una transformación apropiada para estabilizar la varianza?
- c) Si hay réplicas, calcule  $\bar{y}$ ,  $s^2$  y  $s$  para cada conjunto de réplicas y gráfíquelas o construya un boxplot para ver si la varianza es constante y cuál transformación pueda ayudar a estabilizarla.
- d) Examine los boxplots para simetría y para valores separados y gráfícalos en diferentes escalas, seleccione una transformación apropiada para estabilizar la varianza para cada conjunto y luego una transformación que sea buena.
- e) Grafique  $s^2$  vs.  $\bar{Y}$  y  $s$  vs.  $\bar{Y}$  en varias escalas para obtener relaciones simples. Luego utilice la teoría de transformaciones para estabilizar varianza para seleccionar una buena transformación o retome algunas combinaciones.
- f) Si es posible, proporcione una justificación para la selección de una transformación en términos del comportamiento de la respuesta, o dé razones físicas, biológicas u otra para su selección.
- g) Si no hay un comportamiento consistente de  $s^2$  o  $s$  vs.  $\bar{Y}$ , o si la transformación para la estabilización de la varianza causa más complejidad en el modelo resultante que los beneficios obtenidos por la estabilización de la varianza, el analista puede decidirse a usar mínimos cuadrados ponderados. Si es así, los pesos o la forma funcional de los pesos deben determinarse.

**5. Seleccione una forma apropiada del modelo.**

- a) Determine cuáles variables predictoras son posiblemente las más influyentes y, si es posible, qué forma tendrán. Por ejemplo, ¿será  $Y$  aproximadamente lineal en  $x$ ? ¿Deberá  $Y$  incrementarse cuando  $x$  disminuye (consideraciones teóricas)? ¿Será posible que existan interacciones?
- b) ¿Existen características inusuales en  $Y$ , por ejemplo,  $Y$  debe ser siempre positivo? Si es así, trate de incluir esta información en el modelo.
- c) ¿Pueden algunas variables causar un comportamiento inusual? por ejemplo, si  $x_1$  y  $x_2$  se vuelven grandes, entonces  $y$  se sale de control. ¿Hay sensibilidad a las interacciones?
- d) ¿Habrà alguna predictora que tiene limitaciones o un comportamiento especial que sugiera una transformación que proporcione un modelo simple, efectivo y sensible?
- e) Ingrese los datos y grafique  $Y$  vs. cada una de las  $x$ , o use marqui-llas o múltiples gráficos para obtener ideas de cómo se comporta y qué términos deben incluirse en el modelo. También use gráficos de contorno bivariantes de  $Y$  vs.  $x$ .
- f) Grafique  $Y$  o las transformaciones de  $Y$  vs. las  $x$ 's transformadas para obtener relaciones simples (sin considerar si la varianza es estable) y relacione estos modelos simples con las variables originales.
- g) Calcule las correlaciones de las  $Y$  y las  $x$ 's y las  $Y$  transformadas y las  $x$ 's transformadas para que posiblemente permita seleccionar las variables importantes y detectar tendencias para colinealidad.
- h) Utilice técnicas computarizadas tales como regresión stepwise, splines o técnicas de automodelación.
- i) Use la información anterior para especificar un modelo tentativo.

**6. Estime los parámetros.**

- a) En una regresión no lineal determine valores razonables para iniciar los parámetros mediante gráficos de los datos (en diferentes escalas), o mediante transformaciones a una forma lineal y corriendo una regresión lineal y luego derivar valores de los parámetros, o calculando y relacionando los parámetros a valores asintóticos, pendientes iniciales, valores de picos, etc.

- b) Seleccione parametrizaciones ventajosas para ayudar al modelo a converger.
- c) Calcule las derivadas con respecto a los parámetros y codifique tanto el modelo como las derivadas para el procesamiento en el computador.
- d) Digite los datos y use algún software de calidad para estimar los parámetros en un modelo lineal o no lineal.

#### 7. Refine el modelo.

- a) Examine los residuales del modelo ajustado graficándolos contra diferentes variables tales como orden,  $\bar{Y}$ ,  $x$ 's, lo que permite detectar problemas del modelo, outliers, varianza no constante, o correlaciones de los residuales.
- b) Haga gráficos normales o medio-normales (half-normal) o boxplots para detectar outliers o si el supuesto de normalidad de los errores es inapropiado.
- c) Si hay outliers, o bien elimine las observaciones luego de haber justificado satisfactoriamente tal acción y re-estime el modelo, o posiblemente cambie la forma de realizar inferencias, digamos una regresión robusta, o incorpore nuevas variables para tener en cuenta este comportamiento.
- d) Calcule las correlaciones de los residuales con otras variables que sean posibles predictores, y con ellos mismos determine si hay autocorrelación en los residuales.
- e) Examine la tabla con los parámetros estimados para determinar si una relación informativa (sustancial) ha sido demostrada. Pruebe el ajuste del modelo. Si hay necesidad, transforme al menos unas de las variables del modelo a fin de acercarse al cumplimiento de los supuestos del modelo.
- f) Examine las correlaciones de los parámetros o los estadísticos  $t$  para decidir si algunas variables pueden eliminarse. Analice los estadísticos de multicolinealidad antes de eliminar cualquier variable, y calcule los tamaños muestrales efectivos.
- g) Resuma la salida del computador y sus hallazgos y decida si acepta o no el modelo ajustado.
- h) Compare los errores medios cuadrados, las razones  $F$ , las correlaciones y las varianzas de los parámetros, y los residuales para

este modelo y los modelos ajustados previamente que le ayuden a decidir si aceptar o no el modelo.

- i)* Si el modelo ha de ser cambiado, sugiera cómo debe cambiarse y porqué. Regrese al tercer paso (si es necesario) o al cuarto paso.
- j)* Si hay fuerte colinealidad y si se puede conseguir información adicional, especialmente de algún diseño experimental, sugiera cómo deben obtenerse los datos para reducir la colinealidad y por lo tanto ser capaces de remover algunas variables o retenerlas y proporcionar mejores estimaciones de los parámetros.

#### 8. **Evalúe el modelo.**

- a)* Use la salida del computador para determinar regiones de confianza conjuntas y marginales de los parámetros o resúmenes de información que sea equivalente y útil.
- b)* Establezca el modelo y presente información sobre la precisión tal como errores estándar o regiones de confianza.
- c)* Calcule los tamaños muestrales efectivos para las variables para indicar la extensión de la colinealidad. Calcule las eficiencias del diseño y explíquelas.
- d)* Calcule los límites de confianza para la función respuesta en puntos seleccionados y proporcione expresiones para los límites con cálculos simples para otros puntos.
- e)* Calcule límites de confianza para observaciones futuras en puntos seleccionados. Proporcione expresiones para los límites con cálculos simples para otros puntos.
- f)* Calcule eficiencias de predictibilidad para el modelo y diseño y para puntos especificados.
- g)* Interprete el modelo y hallazgos en términos de la situación experimental. Si algo no tiene sentido o es sorprendente, examine este aspecto cuidadosamente hasta que pueda decidir que el modelo es correcto y su intuición (o la del investigador) está equivocada, o el modelo es incorrecto y necesita modificarse, o si hay que examinarlo más pero una conclusión solo puede lograrse consiguiendo más datos, en cuyo caso describa qué clase de experimento sería más informativo.

## 9. Escriba un reporte.

- a) Arme y organice la salida del computador, los gráficos y la información de resumen.
- b) Desarrolle un boceto del reporte.
- c) Prepare versiones preliminares del reporte, describiendo las partes importantes del análisis, los hallazgos (incluyendo el modelo ajustado), y cualquier punto especialmente importante tal como cualquier variable explicatoria inesperada, cualquier limitación del diseño experimental o del modelo y sugerencias para futuros experimentos.
- d) Modifique y revise los borradores hasta lograr un reporte satisfactorio, preferiblemente juzgado por un colega o el investigador. Incluya partes importantes de la salida del computador, gráficos y análisis. Las derivaciones se presentan como apéndices. También incluya un resumen de hechos claro. Marque los gráficos y figuras adecuadamente y use buenas referencias de las secciones en el reporte. Si es necesario incluya una lista de referencias.

## 1.4. Ejemplos de motivación

En esta sección se presentan algunos ejemplos típicos que servirán no solo para contextualizar los modelos lineales mixtos y resaltar su potencial en la práctica, sino también para ilustrar las diversas técnicas que se presentarán en este libro. Algunos de estos conjuntos de datos fueron recopilados durante el desarrollo de investigaciones en distintas disciplinas en Colombia.

### 1.4.1. Electroplateado

Este es un ejemplo de datos no longitudinales y medidas repetidas. El electroplateado es un método para galvanizar placas de acero y se usa en una amplia variedad de productos de acero. Un ingeniero está estudiando tres procesos de electroplateado (tratamientos I, II y III) y para esto toma 18 placas uniformes en tamaño y grosor y las somete a cada uno de los procesos. En este ejemplo, la respuesta de interés para el investigador es el grosor (en m.m.) medido en tres regiones de la placa seleccionadas al azar. De esta manera hay tres mediciones para cada una de las 18 placas. Los datos aparecen en el apéndice. A continuación se presenta un análisis exploratorio de

estos datos de electroplateado con el fin de describir algunas características relevantes asociadas al proceso.

Placa	Promedio	Mediana	Desv. estándar
1	0.500	0.490	0.0173
2	0.513	0.510	0.0152
3	0.507	0.510	0.0152
4	0.520	0.520	0.010
5	0.520	0.520	0.020
6	0.517	0.520	0.0055
7	0.527	0.530	0.0152
8	0.547	0.540	0.0114
9	0.533	0.530	0.0152
10	0.527	0.530	0.0055
11	0.537	0.540	0.0055
12	0.523	0.520	0.0152
13	0.567	0.570	0.0055
14	0.567	0.570	0.0152
15	0.573	0.570	0.0055
16	0.567	0.560	0.0114
17	0.567	0.570	0.0152
18	0.557	0.560	0.0152

Tabla 1.1. Estadísticos básicos para cada una de las placas (basados en 3 mediciones)

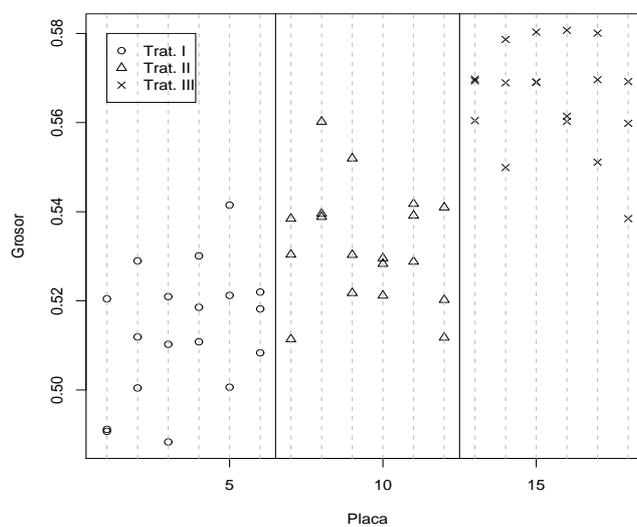
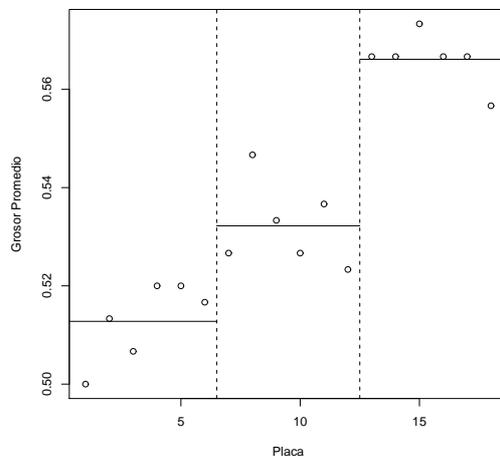
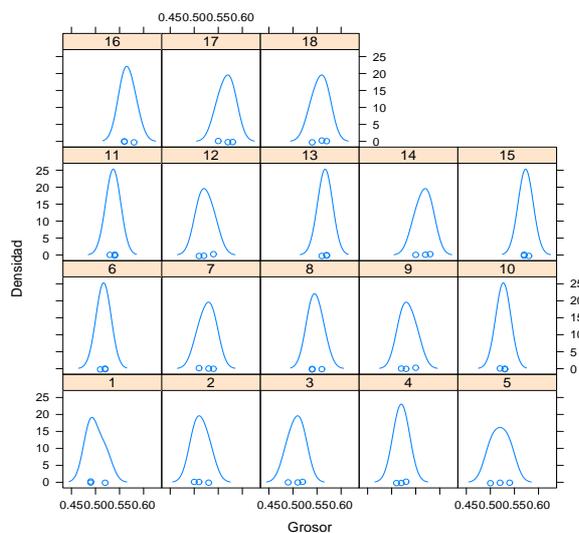


Figura 1.1. Mediciones de las placas sometidas al electroplateado. Las líneas verticales continuas separan las placas por tratamiento.



**Figura 1.2.** Valores medios en cada placa. Las líneas verticales separan las placas por tratamiento y las líneas horizontales muestran el promedio para cada tratamiento.



**Figura 1.3.** Densidades estimadas de las mediciones de las placas sometidas al electroplateado. Se tienen muy pocas observaciones por sujeto y este gráfico realmente es muy poco fiable. La aparente normalidad de las densidades se debe principalmente a la selección del ancho de banda.

### 1.4.2. Hueso ramus

A una cohorte de niños se les registró la altura (en m.m.) del hueso ramus (ubicado en la mandíbula). Estas mediciones se realizaron a los 8, 8 y medio, 9 y 9 y medio años. El objetivo era establecer una curva de crecimiento normal para ser usada por ortodoncistas. Los datos aparecen en el apéndice.

Se presenta un análisis exploratorio de los datos del hueso ramus en el cual se puede considerar que cada sujeto tiene su propio modelo de regresión.

$$Y_{it} = \beta_{i0} + \beta_{i1}Edad_{it} + \epsilon_{it}, \quad t = 1, 2, 3, 4; i = 1, \dots, 20$$

Para los 20 sujetos se construyen 20 regresiones y de este forma se obtiene 20 interceptos, 20 pendientes y 20 estimadores del error. Un problema que se observa, es que las respuestas dentro de cada sujeto no son independientes y ya que ellas se toman en el tiempo, solo dependen del tiempo previo. Esto lleva a trabajar cada individuo con regresiones con autocorrelación. Si se construyen los modelos individuales sin considerar la autocorrelación se obtiene:

Individuo	$\hat{\beta}_0$	$\hat{\beta}_1$
1	38.50	1.18
2	36.25	1.28
3	34.05	1.52
4	33.50	1.42
5	38.95	1.10
6	46.70	0.74
7	33.65	2.24
8	34.95	1.80
9	15.70	4.08
10	29.05	2.04
11	47.50	0.46
12	8.15	4.96
13	36.60	1.92
14	39.95	1.04
15	37.80	1.48
16	36.70	1.76
17	34.90	1.52
18	43.20	1.30
19	34.95	1.44
20	13.90	4.04

**Tabla 1.2.** Datos hueso ramus. Interceptos y pendientes estimadas para cada sujeto asumiendo independencia intra-sujeto.

Considerando la autocorrelación, se obtiene:

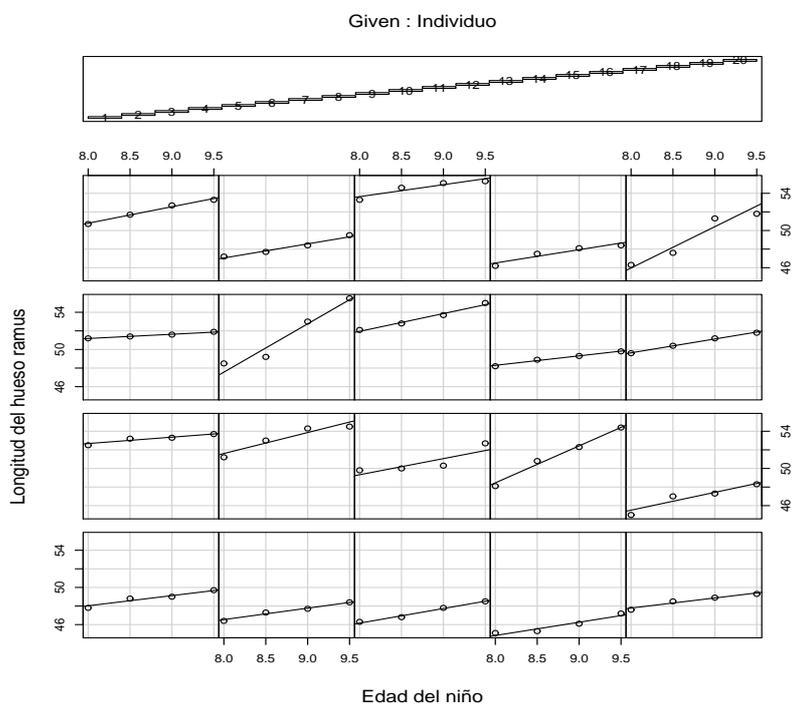
Individuo	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}^2$	$\hat{\rho}$
1	39.26	1.09	0.02	-0.64
2	36.71	1.23	0.01	-0.63
3	33.59	1.57	0.01	-0.63
4	33.40	1.43	0.09	-0.26
5	39.12	1.08	0.03	-0.33
6	47.18	0.69	0.02	-0.57
7	33.56	2.26	0.29	-0.26
8	35.45	1.73	0.58	-0.32
9	16.65	3.98	0.06	-0.57
10	30.21	1.91	0.15	-0.52
11	47.53	0.46	0.00	-0.33
12	6.01	5.19	0.50	-0.52
13	36.63	1.91	0.04	-0.26
14	40.13	1.02	0.01	-0.47
15	37.75	1.49	0.00	-0.33
16	36.60	1.77	0.02	-0.33
17	34.93	1.51	0.04	-0.26
18	43.35	1.29	0.14	-0.27
19	35.07	1.43	0.11	-0.27
20	10.45	4.44	0.22	-0.70

**Tabla 1.3.** Datos hueso ramus. Interceptos, pendientes, varianza y coeficiente de correlación estimados para cada sujeto asumiendo dependencia intra-sujeto.

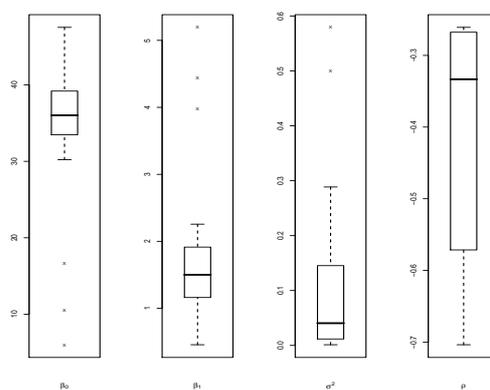
Si se calculan las medianas de los interceptos y de las pendientes (para eliminar el efecto de los posibles outliers) estas son 36.02072 y 1.500743, respectivamente, estos serían los estimadores del modelo poblacional. La diferencia entre los parámetros estimados de un niño y estos valores dan los efectos individuales.

Individuo	efecto intercepto	efecto pendiente
1	3.24	-0.41
2	0.69	-0.27
3	-2.43	0.07
4	-2.62	-0.07
5	3.10	-0.42
6	11.15	-0.81
7	-2.46	0.76
8	-0.57	0.23
9	-19.37	2.47
10	-5.81	0.41
11	11.51	-1.04
12	-30.01	3.69
13	0.61	0.41
14	4.11	-0.48
15	1.73	-0.01
16	0.57	0.27
17	-1.09	0.01
18	7.33	-0.21
19	-0.95	-0.07
20	-25.57	2.94

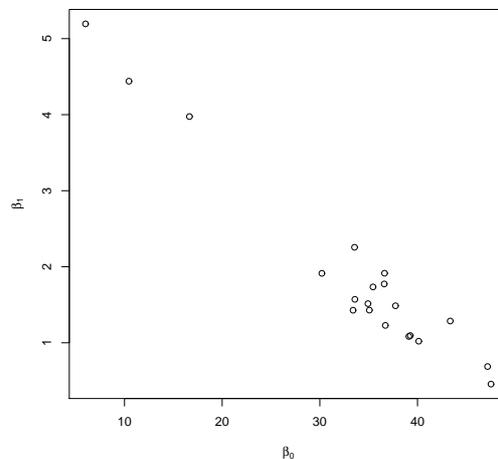
**Tabla 1.4.** Datos hueso ramus. Efectos individuales estimados.



**Figura 1.4.** Crecimiento del hueso ramus en 20 niños. A cada niño se le ajustó un modelo de regresión que considera la correlación entre las mediciones.



**Figura 1.5.** Boxplots de los parámetros estimados para cada niño. Se observan posibles valores extremos.



**Figura 1.6.** Datos hueso ramus. Distribución conjunta de los parámetros estimados para cada niño.

### 1.4.3. Secado de madera

Los siguientes datos corresponden a un experimento relacionado con el secado al aire de unos tablones de ciprés (Botero, 1993). El experimento consistió en poner piezas de madera en apiladas en un patio y protegidos de la lluvia para que fueran perdiendo humedad hasta alcanzar un secado natural. Los datos se encuentran en el apéndice. La madera para este experimento se dimensionó como se muestra en el siguiente cuadro.

Nº de tablones	Ancho (mm)	Espesor (mm)	Largo (m)
20	180	50	2.54
20	180	40	2.54
20	180	30	2.54
20	180	25	2.54
20	180	20	2.54

**Tabla 1.5.** Dimensiones de los tablones en el experimento de secado de madera.

Los espesores seleccionados obedecen a las diferentes especificaciones de uso en la que se emplea la madera de Ciprés. Cada tablón se pesó el primer

día del experimento, luego en los días 7, 14, 21, 28, 35, 42, 49, 56, 62, 70 y 92 de iniciado el experimento. Así, estos datos corresponden al tipo de datos longitudinales. Observe que cada tablón posee mediciones sucesivas con intervalos de medición variables.

#### 1.4.4. Toros

Se seleccionaron 6 toros al azar de varios criaderos y a cada animal se le registró el porcentaje de inseminaciones artificiales exitosas por eyaculación. Los datos aparecen a continuación.

Toro	Porcentaje/100									
1	0.46	0.31	0.37	0.62	0.30					
2	0.70	0.59								
3	0.52	0.44	0.57	0.40	0.67	0.64	0.70			
4	0.47	0.21	0.70	0.46	0.14					
5	0.42	0.64	0.50	0.69	0.77	0.81	0.87			
6	0.35	0.68	0.59	0.38	0.57	0.76	0.57	0.29	0.60	

**Tabla 1.6.** Porcentaje de inseminaciones artificiales exitosas.

#### 1.4.5. Conteo de pasos en humanos

A un grupo de estudiantes de la Universidad Nacional de Colombia-Sede Medellín, se les solicitó caminar una distancia de 50 metros tres veces. En cada recorrido se les registró el número de pasos y la longitud promedio del paso (la marcha del caminado es a paso normal). Adicionalmente, a cada persona se le registraron los valores de algunas variables. Los datos obtenidos se encuentran en el apéndice.

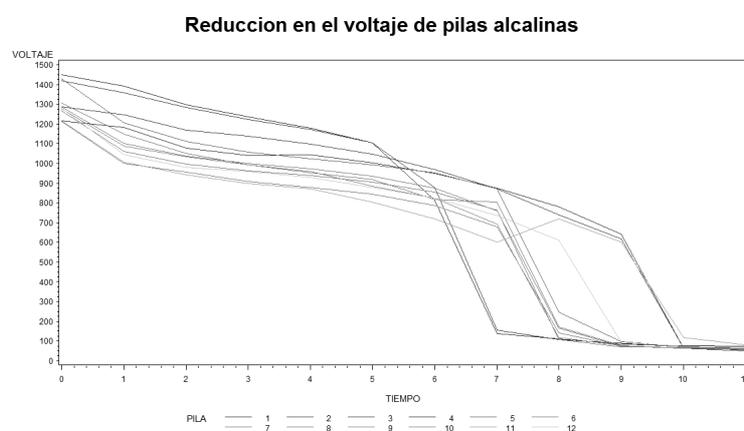
#### 1.4.6. Duración de pilas alcalinas

Se seleccionaron 6 tipos de pilas alcalinas: RAYOVAC<sup>®</sup>, DURACELL<sup>®</sup>, VARTA<sup>®</sup>, EVEREADY ALKALINE<sup>®</sup>, PANASONIC POWER ALKALINE<sup>®</sup> y ENERGIZER MAX<sup>®</sup>. De cada una se tomaron 2 pilas y se conectaron a un sistema con resistencias paralelas que permitió descargar cada una de las 12 pilas. Cada hora y durante 11 horas se registró la disminución en el voltaje (milivoltios). Los datos están en las siguientes tablas:

	RAYOVAC		DURACELL		VARTA	
	VOLTAJE		VOLTAJE		VOLTAJE	
TIEMPO	PILA 1	PILA 2	PILA 1	PILA 2	PILA 1	PILA 2
0	1217	1288	1419	1450	1430	1308
1	1181	1248	1359	1393	1207	1150
2	1077	1170	1284	1296	1112	1052
3	1040	1138	1222	1235	1056	993
4	1043	1099	1173	1178	1025	953
5	1004	1047	1103	1106	992	917
6	949	968	877	814	952	818
7	873	870	156	138	870	805
8	780	740	107	112	246	172
9	641	618	89	84	96	77
10	62	62	71	78	69	65
11	50	58	61	75	64	60

	EVEREADY		PANASONIC		ENERGIZER	
	VOLTAJE		VOLTAJE		VOLTAJE	
TIEMPO	PILA 1	PILA 2	PILA 1	PILA 2	PILA 1	PILA 2
0	1276	1212	1217	1279	1286	1268
1	1086	1001	1005	1044	1102	1060
2	1034	955	943	980	1036	995
3	992	910	897	958	1000	962
4	959	879	871	930	972	940
5	886	844	804	879	934	905
6	825	788	718	823	874	853
7	693	679	602	736	759	763
8	118.8	142.4	718	612	166	109.2
9	75.5	75.1	602	95.5	75.4	70.4
10	64.7	64.3	117.6	73	62.6	66
11	61.2	61.5	81.7	70.3	60.1	65.1

Un gráfico de espagueti o de perfiles de los datos de duración de pilas alcalinas:



**Figura 1.7.** Gráfico de perfiles o espagueti de los datos del experimento de la duración de pilas alcalinas. Alrededor de las 5 horas y media se observa una caída brusca de las curvas asociadas a las distintas baterías.

#### 1.4.7. 100 metros planos Olímpicos de Londres 2012

A continuación se presentan datos de una prueba final de 100 metros planos de 5 hombres y 6 mujeres registrados en los Juegos Olímpicos de Londres

2012 junto con algunos gráficos descriptivos. Se registró el tiempo de cada atleta cada 10 metros. Los datos son:

Sexo	Atleta	t10	t20	t30	t40	t50	t60	t70	t80	t90	t100
H	Tyson Gay	1.90	2.94	3.86	4.73	5.59	6.44	7.28	8.13	8.99	9.85
H	Derrick Atkins	1.89	2.93	3.87	4.75	5.62	6.47	7.32	8.18	9.04	9.91
H	Asafa Powell	1.88	2.91	3.83	4.71	5.57	6.42	7.29	8.16	9.04	9.96
H	Churandy Martina	1.98	3.03	3.98	4.88	5.75	6.62	7.47	8.34	9.20	10.08
H	Marlon Devonish	1.96	3.00	3.95	4.85	5.73	6.61	7.48	8.36	9.23	10.14
M	Veronica Campbell	2.01	3.14	4.15	5.11	6.07	7.02	7.97	8.96	9.97	11.01
M	Lauryn Williams	2.01	3.13	4.13	5.10	6.06	7.03	8.00	8.98	9.99	11.01
M	Torri Edwards	2.00	3.12	4.15	5.12	6.08	7.04	8.01	9.00	10.02	11.05
M	Kim Gevaert	1.98	3.10	4.11	5.10	6.08	7.04	8.02	9.01	10.01	11.05
M	Christine Arron	2.03	3.15	4.16	5.13	6.09	7.06	8.03	9.02	10.04	11.08
M	Veronica Campbell	2.00	3.12	4.12	5.10	6.07	7.02	7.98	8.96	9.93	10.99

Tabla 1.7. Datos de registros de tiempos cada 10 metros. Juegos Olímpicos, Londres 2012.

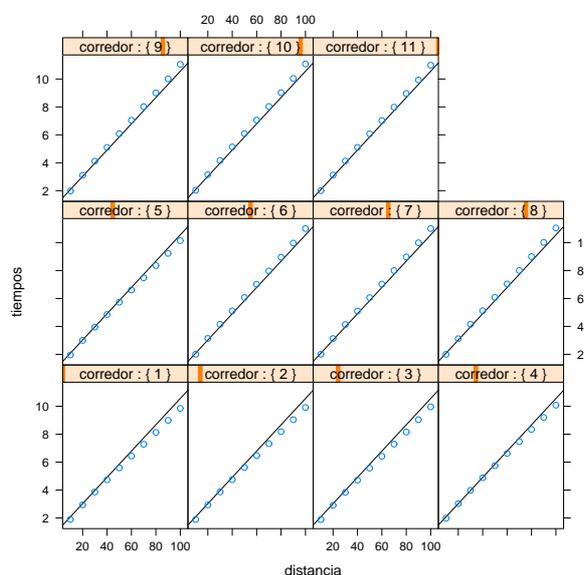


Figura 1.8. Gráfico de la distancia versus el tiempo en la prueba de 100 metros planos, Olímpicos de Londres 2012. Se observa una clara tendencia lineal.

**Ejercicio.** Se deja como ejercicio al lector describir los conjuntos de datos de esta sección a los cuales no se les realizó el respectivo análisis exploratorio.

Ahora se procederá a discutir y exponer algunos elementos teóricos fundamentales relacionados con el objetivo de este trabajo que consiste en presentar una introducción al modelamiento mixto.

# Capítulo 2

## ¿Qué es un modelo mixto?

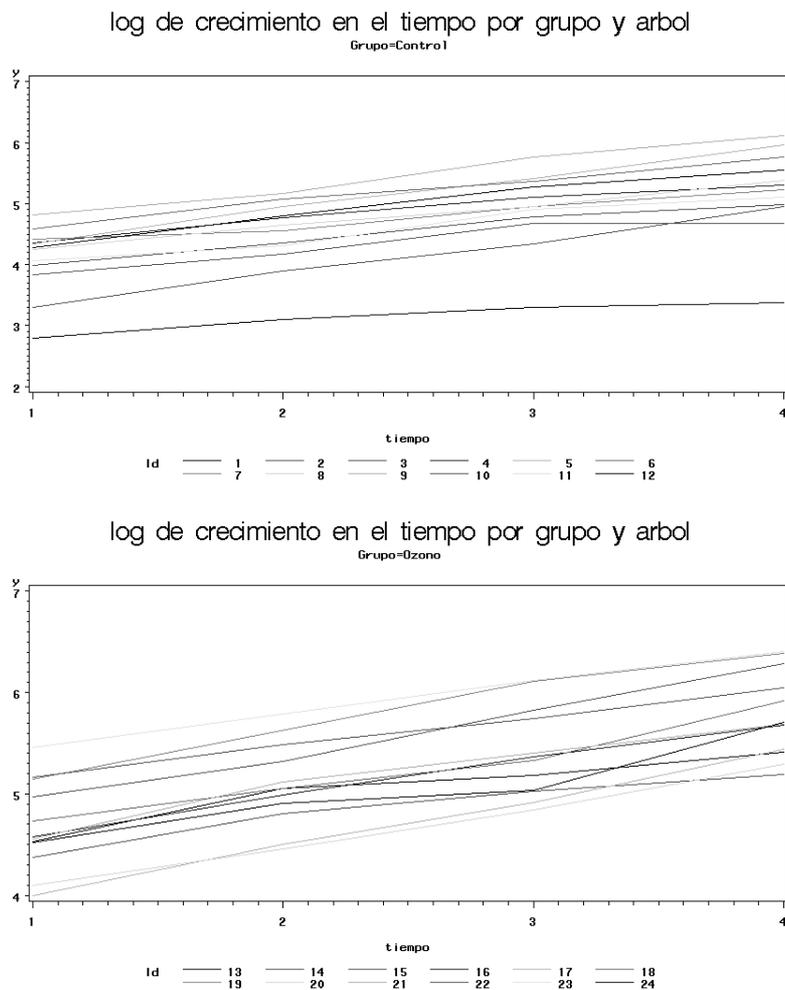
Los datos presentados anteriormente se pueden modelar de tal manera que se tengan en cuenta estructuras de asociación entre observaciones tomadas al mismo sujeto. Esto se puede hacer usando modelos mixtos. Pero ¿Qué es un modelo mixto? Para proporcionar una posible respuesta a esta pregunta considere el siguiente conjunto de datos acerca de logaritmos del crecimiento de árboles de abeto,

Obs	Grupo	Id	y <sub>152</sub>	y <sub>174</sub>	y <sub>201</sub>	y <sub>227</sub>
1	Control	1	2.79	3.10	3.30	3.38
2	Control	2	3.30	3.90	4.34	4.96
3	Control	3	3.98	4.36	4.79	4.99
4	Control	4	4.36	4.77	5.10	5.30
5	Control	5	4.34	4.95	5.42	5.97
6	Control	6	4.59	5.08	5.36	5.76
7	Control	7	4.41	4.56	4.95	5.23
8	Control	8	4.24	4.64	4.95	5.38
9	Control	9	4.82	5.17	5.76	6.12
10	Control	10	3.84	4.17	4.67	4.67
11	Control	11	4.07	4.31	4.90	5.10
12	Control	12	4.28	4.80	5.27	5.55

Obs	Grupo	Id	y <sub>152</sub>	y <sub>174</sub>	y <sub>201</sub>	y <sub>227</sub>
13	Ozono	13	4.53	5.05	5.18	5.41
14	Ozono	14	4.97	5.32	5.83	6.29
15	Ozono	15	4.37	4.81	5.03	5.19
16	Ozono	16	4.58	4.99	5.37	5.68
17	Ozono	17	4.00	4.50	4.92	5.44
18	Ozono	18	4.73	5.05	5.33	5.92
19	Ozono	19	5.15	5.63	6.11	6.39
20	Ozono	20	4.10	4.46	4.84	5.29
21	Ozono	21	4.56	5.12	5.40	5.69
22	Ozono	22	5.16	5.49	5.74	6.05
23	Ozono	23	5.46	5.79	6.12	6.41
24	Ozono	24	4.52	4.91	5.04	5.71

Observe que en este conjunto de datos, se tiene el mismo número de árboles por grupo (12) y además se registraron 4 medidas de crecimiento a los días

152, 174, 201 y 227, respectivamente. Es decir, cada árbol contribuye con 4 mediciones. El objetivo del estudio es comparar el crecimiento de los árboles en dos ambientes diferentes: Ambiente de control y Ambiente rico en ozono. A continuación se presenta un gráfico con los perfiles de cada árbol.



**Figura 2.1.** Gráfico de líneas del crecimiento de árboles.

En estos gráficos se observa un crecimiento lineal a través del tiempo (específicamente, en el intervalo 152-227 días) con un intercepto aleatorio (parece que cada árbol tiene su propio intercepto pero las pendientes son muy similares). Para analizar este tipo de datos existen algunas técnicas. Una técnica muy usada, que no representa necesariamente una buena estrategia de análisis, consiste en 'resumir' o 'comprimir' los datos por medio de estadísticos

de resumen. Algunas de estas técnicas se presentan a continuación.

## 2.1. Análisis estadístico: estadísticos de resumen

Una de las técnicas más comunes para describir datos es por medio de medidas básicas de resumen. Algunas de ellas sirven además para imputar datos faltantes. A continuación se muestra el procedimiento usual.

### Procedimiento general:

1. Resuma los datos de cada sujeto usando algún estadístico (por ejemplo la media muestral)
2. Haga un análisis con los estadísticos de resumen

De esta manera, el análisis de datos longitudinales, se reduce al análisis de observaciones independientes para los cuales se cuenta con una amplia gama de métodos estadísticos clásicos. Entre los métodos comúnmente más usados se encuentran:

1. Análisis de cada punto en el tiempo por separado. Por ejemplo, se puede usar una prueba t de Student para el punto de 152 días entre el ambiente control y el de ozono.
2. Área bajo la curva (Area under curve AUC). Comúnmente se usa como una medida de la eficacia de una droga o tratamiento. El problema en su uso radica en que dos curvas muy distintas pueden producir áreas iguales.
3. Análisis basado en el último punto (Last Observation Carried Forward LOCF.). Se recomienda en situaciones donde hay datos faltantes.
4. Análisis de incrementos. Este tipo de análisis es frecuente en estudios de crecimiento, por ejemplo crecimiento craneofacial.
5. Regresión lineal simple. Para cada sujeto se estima una recta de regresión usando mínimos cuadrados ordinarios (OLS) y luego se comparan los interceptos y las pendientes promedio de cada grupo.

Estos métodos presentan serias desventajas y limitaciones:

1. Con su uso se pueden perder grandes cantidades de información.

2. No dicen nada de la tendencia desde el primero hasta el último punto.
3. Se dificulta el ajuste por datos faltantes.
4. Ignoran la posible correlación entre las medidas repetidas realizados sobre el mismo sujeto.

Por las razones anteriores, su uso no es muy recomendable. Sin embargo, una modificación del método de estadísticos de resumen basado en regresión lineal simple, puede permitir aproximar bien los perfiles asociados a cada sujeto. Esta adaptación de este método de resumen se conoce como *modelo de dos etapas*.

## 2.2. Modelo de dos etapas

1. **Etapa 1:** Ajuste un modelo de regresión para cada sujeto por separado.
2. **Etapa 2:** Explique la variabilidad en los coeficientes de regresión, específicos a cada sujeto o sujeto-específicos, usando variables conocidas (efectos fijos).

Se procederá a ampliar estos puntos

### 2.2.1. Etapa 1 (formulación general)

Asuma que  $Y_{ij}$  es la respuesta para el  $i$ -ésimo sujeto medida en el tiempo  $X_{ij}$ ,  $i = 1, 2, \dots, m$  y  $j = 1, 2, \dots, n_i$ . Así, se tiene que  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})^T$  es el vector de respuestas para el sujeto  $i$ .

De esta manera, el modelo  $\mathbf{Y}_i = \mathbf{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ , describe la variabilidad dentro de los sujetos, donde  $\mathbf{Z}_i$  es una matriz de orden  $n_i \times q$  de covariables conocidas,  $\boldsymbol{\beta}_i$  es un vector  $q$ -dimensional con los coeficientes de regresión sujeto-específicos y  $\boldsymbol{\epsilon}_i$  se asume  $N(0, \Sigma_i)$  donde  $\Sigma_i$  es una matriz de varianzas y covarianzas. En resumen, la etapa 1 describe la variabilidad intra-sujeto (Within-Subject Variability).

### 2.2.2. Etapa 2

En esta etapa se modela la variabilidad entre los sujetos (Between-Subject Variability). Esta variabilidad puede modelarse, si los  $\beta_i$ 's se relacionan con

variables conocidas. En otras palabras si se hace  $\beta_i = \mathbf{K}_i\beta + \mathbf{b}_i$ , donde

$\mathbf{K}_i$  : es una matriz  $q \times p$  de covariables conocidas

$\beta$  : es un vector  $p$ -dimensional de parámetros de regresión desconocidos

$\mathbf{b}_i \sim N(0, \mathbf{D})$

donde  $\mathbf{D}$  es una matriz de orden  $q \times q$  de varianzas y covarianzas.

### 2.2.3. Ejemplo: Abetos

Como se vió en una sección anterior, para cada grupo de árboles, un modelo razonable incluye interceptos aleatorios y una pendiente común. Entonces, si se quiere usar el método de dos etapas, en la primera etapa se debe formular el modelo:

$$\mathbf{Y}_i = \mathbf{Z}_i\beta_i + \epsilon_i$$

donde

$$\mathbf{Z}_i = \begin{pmatrix} 1 & 152 \\ 1 & 174 \\ 1 & 201 \\ 1 & 227 \end{pmatrix}$$

y

$$\beta_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix}$$

aquí,  $\beta_{1i}$  es el intercepto desconocido para el sujeto  $i$  y  $\beta_{2i}$  es la pendiente desconocida para el sujeto  $i$

En la segunda etapa, los interceptos sujeto-específicos y los efectos del tiempo se relacionan con el tratamiento ambiental (*Control* u *Ozono*) de la siguiente forma:

$$\beta_{1i} = \beta_0 + b_{1i}$$

donde  $b_{1i}$  es un efecto aleatorio que modifica el intercepto y

$$\beta_{2i} = \beta_1 C_i + \beta_2 O_i$$

Son los efectos fijos ya que las pendientes se asumen iguales. Además,

$$C_i = \begin{cases} 1 & \text{Si } Control \\ 0 & \text{En otro caso} \end{cases}$$

y

$$O_i = \begin{cases} 1 & \text{Si el árbol se trata con Ozono} \\ 0 & \text{En otro caso} \end{cases}$$

La interpretación de parámetros es como sigue:

1.  $\beta_0$ : La respuesta promedio al comienzo del tratamiento.
2.  $\beta_1$  y  $\beta_2$ : Efectos promedio del tiempo para los grupos de *Control* y *Ozono* respectivamente.

De acuerdo a la manera en que se formulan los modelos, unas elecciones adecuadas para  $\beta$ ,  $\mathbf{K}_i$  y  $\mathbf{b}_i$  son:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\mathbf{K}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & C_i & O_i \end{pmatrix}$$

$$\mathbf{b}_i = \begin{pmatrix} \beta_i \\ 0 \end{pmatrix}$$

Una aproximación de dos etapas se puede llevar a cabo de manera explícita en el análisis. Sin embargo, este es solo otro ejemplo del uso de estadísticos de resumen, ya que  $\mathbf{Y}_i$  está resumido por medio de  $\beta_i$  el cual es un estadístico de resumen que se analiza en la segunda etapa. Los problemas asociados con esta estrategia basada en estadísticos de resumen se puede evitar combinando las dos etapas:

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{Z}_i \beta_i + \epsilon_i \\ \beta_i &= \mathbf{K}_i \beta + \mathbf{b}_i \end{aligned}$$

en un solo modelo:

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{Z}_i \mathbf{K}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \epsilon_i \\ &= \mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \epsilon_i \end{aligned}$$

Este modelo lineal mixto (MLM) se puede enunciar de una manera más

precisa:

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \quad (\mathbf{Z}_i\mathbf{K}_i = \mathbf{X}_i) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \\ \boldsymbol{\epsilon}_i &\sim N(\mathbf{0}, \boldsymbol{\Sigma}) \\ \mathbf{b}_1, \dots, \mathbf{b}_m, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_m &: \text{ mutuamente independientes} \end{aligned}$$

**Terminología:**

- $\boldsymbol{\beta}$ : Efectos fijos.
- $\mathbf{b}_i$  y  $\boldsymbol{\epsilon}_i$ : Efectos aleatorios.
- Componentes de varianza: Elementos de las matrices  $\mathbf{D}$  y  $\boldsymbol{\Sigma}_i$

Note que la matriz de varianzas y covarianzas del vector aleatorio  $\mathbf{Y}_i$  es

$$\text{Cov}(\mathbf{Y}_i) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \boldsymbol{\Sigma}_i$$

### 2.3. Forma matricial del MLM

Sean

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix}$$

(donde el vector asociado al sujeto  $i$ ,  $\mathbf{Y}_i$  tiene  $m_i$  componentes)

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}$$

$$\boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_m \end{pmatrix}$$

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_m \end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{pmatrix}$$

Así, la forma matricial del MLM queda definida como<sup>1</sup>:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$$

Donde  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  y

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \begin{pmatrix} \boldsymbol{\epsilon}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\epsilon}_2 & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\epsilon}_m \end{pmatrix}$$

$$\boldsymbol{\Sigma}_{\mathbf{b}} = \begin{pmatrix} \mathbf{D} & 0 & \cdots & 0 \\ 0 & \mathbf{D} & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{D} \end{pmatrix} = \mathbf{I}_m \otimes \mathbf{D}$$

( $\otimes$  Producto cruzado de Kronecker)

## 2.4. Ejemplo: Abetos (Continuación)

En el ejemplo de los árboles de abeto, la formulación del MLM es:

$$y_{ij} = \beta_o + \beta_1 C_i x_j + \beta_2 O_i x_j + b_i + \epsilon_{ij}$$

y puesto que  $C_i = 1 - O_i$ ,

$$Y_{ij} = \beta_o + \beta_1 x_j + (\beta_2 - \beta_1) O_i x_j + b_i + \epsilon_{ij}$$

---

<sup>1</sup>Compare con la forma matricial del Modelo Lineal Clásico:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

donde  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  y  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \sigma^2 \mathbf{I}$

donde  $b_i \sim N(0, \sigma^2)$  y  $\epsilon_i \sim N(0, \Sigma_i)$  (mutuamente independientes). Este modelo se denomina *Modelo de interceptos aleatorios* y se discute con más detalle en la siguiente sección.

## 2.5. Modelo de Laird-Ware

Laird y Ware (1982) formulan el modelo lineal mixto de la siguiente manera<sup>2</sup>

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \epsilon_i \text{ para } i = 1, 2, \dots, n$$

donde

- $\epsilon_i \sim N(\mathbf{0}, \mathbf{R}_i)$
- $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$
- $Cov(\epsilon_i, \mathbf{b}_i) = \mathbf{0}$

De lo anterior

$$\begin{aligned} E(\mathbf{y}_i | \mathbf{b}_i) &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i \\ Cov(\mathbf{y}_i | \mathbf{b}_i) &= \mathbf{R}_i \end{aligned}$$

**Objetivo:** Estimar  $\boldsymbol{\beta}$  y predecir las variables aleatorias  $\mathbf{b}_i$ .

$$\mathbf{Y}_{T \times 1} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} \quad \mathbf{X}_{T \times p} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \quad \mathbf{b}_{T \times 1} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} \quad \boldsymbol{\epsilon}_{T \times 1} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

donde  $T = \sum_{i=1}^n n_i$  es el número total de observaciones.

$$\mathbf{Z}_{T \times n \cdot k} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \end{bmatrix}$$

---

<sup>2</sup>Esta formulación no pretende ser redundante con la que se introdujo en la sección anterior la cual corresponde más bien a una motivación de la manera como se puede construir y justificar el MLM a partir de un problema con datos reales

$$\mathbf{B}_{n \cdot k \times n \cdot k} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D} \end{bmatrix}$$

$$\mathbf{R}_{T \times T} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_n \end{bmatrix}$$

Con lo anterior se obtiene

$$\text{Cov}(\mathbf{Y}) = \mathbf{V} = \mathbf{ZBZ}' + \mathbf{R}$$

**Lema 2.5.1.** *Bajo las anteriores definiciones se tiene:*

$$\begin{aligned} E(\mathbf{y}_i) &= E\{E(\mathbf{y}_i | \mathbf{b}_i)\} = \mathbf{X}_i \boldsymbol{\beta} \\ \text{Cov}(\mathbf{y}_i) &= E\{\text{Cov}(\mathbf{y}_i | \mathbf{b}_i)\} + \text{Cov}\{E(\mathbf{y}_i | \mathbf{b}_i)\} \\ &= \mathbf{R}_i + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' = \mathbf{V}_i \end{aligned}$$

La f.d.p. conjunta de  $\boldsymbol{\epsilon}_{T \times 1}$  y  $\mathbf{b}_{n \cdot k \times 1}$  es

$$\begin{aligned} f(\mathbf{b}, \boldsymbol{\epsilon}) &= (2\pi)^{-(T+nk)/2} \left| \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right| \\ &\times e \left\{ -\frac{1}{2} \begin{bmatrix} \mathbf{b} \\ \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Zb} \end{bmatrix}' \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{b} \\ \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Zb} \end{bmatrix} \right\} \end{aligned}$$

ya que  $\boldsymbol{\epsilon}$  y  $\mathbf{b}$  se asumen independientes. Aquí, las barras verticales se refieren al determinante de la respectiva matriz.

### 2.5.1. Modelo de interceptos aleatorios

Este es uno de los modelos más usados en modelos lineales mixtos. Cada sujeto es aproximadamente paralelo a la tendencia de su grupo. El modelo se expresa de manera genérica así:

$$y = \text{Tiempo} + \text{Grupo} + (\text{Grupo} \times \text{Tiempo}) + \text{Sujeto} + \text{Error}$$

que en símbolos es

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_e G_i + \beta_3 G_i \times T_{ij} + b_{0i} + \epsilon_{ij}$$

donde

- El sujeto es representado por  $i: i = 1, 2, \dots, N$ .
- Las observaciones para el sujeto  $i: j = 1, 2, \dots, m_i$ .
- $b_{0i} \sim N_{iid}(0, \sigma_b^2)$ .
- $\epsilon_{ij} \sim_{iid} N(0, \sigma_\epsilon^2)$ .

### 2.5.2. Formulaciones especiales del MLM

#### MLM para el modelo de regresión simple

**Caso 1:** Se asume que  $\mathbf{D}$  y  $\Sigma_i$  son conocidos. El Estimador de Máxima Verosimilitud (EMV) de  $\beta$  está dado por:

$$\hat{\beta} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y}$$

(Solución de mínimos cuadrados generalizados). Aquí,

$$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m)'$$

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)'$$

$$\mathbf{V} = \text{diag}\{\mathbf{V}_i\}$$

y

$$\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \Sigma_i$$

En caso de que  $\mathbf{V}_i$  no se conozca se puede usar  $\widehat{\mathbf{V}}_i$ . Como  $\mathbf{V}$  es una matriz diagonal en bloques, se tiene que

$$\hat{\beta} = \left( \sum_{i=1}^m \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^m \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Y}_i \right)$$

También se tiene que  $\widehat{\boldsymbol{\beta}}|\mathbf{b}_i$  es un estimador insesgado para  $\boldsymbol{\beta}$  con matriz de varianzas y covarianzas muestral especificada por:

$$\text{Cov}(\widehat{\boldsymbol{\beta}}|\mathbf{b}_i) = \left( \sum_{i=1}^m \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1}$$

**Caso 2:** Se asume que  $\mathbf{D}$  y/o  $\boldsymbol{\Sigma}_i$  no se conocen. A fin de encontrar un estimador de máxima verosimilitud (EMV) de  $\boldsymbol{\beta}$  se debe trabajar con la distribución conjunta de  $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$ . En esta situación se presentan problemas adicionales que se discutirán más adelante. De momento, se discute un caso que ilustra algunas de estas complicaciones.

**Caso especial:** Este caso especial está definido por un solo grupo que tiene datos balanceados en el modelo de intercepto aleatorio.

1. Datos balanceados donde  $m_i = m$ .
2. Errores de medición independientes:  $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{m \times m}$ .
3. El intercepto aleatorio solo genera una componente para la matriz  $\mathbf{D} = \sigma_b^2$ .

Bajo estas condiciones,

$$\mathbf{V}_i = \mathbf{V}_o$$

donde  $\mathbf{V}_o = \sigma^2 \mathbf{I} + \sigma_b^2 \mathbf{J}$  ( $\mathbf{J}$  es una matriz de orden  $m \times m$  de unos) (es decir, todos los sujetos tienen la misma matriz de varianzas y covarianzas definida por  $\mathbf{V}_o$ ). Cuando  $\sigma_b^2$  se conoce, los EMV del intercepto y la pendiente están dados por:

$$\begin{aligned} \widehat{\beta}_o &= \bar{Y}_{..} - \widehat{\beta}_1 \bar{X}_{..} \\ \widehat{\beta}_1 &= \frac{S_{XY}}{S_{XX}} \end{aligned}$$

aquí,  $S_{XY} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$  y  $S_{XX} = \sum_{i=1}^n (X_i - \bar{X})^2$  **Propiedades de estos estimadores:**

1. Están definidos sin dependencia de  $\sigma$  y  $\sigma_b^2$ .
2. Son idénticos a los estimadores obtenidos cuando  $\sigma_b^2 = 0$  (los obtenidos usando mínimos cuadrados ordinarios (OLS en inglés y MCO en español) sin efecto aleatorio).

3. La matriz de varianzas y covarianzas de estos estimadores se define en términos de  $\sigma$  y  $\sigma_b^2$ :

$$\begin{aligned} \text{Var}(\hat{\beta}_0) &= \frac{1}{m} \left\{ \sigma_b^2 + \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}^2}{S_{XX}} \right\} \\ \text{Var}(\hat{\beta}_1) &= \frac{1}{m} \left\{ \frac{\sigma^2}{S_{XX}} \right\} \end{aligned}$$

y

$$\text{Cov}\{\hat{\beta}_0, \hat{\beta}_1\} = -\bar{X} \cdot \text{Var}(\hat{\beta}_1)$$

4. Si  $\sigma_b^2$  no se conoce, entonces después de considerar la verosimilitud para  $\beta_0$ ,  $\beta_1$  y  $\sigma_b^2$  se encuentra que los EMV de los efectos fijos son los mismos de arriba (y por lo tanto no dependen de las componentes de varianza). Pero, la matriz de varianzas y covarianzas está definida en términos de las componentes de varianza desconocidas. Específicamente, los estimadores de las componentes de varianza son:

$$\hat{\sigma}^2 = \frac{SSR}{m(n-1)}$$

y

$$\hat{\sigma}_b^2 = \frac{1}{n} \left( \frac{SSB}{m} - \hat{\sigma}^2 \right)$$

donde

$$SSR = SSE - m \frac{\hat{\beta}_1^2}{S_{XX}}$$

$$SSB = n \sum_{i=1}^m (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$SSE = \sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i.})^2$$

Esta formulación asume que  $\hat{\sigma}_b^2 > 0$ , sin embargo puede ser que  $\hat{\sigma}_b^2 < 0$  cuando  $\hat{\sigma}^2 > \frac{SSB}{m}$ . A fin de sortear esta dificultad se toma

$$\hat{\sigma}^2 = \frac{SSB + SSR}{mn}$$

y

$$\hat{\sigma}_b^2 = 0$$

cuando  $SSB < \frac{SSR}{(n-1)}$

Teniendo a  $\hat{\sigma}_b^2$  y a  $\hat{\sigma}^2$  se puede obtener el *coeficiente de correlación intraclase* (ICC) que se formula como:

$$\hat{\rho} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}^2}$$

Este ICC mide la covarianza entre medidas tomadas a la misma unidad experimental.

### Consideraciones de diseño para el modelo de regresión simple

¿Cómo planear un experimento longitudinal para comparar dos grupos cuando se especifica un modelo de regresión lineal simple? Para esto es necesario fijar de antemano:

1. Las tasas de error Tipo I y tipo II ( $\alpha$  y  $\beta$ , respectivamente).
2. El tipo de hipótesis alternativa (de una o dos colas).
3. El número de puntos ( $n$ ) en el tiempo para cada sujeto en cada grupo.
4. Especificar la estructura de Simetría Compuesta:

$$\mathbf{V}_o = Cov(\mathbf{Y}_i) = \sigma^2 \mathbf{I} + \sigma_b^2 \mathbf{J}$$

Sea  $\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} = Cov(Y_{ij}, Y'_{ij}) = ICC$  (Correlación entre dos respuestas distintas dentro del mismo sujeto).

5.  $d$ =Una medida del tamaño de la diferencia entre los dos grupos.

Dos situaciones comunes en este caso:

**Situación 1:** Se desean comparar las pendientes entre los grupos usando un modelo de intercepto aleatorio. Sea  $d$  la diferencia hipotética entre las pendientes. Entonces usando  $Var(\hat{\beta}_1) = \frac{1}{m} \left\{ \frac{\sigma^2}{S_{XX}} \right\}$  se tiene que:

$$\begin{aligned} Var(\hat{\beta}_1) &= \frac{1}{m} \left\{ \frac{\sigma^2}{S_{XX}} \right\} \\ &= \frac{Var(Y_{ij})(1 - \rho)}{mS_{XX}} \end{aligned}$$

Ya que la  $Var(Y_{ij}) = \sigma^2 + \sigma_b^2$  y  $(1 - \rho) = 1 - \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} = \frac{\sigma^2}{\sigma_b^2 + \sigma^2}$ , se tiene que  $Var(Y_{ij})(1 - \rho) = \sigma^2 + \sigma_b^2 \times \frac{\sigma^2}{\sigma_b^2 + \sigma^2} = \sigma^2$ . Entonces, puesto que las pendientes se estiman independientemente en cada grupo, el mínimo número de sujetos en cada grupo está dado por:

$$m = 2 \frac{(z_\alpha + z_{1-\beta})^2 Var(Y_{ij})(1 - \rho)}{S_{XX}d^2}$$

Esta fórmula sirve para determinar el número de unidades en cada grupo para comparar dos medias poblacionales.

**Ejemplo 2.5.1.** *Considere un ensayo clínico hipotético diseñado para medir el efecto de una nueva droga para reducir la presión sanguínea. Asuma tres visitas planeadas al inicio (año 0), a los 2 y a los 5 años (es decir  $n=3$  y  $S_{XX} = 12.667^3$ ). Asuma que  $\alpha = 0.054$ ,  $d=0.5$  mm Hg/año, un test de una cola y  $\beta = 0.20$  (para una potencia de 80%). De esta manera  $(1.645+0.84)^2 = 6.18$ . La siguiente tabla muestra el número de sujetos ( $m$ ) que se necesitan para el grupo de Control y el de Tratamiento para distintos valores de  $\rho$  y  $Var(Y_{ij})$ :*

	$Var(Y_{ij})$		
$\rho$	100	200	300
0.2	313	625	937
0.5	195	391	586
0.8	79	157	235

**Tabla 2.1.** Número de sujetos por grupo para valores de  $\rho$  y  $Var(Y_{ij})$ .

Por ejemplo, si  $\rho=0.2$  y  $Var(Y_{ij})=100$ , se tiene que

$$m = \frac{2 \times 6.18 \times 100 \times (1 - 0.2)}{0.5^2 \times 12.667} \approx 313$$

Note que para valores fijos de  $Var(Y_{ij})$ , el mínimo tamaño muestral requerido decrece a medida que  $\rho$  se incrementa.

**Situación 2:** Se desea comparar la diferencia en el tiempo promedio en las respuestas entre los dos grupos o simplemente comparar las medias entre los dos grupos. Esto aplica a situaciones donde la respuesta es no lineal en

<sup>3</sup> $S_{XX} = (0 - 2.333)^2 + (2 - 2.333)^2 + (5 - 2.333)^2 = 12.667$

el tiempo sino que es una constante en el tiempo. De esta manera el MLM se reduce a un ANOVA de una vía con efectos aleatorios. En cada grupo, la estimación de la media real es  $\bar{Y}_{..}$  (la respuesta promedio sobre todos los sujetos y puntos en el tiempo). Observe que:

$$\begin{aligned} \text{Var}(\bar{Y}_{..}) &= \frac{1}{m} \left( \sigma_b^2 + \frac{\sigma^2}{n} \right) \\ &= \frac{1}{m} \text{Var}(Y_{ij}) \left\{ \frac{1 + (n-1)\rho}{n} \right\} \end{aligned}$$

De esta manera, para detectar  $d$ , la diferencia en la media verdadera entre los dos grupos, se requiere un mínimo de

$$m = \frac{1}{n} \left[ 2 \frac{(z_\alpha + z_{1-\beta})^2}{E^2} \right] [1 + (n-1)\rho]$$

donde,

$$E = \text{Tamaño del efecto} = \frac{d}{(\text{Var}(Y_{ij}))^{0.5}}$$

**Ejemplo 2.5.2.** *Asuma  $\alpha=0.05$ , una prueba de una cola, una potencia (o poder) del 80% y  $n=3$  observaciones por sujeto. La tabla de abajo muestra los valores de  $m$  (el mínimo tamaño muestral requerido por grupo para detectar un tamaño de efecto  $E$ )*

	$E$			
$\rho$	20%	30%	40%	50%
0.2	146	68	37	24
0.5	208	93	52	34
0.8	270	1201	68	44

**Tabla 2.2.** *Valores de  $m$  para detectar un tamaño de efecto  $E$ .*

*Observe que en esta situación, el valor requerido para  $m$  crece a medida que  $\rho$  crece, que es opuesto a lo que se observó en la situación 1. Posibles razones de este comportamiento: En la situación 1, el parámetro de interés es la pendiente o tasa de cambio en  $Y$ . La contribución de cada sujeto en la estimación de esta pendiente es un contraste lineal en las respuestas individuales que tienen varianzas que decrecen con  $\rho$ . Sin embargo en la situación 2, el parámetro de interés es el promedio de todas las respuestas individuales que tiene varianzas que se incrementan con  $\rho$ .*

Diggle et al. (1994) argumentan que:

1. Los modelos de intercepto aleatorio no son siempre necesarios en estudios longitudinales.
2. Los estudios longitudinales<sup>4</sup> son usualmente más eficientes que los estudios de corte transversal<sup>5</sup>.

Se discutirán estos dos argumentos. Respecto al argumento 1), considere el modelo de regresión lineal simple

$$Y_{ij} = \beta_o + \beta X_{ij} + \epsilon \quad i = 1, \dots, m; \quad j = 1, \dots, n \quad (1)$$

Puesto que los datos son logitudinales, este modelo se puede escribir como

$$Y_{ij} = \beta_o + \beta X_{i1} + \beta (X_{ij} - X_{i1}) + \epsilon_{ij} \quad (2)$$

Aquí  $X_{i1}$  mide el efecto de corte transversal de la covariable mientras que  $(X_{ij} - X_{i1})$  mide el efecto longitudinal de la covariable. El coeficiente común  $\beta$  asume que ambos efectos son los mismos. Una modificación del modelo (1) consiste en usar un MLM al reemplazar  $\beta_o + \beta X_{i1}$  en (2) por  $\beta_{0i}$  que produce un modelo de intercepto aleatorio que ya se discutió previamente.

Una segunda modificación que evita usar el modelo de intercepto aleatorio consiste en generalizar (2) de la siguiente manera:

$$Y_{ij} = \beta_o + \beta_C X_{i1} + \beta_L (X_{ij} - X_{i1}) + \epsilon_{ij} \quad (3)$$

donde ahora los coeficientes de los efectos de corte transversal  $\beta_C$  y longitudinal  $\beta_L$  son distintos. Se puede ajustar el modelo (3) a los datos y probar la hipótesis nula de que estos coeficientes son equivalentes. También se puede usar (3) para investigar el sesgo en el estimador usual de  $\beta$  en el modelo (1). Recuerde que

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}}$$

---

<sup>4</sup>Un *estudio prospectivo* se fundamenta en la observación y monitoreo en el tiempo de eventos futuros relacionados con la enfermedad. Por ejemplo, un estudio que incluye adultos mayores para ver como evolucionan sus habilidades mentales durante los siguientes 15 años. Son estudios dinámicos.

<sup>5</sup>Un *estudio de corte o transversal* (también llamados de corte transversal o de prevalencia) se preocupa por mirar eventos presentes en un determinado punto en el tiempo. Por ejemplo, las encuestas de prevalencia que muestran la frecuencia y las características con las que se presenta una enfermedad en un momento determinado. Estos estudios solo pueden medir la prevalencia de la enfermedad (es como tomar una 'foto' del problema). Son estudios estáticos

o

$$\begin{aligned} S_{XX}\hat{\beta} &= \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_{..}) (Y_{ij} - \bar{Y}_{..}) \\ &= \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_{..}) Y_{ij} \end{aligned}$$

Tomando valores esperados en (3) y simplificando se obtiene,

$$E(\hat{\beta}) = \beta_L + A(\beta_C - \beta_L)$$

donde

$$S_{XX}A = \sum_{i=1}^m n(\bar{X}_{i.} - \bar{X}_{..})(\bar{X}_{i1} - \bar{X}_{1.})$$

**Conclusión:** El estimador usual de  $\beta$  en el modelo (1) es insesgado dado que  $\beta_C = \beta_L$  o  $X_{i1}$  y  $X_{i.}$  sean ortogonales. Una condición suficiente para que esta ortogonalidad se cumpla es que las observaciones estén igualmente espaciadas (al menos de manera aproximada) en el estudio.

Respecto al argumento 2), asuma que  $\beta_C = \beta_L$ . Ajuste el modelo (1) con  $n=1$  (un estudio de corte transversal real). De esta manera,

$$Var(\hat{\beta}_C) = \frac{\sigma^2}{S_{XX}^{(1)}}$$

donde

$$S_{XX}^{(1)} = \sum_{i=1}^m (X_{i1} - X_{1.})^2$$

Por otra parte, si se está dispuesto a comprometer recursos para llevar a cabo un verdadero estudio longitudinal y ajustar el modelo (1) con  $n > 1$ , entonces

$$Var(\hat{\beta}_C) = \text{Entrada inferior derecha de } \left\{ \sum_{i=1}^m \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{X}_i \right\}^{-1}$$

donde  $\mathbf{V}_i$  es la matriz de varianzas y covarianzas de  $(\epsilon_{i1}, \dots, \epsilon_{in})'$ . Defina la eficiencia de estas dos estimaciones

$$e = \frac{Var(\beta_L)}{Var(\beta_C)}$$

Diggle et al. (1994) establecen que si la  $cor(\epsilon_{i\ell}, \epsilon_{ik}) = \rho$  (simetría compuesta o correlación uniforme), entonces

$$e = \frac{\{1 + (n - 1)\rho\}(1 - \rho)}{B}$$

donde

$$B = n(1 + \delta) \left\{ 1 - \rho + \frac{np\delta}{(1 + d)} \right\}$$

y

$$\delta = \frac{mS_{XX}}{m(n - 1)S_{XX}^{(1)}}$$

**Conclusión:** A menos que  $\delta$  sea pequeño y  $\rho$  cercano a 1, hay mucho que ganar al llevar a cabo estudios longitudinales aún si  $n = 2$

### 2.5.3. Modelo de interceptos y tendencias aleatorias

Esta es otra de las formulaciones más usada en modelos lineales mixtos. Cada sujeto es diferente de los demás en términos del intercepto y la tendencia de su grupo. El modelo se expresa así:

$$y = \text{Tiempo} + \text{Grupo} + (\text{Grupo} \times \text{Tiempo}) + \text{Sujeto} + (\text{Sujeto} \times \text{Tiempo}) + \text{Error}$$

que en símbolos es

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_e G_i + \beta_3 G_i \times T_{ij} + b_{0i} + b_{1i} T_{ij} + \epsilon_{ij}$$

donde

$$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

y

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_0, b_1} \\ \sigma_{b_0, b_1} & \sigma_{b_1}^2 \end{bmatrix} \right)$$

En forma matricial se tiene

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix} = \begin{bmatrix} 1 & T_{i1} & G_i & G_i \times T_{i1} \\ 1 & T_{i2} & G_i & G_i \times T_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & T_{in_i} & G_i & G_i \times T_{in_i} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 & T_{i1} \\ 1 & T_{i2} \\ \vdots & \vdots \\ 1 & T_{in_i} \end{bmatrix} \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{bmatrix}$$

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i$$

$$\text{con } \boldsymbol{\Sigma}_{y_i} = \mathbf{Z}_i \boldsymbol{\Sigma}_b \mathbf{Z}_i' + \sigma_\epsilon^2 \mathbf{I}_{n_i}$$

## 2.6. Modelo de regresión con errores autocorrelacionados

Asuma que el modelo para un sujeto (se tienen  $T$  mediciones por sujeto y  $n$  sujetos independientes) es

$$\begin{aligned} Y_t &= \mathbf{x}'_t \boldsymbol{\beta} + \epsilon_t, \quad t = 1, \dots, T \\ \epsilon_t &= \xi_t + \rho \epsilon_{t-1}, \quad t = 2, \dots, T \\ \rho &\in (-1, 1) \\ \xi_t &\sim N(0, \sigma^2), \quad t = 2, \dots, T \\ \epsilon_t &\sim N\left(0, \frac{\sigma^2}{1 - \rho^2}\right) \end{aligned}$$

1. De estos supuestos se tiene que las innovaciones o shocks  $\{\xi_2, \xi_3, \dots, \xi_T\}$  son variables aleatorias independientes e idénticamente distribuidas (v.a.'s i.i.d).
2.  $\epsilon_1$  es independiente de los  $\{\xi_2, \xi_3, \dots, \xi_T\}$  y  $\epsilon_1 \sim N\left(0, \frac{\sigma^2}{1 - \rho^2}\right)$ .
3. Escriba los términos de error como una serie geométrica de las innovaciones pasadas

$$\begin{aligned} \epsilon_2 &= \xi_2 + \rho \epsilon_1 \\ \epsilon_3 &= \xi_3 + \rho \epsilon_2 = \xi_3 + \rho \xi_2 + \rho^2 \epsilon_1 \\ \epsilon_4 &= \xi_4 + \rho \epsilon_3 = \xi_4 + \rho \xi_3 + \rho^2 \xi_2 + \rho^3 \epsilon_1 \\ &\vdots \\ \epsilon_t &= \xi_t + \rho \epsilon_{t-1} = \xi_t + \rho \xi_{t-1} + \rho^2 \xi_{t-2} + \dots + \rho^{t-1} \epsilon_1 \end{aligned}$$

4. Todas las varianzas de los  $\epsilon_t$  son iguales a  $\sigma_\epsilon^2$

$$\begin{aligned} \text{Var}(\epsilon_t) &= \sigma^2 + \rho^2 \text{Var}(\epsilon_{t-1}) \\ \sigma_\epsilon^2 &= \sigma^2 + \rho^2 \sigma_\epsilon^2 \\ \sigma_\epsilon^2 &= \frac{\sigma^2}{1 - \rho^2} \end{aligned}$$

5. Usando la expresión en series geométrica para  $\epsilon_t$ , las covarianzas son

$$E(\epsilon_t \epsilon_{t-u}) = \sigma^2 \frac{\rho^u}{1 - \rho^2}, \quad u > 0$$

6. La matriz de varianzas y covarianzas de los términos de error

$$E(\epsilon \epsilon') = \Sigma = \frac{\sigma^2}{1 - \rho^2} \Gamma$$

donde

$$\Gamma = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{t-1} \\ \rho & 1 & \rho & \dots & \rho^{t-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{t-1} & \rho^{t-2} & \rho^{t-3} & \dots & 1 \end{bmatrix}$$

- 7.

$$\Gamma^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & \dots & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \dots & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & \dots & \dots & \dots & -\rho & 1 + \rho^2 & -\rho \\ 0 & \dots & \dots & \dots & 0 & -\rho & 1 \end{bmatrix}$$

$$\det(\Gamma) = (1 - \rho^2)^{T-1}$$

8. Defina los residuales

$$r_t = y_t - \mathbf{x}_t' \boldsymbol{\beta}$$

La distribución condicional de  $\mathbf{Y}$  dado  $\boldsymbol{\beta}$ ,  $\sigma$  y  $\rho$  es

$$\begin{aligned}
 f(\mathbf{Y} | \boldsymbol{\beta}, \sigma, \rho) &\propto |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\
 &\propto \frac{(1 - \rho^2)^{1/2}}{\sigma^T} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (1 - \rho^2) r_1^2 + \sum_{t=2}^T (r_t - \rho r_{t-1})^2 \right] \right\}
 \end{aligned}$$

9. Note que

$$r_t - \rho r_{t-1} = y_t - \rho y_{t-1} - (\mathbf{x}_t - \rho \mathbf{x}_{t-1})' \boldsymbol{\beta}$$

En este capítulo se han discutido aspectos relacionados con el modelo lineal mixto y con las propiedades de algunos estimadores. Se presentaron algunas formulaciones especiales del modelo y formas de calcular tamaños muestrales en la etapa de diseño. En el capítulo siguiente se discutirán aspectos relacionados con inferencia y se ampliará la discusión sobre métodos de estimación.

# Capítulo 3

## Inferencias en el Modelo Lineal Mixto

### 3.1. MLM: Modelos jerárquicos versus marginales

Recuerde que el MLM está dado por:

$$\begin{aligned}\mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \\ \boldsymbol{\epsilon}_i &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)\end{aligned}$$

y  $\mathbf{b}_1, \dots, \mathbf{b}_m$  y  $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_m$  son independientes.

Puede formularse un *modelo jerárquico o condicional*:

$$\begin{aligned}\mathbf{Y}_i|\mathbf{b}_i &\sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \boldsymbol{\Sigma}_i) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D})\end{aligned}$$

(Observe que se formula un modelo para  $\mathbf{Y}_i$  dado  $\mathbf{b}_i$  y un modelo para  $\mathbf{b}_i$ ). Los modelos jerárquicos no siempre están restringidos al caso normal.

La función de densidad marginal de  $\mathbf{Y}_i$  está dada por:

$$f(\mathbf{y}_i) = \int f(\mathbf{y}_i|\mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i$$

la cual es la función de densidad asociada a una distribución normal  $n$ -dimensional con vector de medias  $\mathbf{X}_i\boldsymbol{\beta}$  y matriz de varianzas y covarianzas:

$$\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'$$

De esta manera, el modelo marginal implicado por la aproximación de dos etapas, hace supuestos específicos acerca de la dependencia de la estructura de la media y la estructura de covarianza de las covariables  $\mathbf{X}_i$  y  $\mathbf{Z}_i$ . Específicamente:

$$\text{Media implicada : } \mathbf{X}_i\boldsymbol{\beta}$$

y

$$\text{Covarianza implicada : } \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'$$

Observe que el modelo jerárquico implica el modelo marginal (pero no en sentido contrario). El modelo marginal es importante en la medida que la inferencia para el MLM se basa en el modelo marginal (aunque esto no implica que no se pueda basar en el modelo jerárquico). Generalmente, el SAS PROC MIXED© ajusta el modelo condicional a los datos. El SAS PROC GENMOD© por su parte, ajusta el modelo marginal

#### **Comentarios:**

1. En la práctica, basar la inferencia en el modelo marginal puede ocasionar problemas numéricos en el proceso de optimización de la verosimilitud ya que no hay garantía de que siempre exista convergencia a algún valor de un parámetro (esto es válido para el modelo marginal y el jerárquico).
2. Las inferencias basadas en el modelo marginal no asumen de manera explícita la presencia de efectos aleatorios.

## **3.2. Estimación**

### **3.2.1. Modelo clásico**

La maximización de  $f(\mathbf{b}, \boldsymbol{\epsilon})$  con respecto a  $\boldsymbol{\beta}$  y  $\mathbf{b}$  requiere la minimización de

$$\begin{aligned} \mathbf{Q} &= \begin{bmatrix} \mathbf{b} \\ \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b} \end{bmatrix}' \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{b} \\ \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b} \end{bmatrix} \\ &= \mathbf{b}'\mathbf{B}\mathbf{b} + (\mathbf{r} - \mathbf{Z}\mathbf{b})' \mathbf{R}^{-1} (\mathbf{r} - \mathbf{Z}\mathbf{b}) \end{aligned}$$

donde

$$\mathbf{r} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$$

Aquí,

$$\mathbf{B}_{n \cdot k \times n \cdot k} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D} \end{bmatrix}$$

Tomando derivadas con respecto a  $\boldsymbol{\beta}$  y  $\mathbf{b}$  e igualando a  $\mathbf{0}$  se obtiene el siguiente sistema de ecuaciones:

$$\frac{\partial \mathbf{Q}}{\partial \boldsymbol{\beta}} = \mathbf{0} \quad (1)$$

$$\frac{\partial \mathbf{Q}}{\partial \mathbf{b}} = \mathbf{0} \quad (2)$$

El anterior sistema lleva a lo que se conoce como *ecuaciones del modelo mixto de Henderson* (Henderson, 1975):

$$\begin{aligned} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{b}} &= \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})\hat{\mathbf{b}} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{aligned}$$

Lo anterior en forma matricial queda como

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix}$$

y se calcula

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix}$$

Para hallar las soluciones, de (2) se tiene

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1}) \hat{\mathbf{b}} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y}$$

lo que lleva a

$$\hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (3)$$

**Lema 3.2.1.**

$$\begin{aligned} (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})^{-1} &= \mathbf{B} - \mathbf{BZ}'(\mathbf{ZBZ}' + \mathbf{R})^{-1}\mathbf{ZB} \\ &= \mathbf{B} - \mathbf{BZ}'\mathbf{V}^{-1}\mathbf{ZB} \end{aligned}$$

Sustituyendo el anterior resultado en (3)

$$\begin{aligned} \hat{\mathbf{b}} &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= (\mathbf{B} - \mathbf{BZ}'(\mathbf{ZBZ}' + \mathbf{R})^{-1}\mathbf{ZB})^{-1} \mathbf{Z}'\mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{BZ}' [\mathbf{R}^{-1} - (\mathbf{ZBZ}' + \mathbf{R})^{-1} \mathbf{ZBZ}'\mathbf{R}^{-1}] (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{BZ}' (\mathbf{ZBZ}' + \mathbf{R})^{-1} [(\mathbf{ZBZ}' + \mathbf{R}) \mathbf{R}^{-1} - \mathbf{ZBZ}'\mathbf{R}^{-1}] (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{BZ}' (\mathbf{ZBZ}' + \mathbf{R})^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{BZ}'\mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned}$$

Para hallar una solución para  $\boldsymbol{\beta}$  se encuentra una solución para  $\mathbf{b}$  y se reemplaza en (1). Sustituyendo (3) en (1) se obtiene

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y}$$

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{R}^{-1} [\mathbf{I} - \mathbf{Z} (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})] \mathbf{Y}$$

$$\mathbf{X}' [\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z} (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1}] \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}' [\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z} (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1}] \mathbf{Y}$$

**Lema 3.2.2.**

$$(\mathbf{Z}'\mathbf{B}\mathbf{Z} + \mathbf{R})^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{B}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}$$

Aplicando este resultado, lo anterior se reduce a

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

De donde se obtiene

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

La solución para  $\mathbf{b}$  será

$$\hat{\mathbf{b}} = \mathbf{B}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Este es el mejor predictor lineal insesgado (BLUP) para  $\mathbf{b}^1$

**Definición 3.2.1.** En el MLM, sea

$$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m)'$$

Un contraste de error es cualquier combinación lineal  $\mathbf{a}'\mathbf{Y}$  tal que  $E(\mathbf{a}'\mathbf{Y}) = 0$  para todo  $\boldsymbol{\beta}$ . Es decir, se debe satisfacer la condición de ortogonalidad

$$\mathbf{a}'\mathbf{X} = \mathbf{0}$$

donde  $\mathbf{X}$  es la matriz de diseño. En el MLM, si  $\text{rango}(\mathbf{X}) = p$  entonces cualquier conjunto de errores de contraste contiene a lo más  $n - p$  contrastes de error linealmente independientes ( $n$  es el número total de observaciones).

Es una práctica común, definir un conjunto de errores seleccionando una matriz cuadrada  $n \times n$  de rango  $n - p$  (lo cual implica que el conjunto de errores de contraste contiene redundancias).

---

<sup>1</sup>Los BLUP se usan para estimar efectos aleatorios. Los BLUP son similares a los BLUE (Best Linear Unbiased Estimator) de efectos fijos. La diferencia radica en que se habla de *estimar* efectos fijos y *predecir* efectos aleatorios, sin embargo las ecuaciones para uno u otro son diferentes. Los BLUP son similares a los estimadores empíricos de Bayes de efectos aleatorios en el MLM, excepto que en este último caso los pesos dependen de valores desconocidos de componentes de varianza que se reemplazan por estimaciones basadas en la muestra.

### 3.2.2. Estimación basada en la verosimilitud perfilada

Si  $v(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{Y})$  es dos veces el negativo de la log-verosimilitud marginal

$$-2 \log(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{Y}) = \log(|\mathbf{V}(\boldsymbol{\theta})|) + (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + T \log(2\pi) \quad (4)$$

$v$  debe ser minimizada con respecto a  $\boldsymbol{\beta}$  y  $\boldsymbol{\theta}$  (el término  $T \log(2\pi)$ , al ser constante, no afecta el proceso de maximización de la log-verosimilitud). Esto se hace mediante el perfilamiento (profiling) con respecto a  $\boldsymbol{\beta}$ :

- En este proceso se halla una expresión cerrada para  $\hat{\boldsymbol{\beta}}$  y se sustituye en  $v(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{Y})$ .
- La expresión resultante será una función de  $\boldsymbol{\theta}$  y es dos veces el negativo de la verosimilitud perfilada.

La minimización de (4) con respecto a  $\boldsymbol{\beta}$  requiere la minimización de

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

que es la suma de cuadrados generalizada. Se tiene entonces

$$\frac{\partial v(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{Y})}{\partial \boldsymbol{\beta}} = 2\mathbf{X}'\mathbf{V}(\boldsymbol{\theta}) - 2\mathbf{X}'\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Y}$$

Luego

$$v(\boldsymbol{\theta} | \mathbf{Y}) = \log(|\mathbf{V}(\boldsymbol{\theta})|) + (\mathbf{Y} - \mathbf{X}\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{X}'\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Y})' \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{Y} - \mathbf{X}\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{X}'\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Y}) + T \log(2\pi)$$

La ecuación estimadora para el  $j$ -ésimo elemento de  $\boldsymbol{\theta}$  es

$$\text{traza} \left[ \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right] + \mathbf{Y}' \frac{\partial \mathbf{Q}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{Y} = 0$$

donde

$$\mathbf{Q}(\boldsymbol{\theta}) = \mathbf{V}^{-1}(\boldsymbol{\theta}) - \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{X}'\mathbf{V}^{-1}(\boldsymbol{\theta})$$

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

Sustituyendo en (4) se obtiene la función objetivo perfilada

La solución de estas ecuaciones produce  $\hat{\boldsymbol{\theta}}_M$  que se conoce como *estimador de mínimos cuadrados generalizados* (generalized least squares). Sustituyendo en la ecuación (4) para la estimación de  $\boldsymbol{\beta}$  produce

$$\begin{aligned}\hat{\boldsymbol{\beta}}_M &= (\mathbf{X}'\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}}_M)\mathbf{X})\mathbf{X}'\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}}_M)\mathbf{Y} \\ &= \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}_M)\mathbf{X}'\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}}_M)\mathbf{Y}\end{aligned}$$

**Nota:** Frecuentemente, la

$$Cov(\mathbf{Y}) = \sigma^2\mathbf{ZB}^*\mathbf{Z}' + \sigma^2\mathbf{R}^*$$

donde  $\sigma^2$  es un parámetro de escala. En este caso  $\sigma^2$  puede ser perfilado de la verosimilitud. La solución es

$$\hat{\sigma}^2 = \frac{1}{T} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{ZB}^*\mathbf{Z}' + \mathbf{R}^*)^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

que se sustituye en (4).

### 3.2.3. Estimación por máxima verosimilitud (ML)

**Notación:**  $\boldsymbol{\beta}$ : Vector de efectos fijos

$\boldsymbol{\alpha}$ : Vector de todas las componentes de varianza en  $\mathbf{D}$  y  $\boldsymbol{\Sigma}_i$

$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ : Vector de todos los parámetros en el modelo marginal.

**Función de verosimilitud marginal:**

$$L_{ML}(\boldsymbol{\theta}) = \prod_{i=1}^m \left\{ (2\pi)^{-\frac{n_i}{2}} |\mathbf{V}_i(\boldsymbol{\alpha})|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\boldsymbol{\alpha})(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})} \right\}$$

Si  $\boldsymbol{\alpha}$  se conoce, los EMV de  $\boldsymbol{\beta}$  son:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left( \sum_{i=1}^m \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i \right)^{-1} \sum_{i=1}^m \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i$$

donde  $\mathbf{W}_i = \mathbf{V}_i^{-1}$ . Sin embargo, en general  $\boldsymbol{\alpha}$  no se conoce y es necesario reemplazarlo por un estimador  $\hat{\boldsymbol{\alpha}}$  que se obtiene ya sea por máxima verosimilitud (ML) o por máxima verosimilitud restringida (REML) (se discute

más adelante). Usando el método ML, se puede obtener a  $\hat{\alpha}$  maximizando, con respecto a  $\alpha$ , la función  $L_{ML}(\alpha, \hat{\beta}(\alpha))$ . El estimador resultante  $\hat{\beta}(\hat{\alpha})$  para  $\beta$  se denota  $\hat{\beta}_{ML}$ . Note que es posible obtener estimaciones para  $\beta$  y  $\alpha$  maximizando  $L_{ML}(\theta)$  respecto a  $\theta$  (es decir, respecto a  $\beta$  y  $\alpha$  simultáneamente).

En el pasado, la estimación de los efectos fijos y las componentes de varianza por máxima verosimilitud no era común debido a dificultades de tipo computacional que requerían la solución numérica de un problema de optimización no lineal con restricciones. Harville (1977) revisó trabajos previos, unificó la metodología y describió algoritmos iterativos de ML para obtener estimaciones de los parámetros.

Adicional a los problemas numéricos, los estimadores de las componentes de varianza de ML tienden a ser sesgados. Para ilustrar, considere el problema de estimar a  $\sigma^2$  en el modelo lineal clásico  $\mathbf{Y} = \mathbf{X}\beta + \epsilon$  donde  $\mathbf{Y}$  es un vector  $n \times 1$ ,  $\mathbf{X}$  es una matriz de diseño de orden  $n \times p$  de rango  $p$  y

$$\epsilon \sim N_n(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$$

El EMV de  $\sigma^2$  es

$$\tilde{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})}{n}$$

donde  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ . El estimador REML de  $\sigma^2$  es el estimador insesgado de mínima varianza:

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})}{n - p}$$

el sesgo del estimador ML es

$$E(\tilde{\sigma}^2 - \hat{\sigma}^2) = -\sigma^2 \left(\frac{p}{n}\right) \quad (\text{Ejercicio})$$

el cuál es negativo y se incrementa a medida que  $p$  se incrementa.

#### 3.2.4. Máxima verosimilitud restringida (REML)

Ya se mencionó que adicional a los problemas numéricos, los estimadores de las componentes de varianza de ML tienden a ser sesgados. Patterson y Thompson (1971) propusieron un método alternativo llamado *Estimación*

por *Máxima Verosimilitud Restringida (REML)* que también se conoce como *Máxima verosimilitud residual*. Este procedimiento utiliza técnicas de estimación por máxima verosimilitud a una función de verosimilitud asociada a un conjunto de errores de contraste en vez de aquella asociada a las observaciones originales. Este procedimiento compensa la pérdida de grados de libertad que resulta de la estimación de los efectos fijos y produce estimaciones menos sesgadas de las componentes de varianza. El asunto relacionado con este sesgo es importante cuando el número de parámetros no es pequeño en relación al número total de observaciones. El procedimiento REML produce los estimadores estándar del ANOVA en modelos balanceados.

Este procedimiento se basa en la idea de maximizar la log-verosimilitud de  $\mathbf{KY}$  donde  $\mathbf{K}$  es una matriz de constantes, o sea  $\mathbf{KX} = \mathbf{0}$ . Esta condición implica

$$E[\mathbf{KY}] = E[\mathbf{KX}\boldsymbol{\beta} + \mathbf{KZb} + \mathbf{K}\boldsymbol{\epsilon}] = \mathbf{0} + \mathbf{0} + \mathbf{0}$$

Si  $\mathbf{X}_{T \times p}$  es de rango  $r$ , ( $r \leq p$ ), hay  $T - r$  contrastes linealmente independientes. Selecciones  $\mathbf{K} = \mathbf{PM}$  donde

- $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  (También puede formularse como  $\mathbf{M} = \widehat{\mathbf{V}}^{-1} - \widehat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\widehat{\mathbf{V}}^{-1})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}$ , donde  $\widehat{\mathbf{V}}$  es algún estimador de  $\mathbf{V} = Cov(\mathbf{Y})$ . Note que si se transforma los datos originales ortogonales a  $\mathbf{X}$ , es decir,  $\mathbf{U} = \mathbf{M}'\mathbf{Y}$ , entonces  $\mathbf{U}$  tendrá una distribución normal con media  $E(\mathbf{U}) = \mathbf{0}$  y  $Cov(\mathbf{U}) = \mathbf{M}'\mathbf{V}\mathbf{M}$  (esta matriz depende solo del vector de componentes de varianza  $\boldsymbol{\alpha}$  y no del vector  $\boldsymbol{\beta}$ )).
- $\mathbf{M}$  es de dimensión  $T \times T$  y rango  $T - r$
- $\mathbf{P}$  es una matriz de rango  $T - r$

La  $-2(\log\text{verosimilitud})$  de  $\mathbf{KY}$  se vuelve

$$v(\boldsymbol{\theta}|\mathbf{KY}) = \log(|\mathbf{KV}(\boldsymbol{\theta})\mathbf{K}'|) + \mathbf{Y}'\mathbf{K}'[\mathbf{KV}(\boldsymbol{\theta})\mathbf{K}']^{-1}\mathbf{KY} + (T - r)\log(2\pi)$$

**Lema 3.2.3.** Si  $\mathbf{KX} = \mathbf{0}$  con  $\mathbf{V}(\boldsymbol{\theta})$  definida positiva, la siguiente identidad

$$\mathbf{K}'[\mathbf{KV}(\boldsymbol{\theta})\mathbf{K}']^{-1}\mathbf{K} = \mathbf{V}(\boldsymbol{\theta})^{-1} - \mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1} = \mathbf{Q}(\boldsymbol{\theta})$$

se satisface. Recuerde que  $\boldsymbol{\Omega}(\boldsymbol{\theta}) = (\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X})^{-1}$ .

Usando el resultado anterior y ya que  $\widehat{\boldsymbol{\beta}} = \boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{Y}$ , entonces

$$\mathbf{Y}'\mathbf{K}' [\mathbf{K}\mathbf{V}(\boldsymbol{\theta})\mathbf{K}']' \mathbf{K}\mathbf{Y} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Si  $\mathbf{K}_{(T-r) \times T}$  satisface

- $\mathbf{K}'\mathbf{K} = \mathbf{M}$  (Recuerde que  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ )
- $\mathbf{K}\mathbf{K}' = \mathbf{I}$

entonces la -2log-verosimilitud puede escribirse como

$$\begin{aligned} v(\boldsymbol{\theta}|\mathbf{K}\mathbf{Y}) &= \log(|\mathbf{V}(\boldsymbol{\theta})|) + \log(|\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}|) - \log(|\mathbf{X}'\mathbf{X}|) + \mathbf{Y}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{Y} + (T-r)\log(2\pi) \\ &= \log(|\mathbf{V}(\boldsymbol{\theta})|) + \log(|\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}|) - \log(|\mathbf{X}'\mathbf{X}|) \\ &\quad + (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (T-r)\log(2\pi) \quad (**) \end{aligned}$$

Hallar una matriz  $\mathbf{K}$  no es problema: Cualquier matriz de dimensión  $(T-r) \times T$  cuyas filas sean  $T-r$  filas linealmente independientes de  $M$  sirve.

La diferenciación de (\*\*) con respecto a  $\boldsymbol{\theta}$  produce la ecuación estimadora para  $\theta_j$

$$\text{traza} \left[ \mathbf{Q}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right] + \mathbf{Y}' \frac{\partial \mathbf{Q}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{Y} = 0$$

Estas ecuaciones REML producen un estimador para  $\boldsymbol{\theta}$  solamente. Los estimadores para los efectos fijos se estiman mediante mínimos cuadrados generalizados, una vez se tiene  $\hat{\boldsymbol{\theta}}_R$ , como

$$\hat{\boldsymbol{\beta}}_R = \left( \mathbf{X}'\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}}_R)\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}}_R)\mathbf{Y}$$

Note que la matriz  $\mathbf{K}$  no se usa en el proceso de estimación.

### Dos Métodos para Construir $\mathbf{K}$

**Método 1:** Para hallar  $\mathbf{K}$  tal que  $\mathbf{K}'\mathbf{K} = \mathbf{M}$  y  $\mathbf{K}\mathbf{K}' = \mathbf{M}$ , podemos considerar la descomposición espectral de  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . Ya que  $\mathbf{M}$  es simétrica podemos escribir  $\mathbf{M} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}'$  donde

- $\mathbf{Q}$  es una matriz ortogonal y
- $\boldsymbol{\Lambda}$  es una matriz diagonal con los valores propios de  $\mathbf{M}$ .

Ya que  $\mathbf{M}$  es idempotente,  $T - r$  de los valores propios son 1 y el resto son 0. Seleccione  $\mathbf{K}'$  como las  $T - r$  columnas de  $\mathbf{Q}$  que correspondan a los valores propios de  $\mathbf{M}$  no ceros.

**Método 2:** Se debe satisfacer que  $\mathbf{KX} = \mathbf{0}$ . Piense en  $\mathbf{K} = \mathbf{TM}$  donde  $\mathbf{T}$  es de dimensión  $(T - r) \times T$ .

Entonces  $\mathbf{K}'\mathbf{K} = \mathbf{M}\mathbf{T}'\mathbf{T}\mathbf{M}$  implica que  $\mathbf{T}'\mathbf{T}$  debe ser una inversa (generalizada) de  $\mathbf{M}$ , de donde  $\mathbf{TMT}' = \mathbf{I}$ .

Asuma que se particiona  $\mathbf{M}$  como

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}$$

tal que  $\mathbf{M}_{11}$  es el menor no singular de rango  $T - r$ . Esto implica que  $\mathbf{M}_{11}^{-1}$  existe y entonces  $\mathbf{M}_{11}^{-1}$  puede ser factorizada como  $\mathbf{M}_{11}^{-1} = \mathbf{U}'\mathbf{U}$ , donde  $\mathbf{U}$  es una matriz triangular superior.

Note que  $\mathbf{M}_{11}^{-1} = \mathbf{U}^{-1}\mathbf{U}'^{-1}$ . Haga  $\mathbf{T} = [\mathbf{U}, \mathbf{0}]$ , con  $\mathbf{0}$  es una matriz nula de dimensión  $(T - r) \times r$ . Entonces

$$\mathbf{T}'\mathbf{T} = \begin{bmatrix} \mathbf{M}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Esta es la inversa generalizada de  $\mathbf{M}$ . Por lo tanto

$$\mathbf{TMT}' = [\mathbf{U}, \mathbf{0}] \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{U}' \\ \mathbf{0}' \end{bmatrix} = \mathbf{UM}_{11}\mathbf{U}' = \mathbf{UU}^{-1}\mathbf{U}'^{-1}\mathbf{U}' = \mathbf{I}$$

Note que  $\mathbf{U}$  es única (excepto por el signo).

### Otra formulación de la verosimilitud

Harville (1974), muestra que la función log de la verosimilitud para el conjunto de errores de contraste se puede escribir como,

$$\ell(\boldsymbol{\alpha}, \mathbf{M}'\mathbf{Y}) = c - \frac{1}{2} \log \left| \sum_{i=1}^m \mathbf{x}'_i \mathbf{V}_i^{-1} \mathbf{x}_i \right| - \frac{1}{2} \sum_{i=1}^m |\mathbf{V}_i| - \frac{1}{2} \sum_{i=1}^m (\mathbf{Y}_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})$$

donde  $c$  es una constante y  $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$  tal como se definió previamente.

Este resultado establece dos propiedades claves de los estimadores REML:

1. La verosimilitud, y por lo tanto el estimador REML para  $\alpha$ , no depende de la elección de  $\mathbf{M}$ .
2. Si se toma  $\hat{\beta}_{REML} = \hat{\beta}(\alpha_{REML})$ , entonces  $\ell(\alpha, \mathbf{M}'\mathbf{Y})$  se interpreta como el logaritmo de la verosimilitud para  $\theta$ , o simplemente  $\ell_{REML}(\theta)$ .

**Relación entre  $\ell_{REML}(\theta)$  y  $\ell_{ML}(\theta)$**

Como  $\ell_{ML}(\theta) = \log L_{ML}(\theta)$  se puede probar que

$$\ell_{REML}(\theta) = -\frac{1}{2} \log \left| \sum_{i=1}^m \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right| + \ell_{ML}(\theta)$$

De esta forma, se establece otra propiedad: Aún si  $\ell(\alpha, \mathbf{M}'\mathbf{Y})$  es la verdadera verosimilitud para  $\alpha$  NO es una log-verosimilitud para  $\theta$ .

Observe que 2) implica que mientras los estimadores ML para  $\alpha$  y  $\beta$  y el estimador REML para  $\alpha$  satisfacen las propiedades con muestras grandes de los estimadores basados en el principio de verosimilitud (*Consistencia, Normalidad asintótica y Eficiencia*), los estimadores REML para los efectos fijos NO. Por otro lado, para diseños ANOVA mixtos balanceados, se sabe que los estimadores REML de las componentes de varianza son iguales a los estimadores clásicos del ANOVA obtenidos vía el método de los momentos. Esto genera una propiedad o característica de 'robustez' de los estimadores REML contra violaciones del supuesto de normalidad.

**Justificación teórica para los estimadores REML**

De acuerdo a Patterson y Thompson (1971) y Harville y Jeske (1992), en ausencia de información sobre  $\beta$ , no se pierde información sobre  $\alpha$  cuando la inferencia se basa en  $\mathbf{M}'\mathbf{Y}$  en vez de  $\mathbf{Y}$  solamente. Es decir,  $\mathbf{M}'\mathbf{Y}$  es *marginalmente suficiente* para  $\alpha$ .

Desde el punto de vista bayesiano, usar solo contrastes de errores para hacer inferencia sobre  $\alpha$ , es equivalente a ignorar cualquier información apriori sobre  $\beta$  y hacer uso de todos los datos para realizar esas inferencias.

### 3.3. Métodos para ajustar MLM a los datos

#### 3.3.1. Algoritmo EM

El algoritmo EM (expectation-Maximization) fue propuesto en 1977 por Dempster, Laird y Rubin (Dempster et al. 1977). Este método tiende a tener una velocidad de convergencia lenta especialmente cuando los estimadores de las componentes de varianza están cercanos a la frontera del espacio paramétrico. En realidad EM es una técnica para encontrar estimaciones de máxima verosimilitud en presencia de datos incompletos. EM es un procedimiento iterativo que encuentra el EMV del vector  $\boldsymbol{\theta}$  basado en dos pasos llamados '*Esperanza* (*Expectation* en inglés)' o *Paso E* y '*Maximización* (*Maximization* en inglés)' o *Paso M*.

Sea  $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)'$  una matriz  $n \times p$  de datos completos los cuales podrían no haber sido completamente observados, de hecho  $n_o$  es el número de observaciones totalmente observadas mientras que  $n_j$  es el número de datos con valores observados para la variable  $\mathbf{Y}_j$ . Específicamente:

1. *Paso E*. Dado un conjunto de parámetros estimados, tales como el vector de medias y la matriz de varianzas y covarianzas y una distribución normal multivariada, el Paso E calcula la esperanza condicional de la log-verosimilitud para datos completos dados los datos observados y los parámetros estimados.
2. *Paso M*. Dada una log-verosimilitud basada en los datos completos, el Paso M encuentra parámetros estimados que maximizan la log-verosimilitud basada en datos completos del Paso E.

Estos dos pasos se repiten hasta alcanzar convergencia de acuerdo a algún criterio. Para datos normales multivariados, suponga que hay  $G$  grupos con distintos patrones de datos perdidos. Entonces la log-verosimilitud para datos observados que se va a maximizar puede expresarse como:

$$\log L(\boldsymbol{\theta} | \mathbf{Y}_{OBS}) = \sum_{g=1}^G \log L_g(\boldsymbol{\theta} | \mathbf{Y}_{OBS})$$

donde  $\log L_g(\boldsymbol{\theta} | \mathbf{Y}_{OBS})$  es el log de la verosimilitud con los datos observados para el grupo  $g$  y está definida como:

$$\log L_g(\boldsymbol{\theta} | \mathbf{Y}_{OBS}) = -\frac{n_g}{2} \log |\boldsymbol{\Sigma}_g| - \frac{1}{2} \sum_{i,g} (\mathbf{y}_{ig} - \boldsymbol{\mu}_g)' \boldsymbol{\Sigma}_g^{-1} (\mathbf{y}_{ig} - \boldsymbol{\mu}_g)$$

donde  $n_g$  es el número de observaciones del grupo  $g$  y  $\mathbf{y}_{ig}$  es un vector de valores observados correspondientes a variables observadas.

En el caso del MLM, este algoritmo se puede adaptar de la siguiente forma. Considere que los efectos aleatorios  $\mathbf{b}$  son los datos faltantes. Denote por  $\mathbf{W}$  los datos completos, esto es  $\mathbf{W} = (\mathbf{Y}, \mathbf{b})$ . La función de verosimilitud para datos completos está dada por

$$\log(L_W) = \sum_i \log(f_{y_i|b}(y_i | \mathbf{b}, \boldsymbol{\beta}, \phi)) + \log(f_b(\mathbf{b} | \mathbf{D})) \quad (3)$$

Esta selección de datos faltantes tiene dos ventajas:

1. Conociendo los  $\mathbf{b}$ , los  $y_i$ 's son independientes.
2. El paso M del algoritmo EM maximiza (3) con respecto a  $\boldsymbol{\beta}$ ,  $\phi$  y  $\mathbf{D}$ .

Debido a que  $\boldsymbol{\beta}$  y  $\phi$  solo entran en el primer término, el paso M con respecto a  $\boldsymbol{\beta}$  y  $\phi$  solo usa  $f_{y|b}$ .

Maximizar con respecto a  $\mathbf{D}$  es simplemente máxima verosimilitud usando la distribución de  $\mathbf{b}$  después de reemplazar los estadísticos suficientes con sus valores esperados condicionales.

El algoritmo EM toma la siguiente forma:

1. Escoja valores iniciales  $\boldsymbol{\beta}^{(0)}$ ,  $\phi^{(0)}$  y  $\mathbf{D}^{(0)}$ . Haga  $m = 0$ .
2. Calcule (con las esperanzas evaluadas bajo  $\boldsymbol{\beta}^{(m)}$ ,  $\phi^{(m)}$  y  $\mathbf{D}^{(m)}$ .)

a)  $\boldsymbol{\beta}^{(m+1)}$ ,  $\phi^{(m+1)}$

$$E[\log(f_{y|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\beta}, \phi)) | \mathbf{Y}]$$

b)  $\mathbf{D}^{(m+1)}$  que maximiza

$$E[\log(f_b(\mathbf{b} | \mathbf{D})) | \mathbf{Y}]$$

c) Haga  $m = m + 1$

3. Si se obtiene convergencia declare  $\boldsymbol{\beta}^{(m+1)}$ ,  $\phi^{(m+1)}$  y  $\mathbf{D}^{(m+1)}$  como los estimadores de máxima verosimilitud, en otro caso vaya a 2)

### 3.3.2. Algoritmo de Newton-Raphson

La implementación de este método está detallada en Lindstrom y Bates (1988). Este método requiere la maximización de  $L_{ML}(\boldsymbol{\theta})$  o  $L_{REML}(\boldsymbol{\theta})$  sobre un espacio paramétrico del vector  $\boldsymbol{\theta}$  el cual produce matrices semi-definidas positivas  $\mathbf{D}$  y  $\boldsymbol{\Sigma}_i$ . Sin embargo, el modelo marginal requiere que  $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \boldsymbol{\Sigma}_i$  sea semi-definida positiva.

Sean  $\mathbf{U}(\beta)$  el vector de primeras derivadas (también se conoce como vector score o gradiente) y  $\mathbf{I}(\beta)$  la matriz de segundas derivadas (también se conoce como Hessiana). El algoritmo de Newton-Raphson está dado por:

$$\beta_{j+1} = \beta_j - \mathbf{I}^{-1}(\beta) \mathbf{U}(\beta_j)$$

donde  $\mathbf{I}^{-1}$  es la inversa de  $\mathbf{I}$ . Este algoritmo requiere valores iniciales  $\beta_0$

Si estos valores se reemplazan en el lado derecho de la expresión anterior, se obtiene un valor para  $\beta_1$ , es decir,

$$\beta_1 = \beta_0 - \mathbf{I}^{-1}(\beta_0) \mathbf{U}(\beta_0)$$

Luego para obtener  $\beta_2$  se reemplaza el lado derecho de esta ecuación por  $\beta_1$ :

$$\beta_2 = \beta_1 - \mathbf{I}^{-1}(\beta_1) \mathbf{U}(\beta_1)$$

Este procedimiento se repite hasta que la distancia entre la actualización  $j+1$  y  $j$  es muy pequeña, por ejemplo  $|\beta_{j+1} - \beta_j| \leq 10^{-4}$ . En el paso  $j+1$ -ésimo se obtiene una estimación del vector  $\beta$ , llámela  $\hat{\beta}$ . Esta estimación permite obtener la matriz de varianzas y covarianzas:  $-\mathbf{I}^{-1}(\hat{\beta})$ .

Una ventaja del algoritmo de Newton-Raphson es que proporciona una estimación de la Hessiana con la cual se pueden obtener estimaciones de los errores estándar de los estimadores. Es decir, se pueden obtener test de hipótesis e intervalos de confianza asintóticamente normales. Sin embargo, cuando las muestras son pequeñas, estos test e intervalos pueden no ser confiables especialmente para componentes de varianza ya que tienden a ser sesgados a derecha.

Este método requiere que la función objetivo tenga derivadas de primer y segundo orden continuas. Cuando el negativo de la Hessiana es definida positiva, el SAS<sup>©</sup> utiliza este algoritmo (que es más eficiente que el algoritmo EM (Lindstrom y Bates (1988))). En caso de que la Hessiana no sea definida positiva, SAS usa una combinación de Ridge y búsqueda basada en líneas (en

realidad un múltiplo de la identidad se adiciona a la Hessiana para volverla definida positiva).

SAS permite especificar otros métodos de optimización entre los cuales se destaca el Simplex de Nelder-Mead que no requiere de derivadas aunque su convergencia puede ser más lenta que la que se presenta con el algoritmo de Newton-Raphson. Otro método muy común es el llamado Fisher's Scoring que es similar al método de Newton-Raphson pero la Hessiana se reemplaza por su valor esperado  $E(\mathbf{I})$  que frecuentemente tiene un forma más simple que la Hessiana  $\mathbf{I}$ .

### 3.3.3. Estimación por mínimos cuadrados generalizados

Si  $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{R}_i$  es conocida y por lo tanto  $\mathbf{V} = \text{diag}(\mathbf{V}_i)$  es conocida, los efectos fijos son estimados por mínimos cuadrados generalizados (Generalized Least Squares GLS siglas en inglés) como

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}' \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{Y}$$

y las predicciones de los efectos aleatorios se calculan como

$$\hat{\mathbf{b}} = \mathbf{B} \mathbf{Z}' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{GLS})$$

Las estimaciones de  $\boldsymbol{\theta}$  ( $\hat{\boldsymbol{\theta}}_M$ , el estimador máxima verosimilitud o  $\hat{\boldsymbol{\theta}}_R$ , el estimador por máxima verosimilitud restringida) pueden usarse para hallar una estimación consistente de  $\mathbf{V}(\boldsymbol{\theta})$ . En otros casos puede usarse otros métodos diferentes.

Resolviendo  $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{R}_i$  para  $\mathbf{D}$  produce

$$\mathbf{D} = (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i' (\mathbf{V}_i - \mathbf{R}_i) \mathbf{Z}_i (\mathbf{Z}_i' \mathbf{Z}_i)^{-1}$$

Si se reemplaza a  $\mathbf{V}_i$  por la estimada

$$\mathbf{W}_i = (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})'$$

y luego de obtener una estimada de  $\mathbf{R}_i$  promediando los  $n$  clusters para obtener la estimada

$$\hat{\mathbf{D}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i' (\mathbf{W}_i - \mathbf{R}_i) \mathbf{Z}_i (\mathbf{Z}_i' \mathbf{Z}_i)^{-1}$$

Si los errores dentro de los clusters son no correlacionados y homoscedásticos (caso de submuestreo)  $\mathbf{R}_i = \sigma^2 \mathbf{I}$ , entonces

$$\begin{aligned}\hat{\mathbf{D}} &= \frac{1}{n} \sum_{i=1}^n (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i (\mathbf{W}_i - \hat{\sigma}^2 \mathbf{I}) \mathbf{Z}_i (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})' \mathbf{W}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})\end{aligned}$$

Construyendo

$$\hat{\mathbf{V}}_i = \mathbf{Z}_i \hat{\mathbf{D}}_i \mathbf{Z}'_i + \hat{\sigma}^2 \mathbf{I}$$

y

$$\hat{\mathbf{V}} = \text{Diag}(\hat{\mathbf{V}}_i)$$

Los estimadores GLS de  $\boldsymbol{\beta}$  y  $\mathbf{b}$  son

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y}$$

y

$$\hat{\mathbf{b}} = \mathbf{BZ}' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{GLS})$$

#### Otro método no iterativo

Asuma que  $\mathbf{R}_i = \sigma^2 \mathbf{I}$ . Comience estimando estimar los efectos aleatorios en el modelo lineal y luego use estas estimaciones para calcular  $\hat{\mathbf{D}}$ . Reordene el modelo

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i$$

como

$$\mathbf{r}_i = \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i$$

donde  $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ .

Considere el modelo con respuesta  $\mathbf{r}_i$  y vector de parámetros  $\mathbf{b}_i$ .

$$\begin{aligned}\hat{\mathbf{b}}_i &= (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{r}_i \\ &= \mathbf{b}_i + (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \boldsymbol{\epsilon}_i \\ &= \mathbf{b}_i + \delta_i\end{aligned}$$

Es importante anotar que hay que asumir que  $n_i > k$  para poder estimar  $\mathbf{b}_i$ .

Puesto que  $\text{VarCov}(\delta_i) = \sigma^2 (\mathbf{Z}'_i \mathbf{Z}_i)^{-1}$ , esto implica que  $E[\hat{\mathbf{b}}_i \hat{\mathbf{b}}'_i] = \mathbf{D} +$

$\sigma^2 (\mathbf{Z}'_i \mathbf{Z}_i)^{-1}$  y el estimador por el método de los momentos de  $\mathbf{D}$  es

$$\hat{\mathbf{D}} = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i' - \frac{1}{n} \hat{\sigma}^2 \sum_{i=1}^n (\mathbf{Z}'_i \mathbf{Z}_i)^{-1}$$

Con  $n$  clusters hay  $n$  regresiones  $\mathbf{r}_i = \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i$ . Si  $s_i^2$  es el error medio cuadrático estimado en la  $i$ -ésima regresión. Entonces  $\sigma^2$  puede estimarse como

$$\hat{\sigma}^2 = \frac{1}{T - nk} \sum_{i=1}^n (n_i - k) s_i^2$$

con  $T = \sum_{i=1}^n n_i$

Esta aproximación produce una estimación (o más exactamente una predicción) de  $\mathbf{b}_i$  diferente del BLUP (mejor predictor lineal insesgado).

### 3.3.4. ¿Y si la estructura de $\mathbf{R}_i$ es más compleja que $\sigma^2 \mathbf{I}$ ?

Una aproximación a este problema es la siguiente:

Comience con los residuales de mínimos cuadrados ordinarios  $\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}$  donde

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i \mathbf{Y}_i$$

La estimada de los efectos aleatorios mínimo cuadráticos ordinarios

$$\hat{\mathbf{b}}_i = (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}_i (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})$$

El residual específico al cluster se define entonces como

$$\mathbf{u}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\mathbf{b}}_i$$

Una posible función es la pseudo-verosimilitud

$$PL(\boldsymbol{\theta}_i | \mathbf{u}_i) = -\frac{1}{2} \sum_{i=1}^n \{ \log(|\mathbf{R}_i(\boldsymbol{\theta})|) + \mathbf{u}'_i \mathbf{R}_i(\boldsymbol{\theta}) \mathbf{u}_i \}$$

que es maximizada iterativamente.

Finalmente, con  $\mathbf{R}_i$  y  $\mathbf{D}$  estimados, la matriz de varianzas y covarianzas

puede estimarse como

$$\widehat{VarCov}(\mathbf{Y}_i) = \hat{\mathbf{V}}_i = \mathbf{Z}_i \hat{\mathbf{D}} \mathbf{Z}_i' + \hat{\mathbf{R}}_i$$

El estimador GLS de  $\beta$  es

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^n \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{Y}_i = \left( \mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{Y}$$

### 3.4. Criterios de información

Sean  $\ell_A$  y  $\ell_0$  las log-verosimilitudes evaluadas en los parámetros estimados obtenidos bajo las hipótesis alternativa y nula, respectivamente. Además, sean  $s$  y  $p$  el número de parámetros libres bajo la hipótesis nula y alternativa. El test de razón de verosimilitud (Likelihood Ratio Test, LRT en inglés)<sup>2</sup> rechaza la hipótesis nula si  $\ell_A - \ell_0$  es grande en comparación a la diferencia en grados de libertad entre dos modelos que van a ser comparados o equivalentemente si

$$\ell_A - \ell_0 > F(p) - F(s)$$

Para una función apropiada  $F(\cdot)$ . Este procedimiento puede interpretarse como un test formal de significancia solo si el modelo bajo la hipótesis nula está anidado<sup>3</sup>.

#### 3.4.1. El criterio de información de Akaike (AIC)

El AIC NO es una prueba de hipótesis sobre el ajuste de un modelo sino más bien un criterio paramétrico comparativo entre modelos y representa por esto una herramienta para selección de modelos. Dado un conjunto de datos, es posible encontrar varios modelos que se ajustan a ellos. La idea es ranquearlos de acuerdo al AIC.

<sup>2</sup>el LRT está definido como:

$$-2 \log(\lambda_N) = -2 \log \left[ \frac{L_{ML}(\hat{\theta}_{ML,0})}{L_{ML}(\hat{\theta}_{ML})} \right]$$

<sup>3</sup>Un modelo se dice *anidado* dentro de otro modelo si este es un caso particular del segundo. Más precisamente, un modelo A está anidado dentro de un modelo B, si A se puede obtener al imponer restricciones sobre los parámetros del modelo B. Por ejemplo, el modelo de solo intercepto está anidado dentro del modelo full o completo

Se define como

$$AIC = -2 \log Lik + 2k$$

donde  $k$  es el número de parámetros estimados. El modelo asociado al menor AIC se considera mejor entre aquellos que se ajustan. El AIC solo dice cuál de los modelos comparados es el mejor, pero no puede decir cuál es el mejor modelo para explicar los datos. El AIC No requiere que los modelos estén anidados.

### 3.4.2. El criterio de información corregido de Akaike (AICC)

El AICC es una versión corregida del criterio AIC y tampoco es una prueba de hipótesis sobre el ajuste de un modelo sino más bien un criterio paramétrico comparativo entre modelos y representa por esto una herramienta para selección de modelos. Dado un conjunto de datos, es posible encontrar varios modelos que se ajustan a ellos. La idea es ranquearlos de acuerdo al AICC.

Se define como

$$AICC = AIC + \frac{2k(k+1)}{n-k-1}$$

donde  $k$  es el número de parámetros estimados.

### 3.4.3. El criterio de información bayesiano de Schwarz (BIC)

El BIC, al igual que el AIC, es un criterio paramétrico para selección de modelos. El BIC No requiere que los modelos estén anidados. El modelo asociado al menor BIC se considera mejor que los otros. El BIC solo dice cuál de los modelos comparados es el mejor, pero no puede decir cuál es el mejor modelo para explicar los datos. Se define como:

$$BIC = -2 \log Lik + k \log n$$

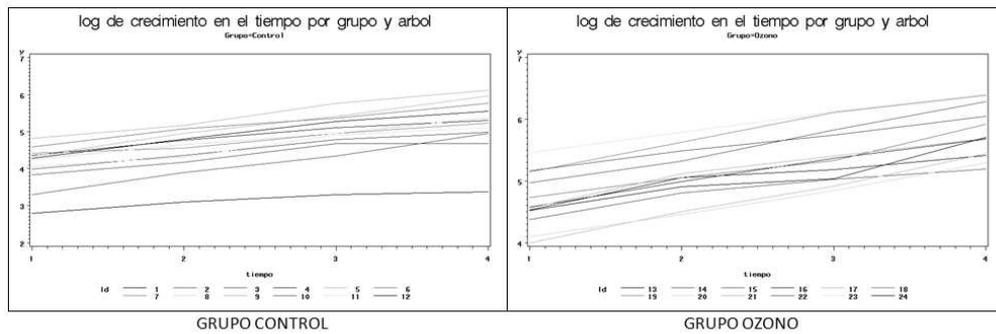
donde  $k$  número de parámetros estimados y  $n$  el número de observaciones que no son “missing” y que fueron usadas en el proceso de estimación.

## 3.5. Ejemplos

**Ejemplo 3.5.1.** *Considere el siguiente conjunto de datos acerca de logaritmos del crecimiento de árboles de abeto*

Obs	Grupo	Id	y <sub>152</sub>	y <sub>174</sub>	y <sub>201</sub>	y <sub>227</sub>
1	Control	1	2.79	3.10	3.30	3.38
2	Control	2	3.30	3.90	4.34	4.96
3	Control	3	3.98	4.36	4.79	4.99
4	Control	4	4.36	4.77	5.10	5.30
5	Control	5	4.34	4.95	5.42	5.97
6	Control	6	4.59	5.08	5.36	5.76
7	Control	7	4.41	4.56	4.95	5.23
8	Control	8	4.24	4.64	4.95	5.38
9	Control	9	4.82	5.17	5.76	6.12
10	Control	10	3.84	4.17	4.67	4.67
11	Control	11	4.07	4.31	4.90	5.10
12	Control	12	4.28	4.80	5.27	5.55
13	Ozono	13	4.53	5.05	5.18	5.41
14	Ozono	14	4.97	5.32	5.83	6.29
15	Ozono	15	4.37	4.81	5.03	5.19
16	Ozono	16	4.58	4.99	5.37	5.68
17	Ozono	17	4.00	4.50	4.92	5.44
18	Ozono	18	4.73	5.05	5.33	5.92
19	Ozono	19	5.15	5.63	6.11	6.39
20	Ozono	20	4.10	4.46	4.84	5.29
21	Ozono	21	4.56	5.12	5.40	5.69
22	Ozono	22	5.16	5.49	5.74	6.05
23	Ozono	23	5.46	5.79	6.12	6.41
24	Ozono	24	4.52	4.91	5.04	5.71

Observe que en este conjunto de datos, se tiene el mismo número de árboles por grupo (12) y además se registraron 4 medidas de crecimiento a los días 152, 174, 201 y 227, respectivamente. Es decir, cada árbol contribuye con 4 mediciones. Los perfiles aparecen en el gráfico de abajo. El objetivo del estudio es comparar el crecimiento de los árboles en dos ambientes diferentes: Ambiente de control y Ambiente rico en ozono. Ajuste un MLM a estos datos para comparar el crecimiento en estos dos ambientes.



**Solución.** *Se implementarán dos modelos distintos pero equivalentes: 1) Un modelo de interceptos aleatorios con estructura de covarianza especificada como componentes de varianza y 2) un modelo de medidas repetidas con estructura de covarianza especificada como simetría compuesta (Modelo de componentes de varianza con las mismas distribuciones marginales).*

**Modelo 1. modelo de interceptos aleatorios con estructura de covarianza especificada como componentes de varianza. El código en SAS:**

```

data abetos;
input Obs Grupo$ Id y152 y174 y201 y227;
cards;
1 Control 1 2.79 3.10 3.30 3.38
2 Control 2 3.30 3.90 4.34 4.96
3 Control 3 3.98 4.36 4.79 4.99
4 Control 4 4.36 4.77 5.10 5.30
5 Control 5 4.34 4.95 5.42 5.97
6 Control 6 4.59 5.08 5.36 5.76
7 Control 7 4.41 4.56 4.95 5.23
8 Control 8 4.24 4.64 4.95 5.38
9 Control 9 4.82 5.17 5.76 6.12
10 Control 10 3.84 4.17 4.67 4.67
11 Control 11 4.07 4.31 4.90 5.10
12 Control 12 4.28 4.80 5.27 5.55
13 Ozono 13 4.53 5.05 5.18 5.41
14 Ozono 14 4.97 5.32 5.83 6.29
15 Ozono 15 4.37 4.81 5.03 5.19
16 Ozono 16 4.58 4.99 5.37 5.68
17 Ozono 17 4.00 4.50 4.92 5.44
18 Ozono 18 4.73 5.05 5.33 5.92
19 Ozono 19 5.15 5.63 6.11 6.39
20 Ozono 20 4.10 4.46 4.84 5.29
21 Ozono 21 4.56 5.12 5.40 5.69
22 Ozono 22 5.16 5.49 5.74 6.05
23 Ozono 23 5.46 5.79 6.12 6.41
24 Ozono 24 4.52 4.91 5.04 5.71
;
run;
data abetos1;
set abetos;
if tiempo=1 then tiempo=152;
if tiempo=2 then tiempo=174;
if tiempo=3 then tiempo=201;
if tiempo=4 then tiempo=227;
array xxx{4} y152 y174 y201 y227;
do i=1 to 4;
y=xxx{i};
tiempo=i;
output;
end;
run;
proc mixed data=abetos1;
class grupo id;
model y=grupo tiempo / noint solution ddfm=betwithin;

```

```
random intercept / subject=id v vcorr;
lsmeans grupo / pdiff;
title 'Modelo de interceptos alaetorios para los datos de abetos';
title2 'Asumiendo Componentes de Varianza';
run;
```

En el primer paso DATA, se leen los datos y se nombran las variables de acuerdo al formato original de los datos. Esta tabla SAS (o base de datos en formato SAS) se denomina abetos. En el segundo paso DATA, se crea otra base de datos SAS llamada abetos1 que tiene una estructura como la que se muestra a continuación (solo se muestran algunos registros):

Obs	Obs	Grupo	Id	y152	y174	y201	y227	tiempo	i	y
1	1	Control	1	2.79	3.10	3.30	3.38	1	1	2.79
2	1	Control	1	2.79	3.10	3.30	3.38	2	2	3.10
3	1	Control	1	2.79	3.10	3.30	3.38	3	3	3.30
4	1	Control	1	2.79	3.10	3.30	3.38	4	4	3.38
5	2	Control	2	3.30	3.90	4.34	4.96	1	1	3.30
6	2	Control	2	3.30	3.90	4.34	4.96	2	2	3.90
7	2	Control	2	3.30	3.90	4.34	4.96	3	3	4.34
8	2	Control	2	3.30	3.90	4.34	4.96	4	4	4.96
.	.	.	.	.	.	.	.	.	.	.
93	24	Ozono	24	4.52	4.91	5.04	5.71	1	1	4.52
94	24	Ozono	24	4.52	4.91	5.04	5.71	2	2	4.91
95	24	Ozono	24	4.52	4.91	5.04	5.71	3	3	5.04
96	24	Ozono	24	4.52	4.91	5.04	5.71	4	4	5.71

Los resultados que produce este programa son:

```
Model Information

Data Set                WORK.ABETOS1
Dependent Variable      y
Covariance Structure    Variance Components
Subject Effect          Id
Estimation Method       REML
Residual Variance Method Profile
Fixed Effects SE Method Model-Based
Degrees of Freedom Method Between-Within
```

```
Class Level Information

Class  Levels  Values
Grupo      2      Control Ozono
Id         24      1 2 3 4 5 6 7 8 9 10 11 12 13
          14 15 16 17 18 19 20 21 22 23
          24
```

```
Dimensions

Covariance Parameters      2
Columns in X                3
```

Columns in Z Per Subject	1
Subjects	24
Max Obs Per Subject	4

Number of Observations

Number of Observations Read	96
Number of Observations Used	96
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	155.39444274	
1	1	-14.21311873	0.00000000

Convergence criteria met.

Estimated V Matrix for Id 1

Row	Col1	Col2	Col3	Col4
1	0.2866	0.2704	0.2704	0.2704
2	0.2704	0.2866	0.2704	0.2704
3	0.2704	0.2704	0.2866	0.2704
4	0.2704	0.2704	0.2704	0.2866

Estimated V Correlation Matrix for Id 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.9435	0.9435	0.9435
2	0.9435	1.0000	0.9435	0.9435
3	0.9435	0.9435	1.0000	0.9435
4	0.9435	0.9435	0.9435	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	Id	0.2704
Residual		0.01620

Fit Statistics

-2 Res Log Likelihood	-14.2
AIC (smaller is better)	-10.2
AICC (smaller is better)	-10.1
BIC (smaller is better)	-7.9

Solution for Fixed Effects

Effect	Grupo	Estimate	Standard Error	DF	t Value	Pr >  t
Grupo	Control	3.7408	0.1540	22	24.29	<.0001
Grupo	Ozono	4.3154	0.1540	22	28.02	<.0001
tiempo		0.3708	0.01162	71	31.91	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Grupo	2	22	664.30	<.0001
tiempo	1	71	1018.25	<.0001

Least Squares Means

Effect	Grupo	Estimate	Standard Error	DF	t Value	Pr >  t
Grupo	Control	4.6677	0.1512	22	30.86	<.0001
Grupo	Ozono	5.2423	0.1512	22	34.66	<.0001

Differences of Least Squares Means

Effect	Grupo	_Grupo	Estimate	Standard Error	DF	t Value	Pr >  t
Grupo	Control	Ozono	-0.5746	0.2139	22	-2.69	0.0135

Ahora se procederá a explicar cada elemento de esta salida. La salida etiquetada **Model Information** contiene algunos elementos: **Data** especifica la base de datos SAS que se usó para implementar el modelo; **Dependent Variable** especifica la variable dependiente; **Covarianza Structure** especifica la estructura de covarianza del modelo, en este caso el “default” es “componentes de varianza”; **Subject Effect** El sujeto se especifica como un efecto; **Estimation Method** especifica el método de estimación que por default es *REML*; **Residual Variance Method** especifica que el método para estimar la varianza residual está basado en la verosimilitud perfilada; **Fixed Effects SE Method** especifica que el método basado para obtener los errores estándar de las estimaciones de los efectos fijos se basa en el modelo y finalmente **Degrees of Freedom Method** especifica que el método para obtener los grados de libertad de la prueba de significancia de los efectos fijos debe tener en cuenta la variabilidad intra y entre sujetos (en el programa se especifica como  $DDFM=BETWHITIN$ , otros métodos están disponibles en SAS).

La salida etiquetada **Class Level Information** contiene información de los niveles de las variables declaradas como categóricas en el modelo. Esta información incluye el número de niveles y los nombres asignados a cada

nivel.

La salida etiquetada **Dimensions** muestra elementos relacionados con las dimensiones de la matriz de varianzas y covarianzas, la matriz de diseño, la matriz de efectos aleatorios (que en este caso es de dimensión 1 ya que se declaró un modelo de interceptos aleatorios); también muestra el número de sujetos y el número máximo de medidas repetidas por sujeto.

La salida etiquetada **Number of Observations** proporciona información sobre el número de observaciones leídas de la base de datos y el número real de observaciones usadas por el procedimiento MIXED. Esta información es útil ya que cuando el número de observaciones no usadas es distinto de cero es porque hay registros con valores perdidos que deben ser revisados.

La salida etiquetada **Iteration History** proporciona información sobre el proceso de iteración. Cuando aparece un mensaje distinto a **Convergence criteria met** es porque se presentaron problemas de convergencia que ocasionan problemas con las estimaciones.

Las salidas etiquetadas **Estimated V Matrix for Id 1** y **Estimated V Correlation Matrix for Id 1** proporciona los bloques de la estimación de la matriz  $\mathbf{V}$  y los bloques de la matriz de correlación correspondiente a los bloques de la matriz estimada  $\hat{\mathbf{V}}$ . Observe que por defecto solo se muestra las matrices para el primer sujeto pero si se quiere por ejemplo obtener estas matrices para otro sujeto o grupo de sujetos se puede usar en el programa la opción  $V=2$  (para el sujeto 2) o  $V=5,7,14$  (para los sujetos 5, 7 y 14). Recuerde que  $\mathbf{V} = \mathbf{ZDZ}' + \Sigma$

La salida etiquetada **Covariance Parameter Estimates** proporciona las estimaciones de las componentes de varianza de los elementos aleatorios del modelo que en este caso son el intercepto individual y el término de error. En realidad,  $\hat{\sigma}_b^2 = 0,2704$  y  $\hat{\sigma}^2 = 0,01620$ . Con estos valores se puede calcular el coeficiente de correlación intraclase.

La salida etiquetada **Fit Statistics** proporciona los valores de algunos estadísticos relacionadas con el modelo estimado. El primer valor de la lista corresponde a -2 veces el valor máximo del logaritmo de la verosimilitud restringida ( $\ell$ ). Aún si un modelo está 'anidado' dentro de otro este valor no se recomienda que se usa para seleccionar un modelo. El segundo valor de la lista corresponde al AIC (Akaike Information Criterion) que se define como  $-2\ell + 2d$  donde  $\ell$  corresponde al máximo valor de la verosimilitud restringida y  $d$  al número de parámetros estimados en el modelo (o dimensión del modelo). El tercer valor de la lista es el AICC o Akaike Information Criterion Corrected (versión del AIC corregida para muestras pequeñas) que se define

como  $-2\ell + \frac{2dn}{n-d-1}$  donde  $n$  es el número de datos. Finalmente, el cuarto miembro de esta lista corresponde al BIC (Bayesian Information Criterion de Schwarz) que se define como  $-2\ell + d\log(n)$ . Tanto el AIC, AICC como el BIC se pueden usar para seleccionar un modelo.

La salida etiquetada **Solution for Fixed Effects** se produce al especificar en el MODEL STATEMENT la opción SOLUTION y tal como su nombre lo indica produce las estimaciones de REML para cada nivel de las variables consideradas como efectos fijos (si estas aparecen en el CLASS STATEMENT) o la estimación del efecto de la covariable declarada como continua (es decir que no se especifica en el CLASS STATEMENT), que en este caso corresponden a los efectos de GRUPO (categórica) y de TIEMPO (continua). El valor del intercepto para el Grupo Control 3.7408 y para el grupo Ozono es 4.3154. El valor de  $t$  Value no corresponde exactamente al de una distribución  $t$  sino que es más bien una pseudo- $t$  al tratarse del cociente entre algo normal y la raíz de algo que no es chi-cuadrado.

La salida etiquetada **Type 3 Tests of Fixed Effects** corresponde a los resultados de la prueba de significancia de cada efecto fijo del modelo. A diferencia de las pruebas anteriores, cada variable se considera globalmente y no de acuerdo a sus niveles. Se observa que tanto el efecto de GRUPO como del TIEMPO son estadísticamente significativos ya que su valor- $p < 0.05$ .

Las salidas etiquetadas **Least Squares Means** y **Differences of Least Squares Means** calculan y comparan medias de mínimos cuadrados de los efectos fijos especificados en la opción LSMEANS. Las LSMEANS son combinaciones lineales de los parámetros estimados de tal manera que correspondan a valores promedios predichos en una población donde los niveles de clasificación están balanceados (son promedios corregidos por el modelo). Permite llevar a cabo comparaciones múltiples y buscar por diferencias significativas entre los niveles de las variables categóricas que representan los efectos fijos. En este caso solo se solicitaron las LSMEANS de la variable GRUPO con sus respectivas comparaciones múltiples. Se observa una diferencia estadísticamente significativa entre los ambientes 'Control' y 'Ozono' siendo mayor el crecimiento de los árboles en ambientes ricos en ozono (pues el valor de la diferencia es negativo, -0.5746). Este valor de -0.5746 se puede obtener de dos formas:  $3.7408 - 4.3154 = -0.5746$  y  $4.6677 - 5.2423 = -0.5746$ .

**Modelo 2. Un modelo de medidas repetidas con estructura de covarianza especificada también como simetría compuesta (Modelo de componentes de varianza con las mismas distribuciones marginales)** El código en SAS:

```

data abetos;
input Obs Grupo$ Id y152 y174 y201 y227;
cards;
1 Control 1 2.79 3.10 3.30 3.38
2 Control 2 3.30 3.90 4.34 4.96
3 Control 3 3.98 4.36 4.79 4.99
4 Control 4 4.36 4.77 5.10 5.30
5 Control 5 4.34 4.95 5.42 5.97
6 Control 6 4.59 5.08 5.36 5.76
7 Control 7 4.41 4.56 4.95 5.23
8 Control 8 4.24 4.64 4.95 5.38
9 Control 9 4.82 5.17 5.76 6.12
10 Control 10 3.84 4.17 4.67 4.67
11 Control 11 4.07 4.31 4.90 5.10
12 Control 12 4.28 4.80 5.27 5.55
13 Ozono 13 4.53 5.05 5.18 5.41
14 Ozono 14 4.97 5.32 5.83 6.29
15 Ozono 15 4.37 4.81 5.03 5.19
16 Ozono 16 4.58 4.99 5.37 5.68
17 Ozono 17 4.00 4.50 4.92 5.44
18 Ozono 18 4.73 5.05 5.33 5.92
19 Ozono 19 5.15 5.63 6.11 6.39
20 Ozono 20 4.10 4.46 4.84 5.29
21 Ozono 21 4.56 5.12 5.40 5.69
22 Ozono 22 5.16 5.49 5.74 6.05
23 Ozono 23 5.46 5.79 6.12 6.41
24 Ozono 24 4.52 4.91 5.04 5.71
;
run;
data abetos1;
set abetos;
if tiempo=1 then tiempo=152;
if tiempo=2 then tiempo=174;
if tiempo=3 then tiempo=201;
if tiempo=4 then tiempo=227;
array xxxx{4} y152 y174 y201 y227;
do i=1 to 4;
y=xxxx{i};
tiempo=i;
output;
end;
run;
proc mixed data=abetos1;
class grupo id;
model y=grupo tiempo / solution;
repeated / subject=id type=cs;
lsmeans grupo / pdiff;
title 'Modelo de componentes de varianza con las mismas marginales';
title2 'Asumiendo Simetría Compuesta';
run;

```

Los resultados que produce este programa son:

The Mixed Procedure

Model Information

```
Data Set                WORK.ABETOS1
Dependent Variable      y
Covariance Structure    Compound Symmetry
Subject Effect          Id
Estimation Method       REML
Residual Variance Method Profile
Fixed Effects SE Method Model-Based
Degrees of Freedom Method Between-Within
```

Class Level Information

```
Class  Levels  Values
Grupo      2  Control Ozono
Id         24  1 2 3 4 5 6 7 8 9 10 11 12 13
                          14 15 16 17 18 19 20 21 22 23
                          24
```

Dimensions

```
Covariance Parameters      2
Columns in X                4
Columns in Z                0
Subjects                    24
Max Obs Per Subject        4
```

Number of Observations

```
Number of Observations Read      96
Number of Observations Used      96
Number of Observations Not Used   0
```

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	155.39444274	
1	1	-14.21311873	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	Id	0.2704
Residual		0.01620

Fit Statistics

```
-2 Res Log Likelihood      -14.2
AIC (smaller is better)    -10.2
AICC (smaller is better)   -10.1
```



tes ricos en ozono (pues el valor de la diferencia es negativo,  $-0.5746$ ). Este valor de  $-0.5746$  se puede obtener de dos formas:  $3.7408 - 4.3154 = -0.5746$  y  $4.6677 - 5.2423 = -0.5746$ .

**Ejemplo 3.5.2.** *De un cargamento de lana se toman dos piezas de cada una de 10 madejas seleccionadas al azar (Youden (1964)). A cada una de estas piezas de cada madeja se les mide el porcentaje de lana pura. Se piensa que el error en la estimación cuantitativa del contenido de la lana pura es pequeño en comparación con la variación muestral entre las piezas. Los datos son:*

Madeja	Porcentaje de lana pura
1	55.51
1	54.32
2	55.63
2	55.58
3	62.13
3	56.51
4	63.56
4	59.84
5	61.57
5	62.68
6	54.68
6	57.64
7	58.05
7	59.70
8	58.59
8	56.70
9	54.94
9	55.30
10	58.25
10	56.61

Usando un MLM responda la siguiente pregunta: ¿Hay evidencia en los datos que apoyen esta creencia de la existencia de una diferencia en las varianzas de las dos piezas?

**Solución.** *Para responder esta pregunta se usará un modelo lineal mixto con interceptos aleatorios (usando SAS y R) que permita estimar las componentes de varianza.*

**Usando SAS:**

```

data lana;
input madeja porcentaje_lana;
cards;
1 55.51
1 54.32
2 55.63
2 55.58
3 62.13
3 56.51
4 63.56
4 59.84
5 61.57
5 62.68
6 54.68
6 57.64
7 58.05
7 59.70
8 58.59
8 56.70
9 54.94
9 55.30
10 58.25
10 56.61
;
proc mixed;
model porcentaje_lana= / solution ddfm=contain;
random intercept / subject=madeja;
run;
quit;

```

*Al ejecutar este programa se obtiene*

The Mixed Procedure

Model Information

Data Set	WORK.LANA
Dependent Variable	porcentaje_lana
Covariance Structure	Variance Components
Subject Effect	madeja
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Dimensions

Covariance Parameters	2
Columns in X	1
Columns in Z Per Subject	1
Subjects	10
Max Obs Per Subject	2

Number of Observations

```
Number of Observations Read      20
Number of Observations Used      20
Number of Observations Not Used   0
```

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	96.66401035	
1	1	92.23035726	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	madeja	5.0708
Residual		3.2974

Fit Statistics

```
-2 Res Log Likelihood      92.2
AIC (smaller is better)   96.2
```

The Mixed Procedure

Fit Statistics

```
AICC (smaller is better)   97.0
BIC (smaller is better)   96.8
```

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	57.8895	0.8197	9	70.62	<.0001

*De esta salida solo interesa la que está etiquetada como **Covariance Parameter Estimates** donde  $\hat{\sigma}_b^2=5.0708$  (varianza entre las piezas de lana) y  $\hat{\sigma}^2=3.2974$  (varianza entre madejas). Puede verse que el error en la estimación cuantitativa del contenido de la lana pura es menor a la variación muestral entre las piezas ( $3.2974 < 5.0708$ )*

**Usando R:**

```
# Modelo Youden, pp. 38
library(nlme)
y<-scan()
5551 5432 5563 5558 6213 5651 6356 5984 6157 6268 5468 5764
5805 5970 5859 5670 5494 5530 5825 5661
```

```

y<-y/100
inter<-rep(1,length(y))
individuo<-rep(1:10,each=2)
lana<-data.frame(y=y,inter=inter,individuo=individuo)
lana.new <- groupedData( y ~ inter | individuo,
                        data = as.data.frame( lana ) )
modelo<-lme(y~1,random=~1,data=lana.new)
summary(modelo)

```

*Se obtiene*

```

Linear mixed-effects model fit by REML
Data: lana.new
      AIC      BIC    logLik
 98.23036 101.0637 -46.11518
Random effects:
Formula: ~1 | individuo
      (Intercept) Residual
StdDev:    2.251844 1.815887

Fixed effects: y ~ 1
              Value Std.Error DF   t-value p-value
(Intercept)  57.8895  0.8197271  10  70.62046     0
Standardized Within-Group Residuals:
      Min         Q1         Med         Q3         Max
-1.3541638 -0.5104775 -0.2850025  0.5108621  1.7407427

Number of Observations: 20
Number of Groups: 10

```

*Los resultados obtenidos con ambos paquetes, en relación a las componentes de varianza, coinciden, solo que el R genera la raíz cuadrada de  $\hat{\sigma}_b^2$  y  $\hat{\sigma}^2$ . La estimación de los efectos fijos obtenida con ambos programas es la misma. La relación entre las varianzas sigue siendo la misma:  $\hat{\sigma}_b^2 > \hat{\sigma}^2$ .*

## 3.6. Aproximación bayesiana

### 3.6.1. Método Bayes empírico

Los efectos aleatorios  $\mathbf{b}_i$  reflejan cómo la evolución del sujeto  $i$  se desvía de la evolución esperada  $\mathbf{X}_i\boldsymbol{\beta}$ . Otro uso importante de las estimaciones de  $\mathbf{b}_i$  es que pueden usarse para detectar perfiles atípicos (outliers). Es de resaltar que esta estimación tiene sentido bajo el modelo condicional (ver siguiente sección):

$$\begin{aligned} \mathbf{Y}_i | \mathbf{b}_i &\sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \boldsymbol{\Sigma}_i) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \end{aligned}$$

En vista de que las  $\mathbf{b}_i$  son variables aleatorias, parece natural usar métodos bayesianos.

### Terminología

*Distribución a priori*  $N(\mathbf{0}, \mathbf{D})$  para  $\mathbf{b}_i$ .

*Distribución posterior*

$$\begin{aligned}
 f(\mathbf{b}_i|\mathbf{y}_i) &\equiv f(\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i) \\
 &= \frac{f(\mathbf{y}_i|\mathbf{b}_i) f(\mathbf{b}_i)}{\int f(\mathbf{y}_i|\mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i} \\
 &\propto f(\mathbf{y}_i|\mathbf{b}_i) f(\mathbf{b}_i) \\
 &\propto \dot{\quad} \text{(Después de algunos cálculos complicados)} \\
 &\propto e^{-\frac{1}{2}(\mathbf{b}_i - \mathbf{DZ}'_i\mathbf{W}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}))' \boldsymbol{\Lambda}_i^{-1}(\mathbf{b}_i - \mathbf{DZ}'_i\mathbf{W}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}))}
 \end{aligned}$$

Para alguna matriz  $\boldsymbol{\Lambda}_i$  definida positiva<sup>4</sup>. Es decir,

$$\mathbf{b}_i|\mathbf{y}_i \sim N(\mathbf{DZ}'_i\mathbf{W}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \boldsymbol{\Lambda}_i)$$

La media posterior estimada para  $\mathbf{b}_i$  está dada por:

$$\begin{aligned}
 \widehat{\mathbf{b}}_i(\boldsymbol{\theta}) &= E[\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i] \\
 &= \int \mathbf{b}_i f(\mathbf{b}_i|\mathbf{y}_i) d\mathbf{b}_i \\
 &= \mathbf{DZ}'_i\mathbf{W}_i(\boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})
 \end{aligned}$$

$\widehat{\mathbf{b}}_i(\boldsymbol{\theta})$  se distribuye normalmente con matriz de covarianzas:

$$\text{Var}(\widehat{\mathbf{b}}_i(\boldsymbol{\theta})) = \mathbf{DZ}'_i \left\{ \mathbf{W}_i - \mathbf{W}_i\mathbf{X}_i \left( \sum_{i=1}^N \mathbf{X}'_i\mathbf{W}_i\mathbf{X}_i \right)^{-1} \mathbf{X}'_i\mathbf{W}_i \right\} \mathbf{Z}_i\mathbf{D}$$

Note que la inferencia para  $\mathbf{b}_i$  debe tener en cuenta la variabilidad de  $\mathbf{b}_i$ . Por lo tanto, la inferencia para  $\mathbf{b}_i$  usualmente se basa en:

$$\text{Var}(\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) - \mathbf{b}_i) = \mathbf{D} - \text{Var}(\widehat{\mathbf{b}}_i(\boldsymbol{\theta}))$$

---

<sup>4</sup>Lindley y Smith (1972) y Smith (1973) proporcionan resultados que permiten deducir esta distribución posterior

Se pueden obtener test tipo Wald. Los parámetros especificados en el vector  $\boldsymbol{\theta}$  se reemplazan por sus estimadores ML o REML obtenidos al ajustar el modelo. A  $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\boldsymbol{\theta})$  se le conoce como *estimador empírico de Bayes* para  $\mathbf{b}_i$ . Los test t y F aproximados que tienen en cuenta la variabilidad introducida al reemplazar  $\boldsymbol{\theta}$  por  $\widehat{\boldsymbol{\theta}}$  son similares a los test de efectos fijos.

### 3.6.2. Estimadores shrinkage

Estos estimadores también se conocen como estimadores de contracción o encogimiento. Considere el problema de la predicción de la evolución del sujeto  $i$ :

$$\begin{aligned}\widehat{\mathbf{Y}}_i &= \mathbf{X}_i \widehat{\boldsymbol{\beta}} + \mathbf{Z}_i \widehat{\mathbf{b}}_i \\ &= \mathbf{X}_i \widehat{\boldsymbol{\beta}} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}) \\ &= (\mathbf{I}_{n_i} - \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i \mathbf{V}_i^{-1}) \mathbf{X}_i \widehat{\boldsymbol{\beta}} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i \mathbf{V}_i^{-1} \mathbf{y}_i \\ &= \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1} \mathbf{X}_i \widehat{\boldsymbol{\beta}} + (\mathbf{I}_{n_i} - \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1}) \mathbf{y}_i\end{aligned}$$

Con lo que  $\widehat{\mathbf{Y}}_i$  es un promedio ponderado del perfil poblacional promedio  $\mathbf{X}_i \widehat{\boldsymbol{\beta}}$  y los datos observados  $\mathbf{y}_i$  con pesos  $\boldsymbol{\Sigma}_i \mathbf{V}_i^{-1}$  y  $\mathbf{I}_{n_i} - \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1}$  respectivamente (los cuales se pueden reemplazar por sus estimaciones  $\widehat{\boldsymbol{\Sigma}}_i \widehat{\mathbf{V}}_i^{-1}$  y  $\mathbf{I}_{n_i} - \widehat{\boldsymbol{\Sigma}}_i \widehat{\mathbf{V}}_i^{-1}$  respectivamente). Observe que  $\mathbf{X}_i \widehat{\boldsymbol{\beta}}$  recibe un peso mayor si la varianza residual es grande en comparación con la variabilidad total. Este fenómeno se conoce como *Shrinkage* (encogimiento, contracción) ya que los datos se ‘contraen’ hacia el perfil promedio a priori  $\mathbf{X}_i \widehat{\boldsymbol{\beta}}$ . Esto también se refleja en el hecho de que para cualquier combinación lineal de efectos aleatorios  $\boldsymbol{\lambda}' \mathbf{b}_i$  se tiene que

$$\text{Var}(\boldsymbol{\lambda}' \widehat{\mathbf{b}}_i) \leq \text{Var}(\boldsymbol{\lambda}' \mathbf{b}_i)$$

Aquí, el vector de escalares  $\boldsymbol{\lambda}$  define la combinación lineal.

Las varianzas aquí comparadas son de distinta naturaleza: Una se relaciona con el modelo (el miembro derecho de la desigualdad) y la otra se relaciona con el proceso de predicción.

### 3.6.3. Chequeo de normalidad de los efectos aleatorios

En la práctica, histogramas a partir de estimaciones de Bayes empíricas se usan con frecuencia para chequear el supuesto de normalidad para los efectos

aleatorios. Sin embargo, puesto que

$$\widehat{\mathbf{b}}_i = \mathbf{D}\mathbf{Z}'_i\mathbf{W}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$$

Entonces

$$\text{Var}(\widehat{\mathbf{b}}_i) = \mathbf{D}\mathbf{Z}'_i \left\{ \mathbf{W}_i - \mathbf{W}_i\mathbf{X}_i \left( \sum_{i=1}^N \mathbf{X}'_i\mathbf{W}_i\mathbf{X}_i \right)^{-1} \mathbf{X}'_i\mathbf{W}_i \right\} \mathbf{Z}_i\mathbf{D}$$

y por lo tanto los estimadores Bayes empíricos se deben estandarizar antes de usarse. Además, debido a la propiedad de ‘Shrinkage’, los estimadores Bayes empíricos no reflejan completamente la heterogeneidad en los datos. De hecho, cuando el ‘shrinkage’ es muy severo esto ocasiona que los estimadores  $\widehat{\mathbf{b}}_i$  satisfagan el supuesto de normalidad. Se puede concluir que si los estimadores Bayes empíricos se obtienen bajo normalidad no pueden usarse para chequear normalidad. Esto sugiere que una alternativa efectiva para chequear normalidad consiste en ajustar un modelo más general, que incluya el modelo lineal clásico mixto como caso especial, y comparar ambos modelos usando métodos basados en pruebas de razón de verosimilitud.

### 3.6.4. Consideraciones de inferencia

Dado un vector aleatorio  $\mathbf{Y}$  con f.d.p.  $f(\mathbf{Y}|\boldsymbol{\theta})$  que depende de una vector de parámetros  $\boldsymbol{\theta}$  sobre el cual se tiene una distribución a priori  $\pi(\boldsymbol{\theta})$ . Las inferencias con respecto a  $\boldsymbol{\theta}$  están basadas en la distribución posterior de  $\boldsymbol{\theta}$  condicionadas en los datos:

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

Una estimación puntual de  $\boldsymbol{\theta}$  puede calcularse como la media o la moda de la distribución posterior.

En este caso se tienen  $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$  bajo el modelo

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

**Modelo jerárquico bayesiano**

- **Etapa 1:** (Variación Intraindividuos o intra-individual)

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

donde  $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$

- **Etapa 2:** (Variación Interindividuo o entre-individuos)

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

- **Etapa 3:** (Distribución hiperapriori)

$$\begin{aligned}\boldsymbol{\beta} &\sim N(\boldsymbol{\beta}^*, \mathbf{H}) \\ \mathbf{D}^{-1} &\sim \text{Wishart} \\ \mathbf{R}_i^{-1} &\sim \text{Wishart}\end{aligned}$$

**Estimación de los efectos asumiendo que la estructura de covarianzas es conocida**

1. **Estimación de  $\boldsymbol{\beta}$ .** La distribución posterior de  $\boldsymbol{\beta}$  está dada por

$$\pi(\boldsymbol{\beta} | \mathbf{Y}) = \frac{\int f(\mathbf{Y} | \boldsymbol{\beta}, \mathbf{b}) \pi(\boldsymbol{\beta}) \pi(\mathbf{b}) d\mathbf{b}}{\int \int f(\mathbf{Y} | \boldsymbol{\beta}, \mathbf{b}) \pi(\boldsymbol{\beta}) \pi(\mathbf{b}) d\mathbf{b} d\boldsymbol{\beta}}$$

donde

$$\begin{aligned}\mathbf{Y} | \boldsymbol{\beta}, \mathbf{b} &\sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{R}) \\ \boldsymbol{\beta} &\sim N(\boldsymbol{\beta}^*, \mathbf{H}) \\ \mathbf{b} &\sim N(\mathbf{0}, \mathbf{D})\end{aligned}$$

Bajo los supuestos de jerarquía

$$\begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \\ \mathbf{Y} \end{bmatrix} \sim N \left( \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{0}_{kn} \\ \mathbf{X}\boldsymbol{\beta}^* \end{bmatrix}, \begin{bmatrix} \mathbf{H} & \mathbf{0}_{p \times kn} & \mathbf{H}\mathbf{X}' \\ \mathbf{0}_{kn \times p} & \mathbf{D} & \mathbf{D}\mathbf{Z}' \\ \mathbf{X}\mathbf{H} & \mathbf{Z}\mathbf{D} & \mathbf{X}\mathbf{H}\mathbf{X}' + \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R} \end{bmatrix} \right)$$

Se tiene entonces que

$$\boldsymbol{\beta} | \mathbf{Y} \sim N(\mathbf{c}^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} + \mathbf{H}^{-1}\boldsymbol{\beta}^*), \mathbf{c}^{-1})$$

donde

$$\mathbf{c} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \mathbf{H}^{-1}$$

A partir de esto se pueden realizar inferencias para  $\boldsymbol{\beta}$ . Si se quiere un estimador puntual, se puede seleccionar la media aposteriori o posterior

$$E[\boldsymbol{\beta} | \mathbf{Y}] = \mathbf{c}^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} + \mathbf{H}^{-1}\boldsymbol{\beta}^*)$$

con matriz de varianzas y covarianzas dada por

$$VarCov(E[\boldsymbol{\beta} | \mathbf{Y}]) = \mathbf{c}^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{c}^{-1}$$

Si se escoge una distribución a priori no informativa para  $\boldsymbol{\beta}$  con  $\mathbf{H}^{-1} = \mathbf{0}$ , entonces

$$\begin{aligned} E[\boldsymbol{\beta} | \mathbf{Y}] &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \\ VarCov(E[\boldsymbol{\beta} | \mathbf{Y}]) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \end{aligned}$$

Estos corresponden a los mínimos cuadrados generalizados para  $\boldsymbol{\beta}$ .

## 2. Estimación de $\mathbf{b}$

$$\mathbf{b} | \mathbf{Y} \sim N\left((\mathbf{Z}'\mathbf{LZ} + \mathbf{D})^{-1}\mathbf{Z}'\mathbf{L}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*), (\mathbf{Z}'\mathbf{LZ} + \mathbf{D})^{-1}\right)$$

donde

$$\mathbf{L} = (\mathbf{R} + \mathbf{X}\mathbf{H}\mathbf{X}')^{-1}$$

Si se escoge una distribución a priori no informativa para  $\boldsymbol{\beta}$  con  $\mathbf{H}^{-1} = \mathbf{0}$  y recordando que  $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}$  y  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$  se tiene que

$$E(\mathbf{b} | \mathbf{Y}) = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

que es el mejor predictor lineal insesgado (BLUP) de  $\mathbf{b}$ .

### Simplificando la estructura de $\epsilon$

Si en la aproximación bayesiana se asume que  $\epsilon \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$  entonces

$$\mathbf{Y} | \beta, \mathbf{b}, \sigma_\epsilon^2 \sim N(\mathbf{X}\beta + \mathbf{Zb}, \sigma_\epsilon^2 \mathbf{I})$$

Ahora, suponga que la a priori para  $\beta$  es

$$\beta | \sigma_\beta^2 \sim N(\mathbf{0}, \sigma_\beta^2 \mathbf{B})$$

donde  $\mathbf{B}$  es una matriz conocida y  $\sigma_\beta^2$  es un hiperparámetro. Además,

$$\mathbf{b} | \sigma_b^2 \mathbf{A} \sim N(\mathbf{0}, \sigma_b^2 \mathbf{A})$$

donde  $\mathbf{A}$  es una matriz conocida y  $\sigma_b^2$  es un hiperparámetro.

Asuma que las distribuciones a priori para los componentes de varianza son chi-cuadradas invertidas <sup>5</sup> escaladas con parámetros  $(s_\epsilon^2, \nu_\epsilon)$  y  $(s_b^2, \nu_b)$  para  $\sigma_\epsilon^2$  y  $\sigma_b^2$  respectivamente.

Las consecuencias de estos supuestos bayesianos son:

- Incondicionalmente de la varianza residual, el modelo muestral, dado  $\beta$  y  $\mathbf{b}$ , es una  $t$  multivariable, con f.d.p. dada por

---

<sup>5</sup>la f.d.p. de una v.a.  $X \sim \chi^2$  con  $\nu$  g.l. está dada por

$$f(x|\nu) = \left[ \Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2} \right]^{-1} x^{\nu/2-1} e^{-\frac{1}{2}x}, \quad x > 0$$

Si  $Y = 1/X$ , entonces se dice que se distribuye *chi-cuadrado invertida* con f.d.p. dada por

$$f(y|\nu) = \left[ \Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2} \right]^{-1} y^{-(\nu/2+1)} e^{-\frac{1}{2y}}, \quad y > 0$$

Sea  $s$  una constante positiva. Si  $X \sim Inv.\chi_\nu^2$  entonces  $Y = sX$  se dice que se distribuye  $Inv.\chi_\nu^2$  escalada con f.d.p. dada por

$$f(y|\nu, s) = \left[ \Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2} \right]^{-1} s^{\nu/2} y^{-(\nu/2+1)} e^{-\frac{s}{2y}}, \quad y > 0, s > 0, \nu > 0$$

Usando el símbolo de proporcionalidad  $\propto$

$$f(y|\nu, s) \propto y^{-(\nu/2+1)} e^{-\frac{s}{2y}}$$

$$f(\mathbf{Y} | \boldsymbol{\beta}, \mathbf{b}, s_\epsilon^2, \nu_\epsilon) \propto \left[ 1 + \frac{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})}{\nu_\epsilon s_\epsilon^2} \right]^{-\frac{n-\nu_\epsilon}{2}} \quad (*)$$

- La distribución marginal de los efectos aditivos incondicional de la varianza  $\sigma_b^2$  es una  $t$  multivariable con densidad dada por

$$f(\mathbf{b} | s_b^2, \nu_b) \propto \left[ 1 + \frac{\mathbf{b}'\mathbf{A}\mathbf{b}}{\nu_b s_b^2} \right]^{-\frac{q-\nu_b}{2}} \quad (**)$$

- En el modelo lineal mixto frecuentista la distribución marginal de las observaciones está indexada por el parámetro fijo  $\boldsymbol{\beta}$  y por las componentes de varianza  $\sigma_b^2$  y  $\sigma_\epsilon^2$  y está dado por:

$$\mathbf{Y} | \boldsymbol{\beta}, \sigma_b^2, \sigma_\epsilon^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_b^2 \mathbf{Z}\mathbf{A}\mathbf{Z}' + \sigma_\epsilon^2)$$

En el modelo bayesiano, de otro lado, la distribución marginal de las observaciones se obtiene integrando sobre todos los parámetros desconocidos  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{b}, \sigma_b^2, \sigma_\epsilon^2)$  que entran en el modelo.

$$f(\boldsymbol{\beta}, \mathbf{b}, \sigma_b^2, \sigma_\epsilon^2 | \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon) = f(\boldsymbol{\beta} | \sigma_\beta^2) f(\mathbf{b} | \sigma_b^2) f(\sigma_b^2 | s_b^2, \nu_b) f(\sigma_\epsilon^2 | s_\epsilon^2, \nu_\epsilon)$$

Entonces la densidad marginal de las observaciones puede escribirse como

$$\begin{aligned} f(\mathbf{Y} | \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon) &= \int \int \int \int f(\mathbf{Y} | \boldsymbol{\beta}, \mathbf{b}, \sigma_\epsilon^2) f(\boldsymbol{\theta} | \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon) d\boldsymbol{\theta} \\ &= \int \int \left[ \int f(\mathbf{Y} | \boldsymbol{\beta}, \mathbf{b}, \sigma_\epsilon^2) f(\sigma_\epsilon^2 | s_\epsilon^2, \nu_\epsilon) d\sigma_\epsilon^2 \right] \\ &\quad \times \left[ \int f(\mathbf{b} | \sigma_b^2) f(\sigma_b^2 | s_b^2, \nu_b) d\sigma_b^2 \right] f(\boldsymbol{\beta} | \sigma_\beta^2) d\boldsymbol{\beta} d\mathbf{b} \quad (***) \end{aligned}$$

Las expresiones entre corchetes evalúan (\*) y (\*\*). La densidad marginal de las observaciones, o predictiva a priori, no puede evaluarse en forma cerrada cuando las componentes de varianza son desconocidas.

Si los parámetros de dispersión fueran conocidos:

$$f(\mathbf{Y} | \sigma_\beta^2, \sigma_b^2, \sigma_\epsilon^2) = \int \int f(\mathbf{Y} | \boldsymbol{\beta}, \mathbf{b}, \sigma_\epsilon^2) f(\boldsymbol{\beta} | \sigma_\beta^2) f(\mathbf{b} | \sigma_b^2) d\boldsymbol{\beta} d\mathbf{b} \quad (B)$$

ya que las tres densidades en el integrando son normales, los exponentes pueden combinarse como

$$\begin{aligned} & \frac{1}{\sigma_\epsilon^2} \left[ (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} + \mathbf{b}' \mathbf{A}^{-1} \mathbf{b} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right] \\ = & \frac{1}{\sigma_\epsilon^2} [(\mathbf{Y} - \mathbf{W}\boldsymbol{\alpha})' (\mathbf{Y} - \mathbf{W}\boldsymbol{\alpha}) + \boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}] \quad (A) \end{aligned}$$

donde

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix}$$

$$\mathbf{W} = [ \mathbf{X} \quad \mathbf{Z} ]$$

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \mathbf{B}^{-1} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \end{bmatrix}$$

Ahora sea

$$\hat{\boldsymbol{\alpha}} = (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})^{-1} \mathbf{W}'\mathbf{Y}$$

Luego

$$\begin{aligned} & (\mathbf{Y} - \mathbf{W}\boldsymbol{\alpha})' (\mathbf{Y} - \mathbf{W}\boldsymbol{\alpha}) + \boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha} \\ = & \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\alpha}' \mathbf{W}'\mathbf{Y} + \boldsymbol{\alpha}' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) \boldsymbol{\alpha} \\ = & \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\alpha}' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) \hat{\boldsymbol{\alpha}} + \boldsymbol{\alpha}' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) \boldsymbol{\alpha} \\ & + \hat{\boldsymbol{\alpha}}' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\alpha}}' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) \hat{\boldsymbol{\alpha}} \end{aligned}$$

donde se suma y se resta  $\hat{\boldsymbol{\alpha}}$  para ayudar a completar un cuadrado. Se puede entonces re-escribir

$$\begin{aligned} & \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) \hat{\boldsymbol{\alpha}} + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \\ = & \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}'\mathbf{W}'\mathbf{Y} + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \end{aligned}$$

Usando esto en (A), el integrando en (B) se puede expresar como

$$\begin{aligned} f(\mathbf{Y} | \sigma_\beta^2, \sigma_b^2, \sigma_\epsilon^2) & \propto \int e^{\left(-\frac{1}{2\sigma_\epsilon^2} [\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}'\mathbf{W}'\mathbf{Y} + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})]\right)} d\boldsymbol{\alpha} \\ & \propto e^{\left(-\frac{\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}'\mathbf{W}'\mathbf{Y}}{2\sigma_\epsilon^2}\right)} \int e^{\left(-\frac{(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})}{2\sigma_\epsilon^2}\right)} d\boldsymbol{\alpha} \end{aligned}$$

Esta última integral es igual a

$$(2\pi)^{(p+q)/2} |\sigma_\epsilon^2 (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})|^{1/2}$$

que no involucra los datos. Se llega entonces a

$$f(\mathbf{Y} | \sigma_\beta^2, \sigma_b^2, \sigma_\epsilon^2) \propto e^{\left(-\frac{\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}'\mathbf{W}'\mathbf{Y}}{2\sigma_\epsilon^2}\right)}$$

Se puede mostrar entonces que

$$\mathbf{Y} | \sigma_\beta^2, \sigma_b^2, \sigma_\epsilon^2 \sim N(\mathbf{0}, \mathbf{X}\mathbf{B}\mathbf{X}'\sigma_\beta^2 + \mathbf{Z}\mathbf{A}\mathbf{Z}'\sigma_b^2 + \mathbf{I}\sigma_\epsilon^2)$$

Esto muestra que la parte bayesiana no puede ser construida como su contraparte frecuentista. La distribución marginal requiere el conocimiento de los efectos fijos y de las varianzas para poder calcular probabilidades.

### Distribución conjunta y condicional

Bajo los supuestos hechos, la f.d.p. posterior de todos los parámetros es

$$\begin{aligned} f(\boldsymbol{\beta}, \mathbf{b}, \sigma_b^2, \sigma_\epsilon^2 | \mathbf{Y}, \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon) & = f(\mathbf{Y} | \sigma_\beta^2, \sigma_b^2, \sigma_\epsilon^2) f(\boldsymbol{\beta} | \sigma_\beta^2) \\ & \quad \times f(\mathbf{b} | \sigma_b^2) f(\sigma_b^2 | s_b^2, \nu_b) f(\sigma_\epsilon^2 | s_\epsilon^2, \nu_\epsilon) \end{aligned} \quad (AA)$$

Todas las distribuciones condicionales posteriores pueden identificarse directamente de la densidad posterior conjunta, mientras se mantienen todos los

demás parámetros y datos fijos.

1. La distribución posterior condicional de  $\sigma_b^2$  dados los otros parámetros es

$$\begin{aligned} f(\sigma_b^2 | \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon, \dots) &\propto f(\mathbf{b} | \sigma_b^2) f(\sigma_b^2 | s_b^2, \nu_b) \\ &\propto (\sigma_b^2)^{-(q+\nu_b+2)/2} e^{\left(-\frac{1}{2\sigma_b^2}(\mathbf{b}'\mathbf{A}^{-1}\mathbf{b} + \nu_b s_b^2)\right)} \end{aligned}$$

que corresponde a una chi-cuadrado invertida escalada  $(\nu'_b, s_b'^2)$ , donde

$$\begin{aligned} \nu'_b &= q + \nu_b \\ s_b'^2 &= \frac{1}{q + \nu_b} (\mathbf{b}'\mathbf{A}^{-1}\mathbf{b} + \nu_b s_b^2) \end{aligned}$$

2. La distribución posterior condicional de  $\sigma_\epsilon^2$  dados los otros parámetros es

$$\begin{aligned} f(\sigma_\epsilon^2 | \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon, \dots) &\propto f(\mathbf{Y} | \sigma_\beta^2, \sigma_b^2, \sigma_\epsilon^2) f(\sigma_\epsilon^2 | s_\epsilon^2, \nu_\epsilon) \\ &\propto (\sigma_\epsilon^2)^{-(n+\nu_\epsilon+2)/2} e^{\left(-\frac{1}{2\sigma_\epsilon^2}((\mathbf{Y}-\mathbf{X}\beta-\mathbf{Zb})'(\mathbf{Y}-\mathbf{X}\beta-\mathbf{Zb}) + \nu_\epsilon s_\epsilon^2)\right)} \end{aligned}$$

que corresponde a una chi-cuadrado invertida escalada  $(\nu'_\epsilon, s_\epsilon'^2)$ , donde

$$\begin{aligned} \nu'_\epsilon &= n + \nu_\epsilon \\ s_\epsilon'^2 &= \frac{1}{n + \nu_\epsilon} ((\mathbf{Y}-\mathbf{X}\beta-\mathbf{Zb})'(\mathbf{Y}-\mathbf{X}\beta-\mathbf{Zb}) + \nu_\epsilon s_\epsilon^2) \end{aligned}$$

**Resultado:** dados todos los otros parámetros, los dos componentes de varianza son condicionalmente independientes.

3. La f.d.p. condicional conjunta de  $\beta$  y  $\mathbf{b}$  dados los componentes de varianza es

$$\begin{aligned}
 f(\boldsymbol{\beta}, \mathbf{b} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon) &\propto f(\mathbf{Y} | \boldsymbol{\beta}, \mathbf{b}, \sigma_\epsilon^2) f(\boldsymbol{\beta} | \sigma_\beta^2) \\
 &\quad \times f(\mathbf{b} | \sigma_b^2) \\
 &\propto e^{\left(\frac{1}{2\sigma_\epsilon^2} [\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}\mathbf{W}'\mathbf{Y} + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})]\right)} \\
 &\propto e^{\left(\frac{1}{2\sigma_\epsilon^2} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})\right)}
 \end{aligned}$$

De donde

$$\boldsymbol{\beta}, \mathbf{b} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon \sim N\left(\hat{\boldsymbol{\alpha}}, \sigma_\epsilon^2 (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})^{-1}\right)$$

Recordando que

$$\hat{\boldsymbol{\alpha}} = (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})^{-1} \mathbf{W}'\mathbf{Y}$$

se tiene que

$$\blacksquare (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})^{-1} \hat{\boldsymbol{\alpha}} = \mathbf{W}'\mathbf{Y}$$

■

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \mathbf{B}^{-1} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \end{bmatrix}$$

$$\blacksquare \mathbf{W} = [\mathbf{X} \quad \mathbf{Z}]$$

Entonces

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} + \mathbf{B}^{-1} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix} \quad (*)$$

Si  $\sigma_\beta^2 \rightarrow \infty$  (conocimiento a priori acerca de  $\boldsymbol{\beta}$  vago o difuso) (\*) se convierte en las ecuaciones de Henderson (Henderson, 1975). Los estimadores BLUP son casos especiales en el marco bayesiano. Si hay alguna información a priori acerca de  $\boldsymbol{\beta}$ ,  $\sigma_\beta^2$  será finita y el modelo bayesiano tiene un efecto de “encogimiento” hacia la media a priori (en este caso se asumió cero).

Sea  $\tilde{\boldsymbol{\alpha}}$  la solución al sistema

$$\mathbf{W}'\mathbf{W}\tilde{\boldsymbol{\alpha}} = \mathbf{W}'\mathbf{Y}$$

Se tiene que  $\tilde{\boldsymbol{\alpha}}$  es el e.m.v. en un modelo en que  $\boldsymbol{\beta}$  y  $\mathbf{b}$  son tratados como fijos (o como la media posterior donde  $\sigma_{\boldsymbol{\beta}}^2 \rightarrow \infty$  y  $\sigma_{\mathbf{b}}^2 \rightarrow \infty$ ).

Si  $\mathbf{W}'\mathbf{W}$  no es de rango completo entonces  $\tilde{\boldsymbol{\alpha}}$  no está definido de manera única. Si  $\hat{\boldsymbol{\alpha}}$  es único, suponiendo que  $\mathbf{X}$  tenga rango completo, o que  $\sigma_{\boldsymbol{\beta}}^2$  sea finita. Entonces, el vector de media posterior

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})^{-1} \mathbf{W}'\mathbf{Y} \\ &= (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})^{-1} (\mathbf{W}'\mathbf{W}\tilde{\boldsymbol{\alpha}} + \boldsymbol{\Sigma}^{-1}\mathbf{0})\end{aligned}$$

que corresponde a un promedio ponderado de  $\tilde{\boldsymbol{\alpha}}$  y  $\mathbf{0}$  (media a priori).

La distribución marginal conjunta  $(\boldsymbol{\beta}, \mathbf{b})$  dados los componentes de varianzas también es normal multivariable con vectores de medias  $\hat{\boldsymbol{\beta}}$  y  $\hat{\mathbf{b}}$ , respectivamente.

Note que la densidad posterior conjunta de  $\boldsymbol{\beta}$  y  $\mathbf{b}$ , dados los parámetros de dispersión, puede escribirse como

$$f(\boldsymbol{\beta}, \mathbf{b} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_{\boldsymbol{\beta}}^2) \propto e^{\left(-\frac{1}{2\sigma_\epsilon^2} \left( (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_{\boldsymbol{\beta}}^2} + \mathbf{b}' \mathbf{A}^{-1} \mathbf{b} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right)\right)}$$

Haga  $a(\boldsymbol{\beta}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$ . Entonces

$$\begin{aligned}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) + \mathbf{b}' \mathbf{A}^{-1} \mathbf{b} \frac{\sigma_\epsilon^2}{\sigma_b^2} &= (a(\boldsymbol{\beta}) - \mathbf{Z}\mathbf{b})' (a(\boldsymbol{\beta}) - \mathbf{Z}\mathbf{b}) + \mathbf{b}' \mathbf{A}^{-1} \mathbf{b} \frac{\sigma_\epsilon^2}{\sigma_b^2} \\ &= a'(\boldsymbol{\beta}) a(\boldsymbol{\beta}) - \hat{\mathbf{b}}(\boldsymbol{\beta})' \mathbf{Z}' a(\boldsymbol{\beta}) + \\ &\quad \left[ \hat{\mathbf{b}}(\boldsymbol{\beta}) - \mathbf{b} \right]' \left( \mathbf{Z}' \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right) \left[ \hat{\mathbf{b}}(\boldsymbol{\beta}) - \mathbf{b} \right]\end{aligned}$$

donde

$$\hat{\mathbf{b}}(\boldsymbol{\beta}) = \left( \mathbf{Z}' \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right)^{-1} \mathbf{Z}' a(\boldsymbol{\beta})$$

Usando este resultado en la densidad de arriba e integrando con respecto a  $\mathbf{b}$ , se obtiene la densidad posterior marginal de  $\boldsymbol{\beta}$ , de los componentes de

varianza

$$\begin{aligned}
 f(\boldsymbol{\beta} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2) &\propto e^{\left\{ -\frac{1}{2\sigma_\epsilon^2} \left( a'(\boldsymbol{\beta})a(\boldsymbol{\beta}) - \hat{\mathbf{b}}(\boldsymbol{\beta})' \mathbf{Z}' a(\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right) \right\}} \\
 &\quad \times \int \exp \left( -\frac{1}{2\sigma_\epsilon^2} \left( [\hat{\mathbf{b}}(\boldsymbol{\beta}) - \mathbf{b}]' \left( \mathbf{Z}' \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right) [\hat{\mathbf{b}}(\boldsymbol{\beta}) - \mathbf{b}] \right) \right) d\mathbf{b} \\
 &\propto \exp \left( -\frac{1}{2\sigma_\epsilon^2} \left( a'(\boldsymbol{\beta})a(\boldsymbol{\beta}) - \hat{\mathbf{b}}(\boldsymbol{\beta})' \mathbf{Z}' a(\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right) \right) \\
 &\quad \times (2\pi)^{q/2} \left| \left( \mathbf{Z}' \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right)^{-1} \sigma_\epsilon^2 \right|^{1/2}
 \end{aligned}$$

De esta manera

$$f(\boldsymbol{\beta} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2) \propto e^{\left( -\frac{1}{2\sigma_\epsilon^2} \left( a'(\boldsymbol{\beta})a(\boldsymbol{\beta}) - \hat{\mathbf{b}}(\boldsymbol{\beta})' \mathbf{Z}' a(\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right) \right)}$$

Ahora

$$\begin{aligned}
 &a'(\boldsymbol{\beta})a(\boldsymbol{\beta}) - \hat{\mathbf{b}}(\boldsymbol{\beta})' \mathbf{Z}' a(\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \\
 = &(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
 &- (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{Z} \left( \mathbf{Z}' \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right)^{-1} \mathbf{Z}' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \\
 = &(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \left[ \mathbf{I} - \mathbf{Z} \left( \mathbf{Z}' \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right)^{-1} \mathbf{Z}' \right] (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2}
 \end{aligned}$$

Usando el siguiente resultado del álgebra lineal:  $(\mathbf{A} + \mathbf{bGb}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{b} (\mathbf{b}' \mathbf{A}^{-1} \mathbf{b} + \mathbf{G}^{-1})^{-1} \mathbf{b}' \mathbf{A}^{-1}$  se puede mostrar que

$$\mathbf{I} - \mathbf{Z} \left( \mathbf{Z}' \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right)^{-1} \mathbf{Z}' = \left( \mathbf{Z} \mathbf{A} \mathbf{Z}' \frac{\sigma_b^2}{\sigma_\epsilon^2} + \mathbf{I} \right)^{-1} = \mathbf{V}^{-1}$$

(Ver<sup>6</sup>). Lo cual implica que

$$f(\boldsymbol{\beta} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2) \propto e^{\left(-\frac{1}{2\sigma_\epsilon^2} \left( (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right)\right)}$$

Note que cuando  $\sigma_\beta^2 \rightarrow \infty$ , el kernel de la densidad posterior tiende a la verosimilitud de  $\boldsymbol{\beta}$  de un modelo lineal mixto normal con componentes de varianzas conocidos y la matriz de covarianzas dada por  $\mathbf{ZAZ}'\sigma_b^2 + \mathbf{I}\sigma_\epsilon^2 = \mathbf{V}$

El E.M.V. de  $\boldsymbol{\beta}$  es  $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ . Entonces

$$\begin{aligned} f(\boldsymbol{\beta} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2) &\propto \exp\left(-\frac{1}{2\sigma_\epsilon^2} \left[ (\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \right. \right. \\ &\quad \left. \left. + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right] \right) \\ &\propto e^{\left(-\frac{1}{2\sigma_\epsilon^2} \left[ (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right] \right)} \end{aligned}$$

Ahora,

$$\begin{aligned} &(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \\ &= (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \left( \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \mathbf{B}^{-1} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ &\quad + \tilde{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \left( \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \mathbf{B}^{-1} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right)^{-1} \mathbf{B}^{-1} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \tilde{\boldsymbol{\beta}} \end{aligned} \tag{3.2}$$

donde

$$\hat{\boldsymbol{\beta}} = \left( \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \mathbf{B}^{-1} \frac{\sigma_\epsilon^2}{\sigma_\beta^2} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

---

<sup>6</sup>Otra forma de obtener  $\mathbf{V}^{-1}$  y que fue sugerida por uno de los revisores anónimos del texto es como sigue: Note que las matrices  $\mathbf{I} + \mathbf{ZUZ}'$  (donde  $\mathbf{Z}$  es fija y  $\mathbf{U}$  es arbitraria) son cerradas respecto de la multiplicación. Por lo tanto, la inversa de  $\mathbf{I} + \gamma\mathbf{ZAZ}'$  tendrá la misma forma que  $\mathbf{I} + \mathbf{ZBZ}'$  y entonces la matriz  $\mathbf{B}$  se halla resolviendo la ecuación  $(\mathbf{I} + \gamma\mathbf{ZAZ}')(\mathbf{I} + \mathbf{ZBZ}') = \mathbf{I}$  cuya solución es única y es exactamente  $\mathbf{V}^{-1}$

Entonces

$$f(\boldsymbol{\beta} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2) \propto e^{\left(-\frac{1}{2\sigma_\epsilon^2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})' \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \mathbf{B}^{-1}\frac{\sigma_\epsilon^2}{\sigma_\beta^2}\right)(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right)}$$

Así, la distribución posterior marginal de  $\boldsymbol{\beta}$  cuando los parámetros de dispersión son conocidos es

$$\boldsymbol{\beta} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2 \sim N\left(\hat{\boldsymbol{\beta}}, \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \mathbf{B}^{-1}\frac{\sigma_\epsilon^2}{\sigma_\beta^2}\right)^{-1} \sigma_\epsilon^2\right)$$

La marginal de  $\mathbf{b}$  dadas las componentes de varianza se pueden hallar en forma similar.

$$\mathbf{b} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2 \sim N\left(\hat{\mathbf{b}}, \left(\mathbf{Z}'\mathbf{T}^{-1}\mathbf{Z} + \mathbf{A}^{-1}\frac{\sigma_\epsilon^2}{\sigma_b^2}\right)^{-1} \sigma_\epsilon^2\right)$$

con

$$\hat{\mathbf{b}} = \left(\mathbf{Z}'\mathbf{T}^{-1}\mathbf{Z} + \mathbf{A}^{-1}\frac{\sigma_\epsilon^2}{\sigma_b^2}\right)^{-1} \mathbf{Z}'\mathbf{T}^{-1}\mathbf{Y}$$

y

$$\mathbf{T} = \mathbf{X}\mathbf{B}\mathbf{X}'\frac{\sigma_\beta^2}{\sigma_\epsilon^2} + \mathbf{I}$$

El vector de medias es igual al del componente  $\mathbf{b}$  de  $\hat{\boldsymbol{\alpha}}$ . Puesto que la distribución de  $\boldsymbol{\beta}$  y  $\mathbf{b}$  es conjuntamente normal, se sigue que

$$\boldsymbol{\beta} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon, \mathbf{b} \sim N\left(\hat{\boldsymbol{\beta}}(\mathbf{b}), \left(\mathbf{X}'\mathbf{X} + \mathbf{B}^{-1}\frac{\sigma_\epsilon^2}{\sigma_\beta^2}\right)^{-1} \sigma_\epsilon^2\right)$$

donde

$$\hat{\boldsymbol{\beta}}(\mathbf{b}) = \left(\mathbf{X}'\mathbf{X} + \mathbf{B}^{-1}\frac{\sigma_\epsilon^2}{\sigma_\beta^2}\right)^{-1} \mathbf{X}'(\mathbf{Y} - \mathbf{Z}\mathbf{b})$$

También

$$\mathbf{b} | \mathbf{Y}, \sigma_b^2, \sigma_\epsilon^2, \sigma_\beta^2, s_b^2, \nu_b, s_\epsilon^2, \nu_\epsilon, \boldsymbol{\beta} \sim N\left(\hat{\mathbf{b}}(\boldsymbol{\beta}), \left(\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\frac{\sigma_\epsilon^2}{\sigma_b^2}\right)^{-1} \sigma_\epsilon^2\right)$$

donde

$$\hat{\mathbf{b}}(\boldsymbol{\beta}) = \left( \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_\epsilon^2}{\sigma_b^2} \right)^{-1} \mathbf{Z}' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

### Distribución marginal de las componentes de varianza

La densidad posterior conjunta de todos los parámetros es

$$\begin{aligned} f(\boldsymbol{\beta}, \mathbf{b}, \sigma_b^2, \sigma_\epsilon^2 | \mathbf{Y}, s_\beta^2, s_b^2, s_\epsilon^2, \sigma_\beta^2) &\propto f(\sigma_b^2 | s_b^2, \nu_b) f(\sigma_\epsilon^2 | s_\epsilon^2, \nu_\epsilon) \\ &\times (\sigma_\epsilon^2)^{-n/2} (\sigma_b^2)^{-q/2} e^{\left( -\frac{1}{2\sigma_\epsilon^2} [\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}_c' \mathbf{W}\mathbf{Y} + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_c)' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_c)] \right)} \\ &\times e^{\left( -\frac{1}{2\sigma_\epsilon^2} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_c)' (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_c) \right)} \end{aligned}$$

donde  $\hat{\boldsymbol{\alpha}}_c = \hat{\boldsymbol{\alpha}} = (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})^{-1} \mathbf{W}'\mathbf{Y}$ , pero se le pone el subíndice para realzar la dependencia entre las componentes de varianza  $\sigma_b^2$  y  $\sigma_\epsilon^2$ .

Los parámetros de localización aparecen en el kernel de la densidad normal multivariable y pueden eliminarse mediante integración. Luego de integrar se obtiene

$$\begin{aligned} f(\sigma_b^2, \sigma_\epsilon^2 | \mathbf{Y}, s_\beta^2, s_b^2, s_\epsilon^2, \sigma_\beta^2) &\propto f(\sigma_b^2 | s_b^2, \nu_b) f(\sigma_\epsilon^2 | s_\epsilon^2, \nu_\epsilon) \\ &\times (\sigma_\epsilon^2)^{-n/2} (\sigma_b^2)^{-q/2} e^{\left( -\frac{1}{2\sigma_\epsilon^2} [\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}_c' \mathbf{W}\mathbf{Y}] \right)} |(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1}) \sigma_\epsilon^2|^{1/2} \end{aligned}$$

La f.d.p. conjunta de las componentes de varianza puede escribirse explícitamente

$$\begin{aligned} f(\sigma_b^2, \sigma_\epsilon^2 | \mathbf{Y}, s_\beta^2, s_b^2, s_\epsilon^2, \sigma_\beta^2) &\propto (\sigma_\epsilon^2)^{-(n-p-q+\nu_\epsilon+2)/2} (\sigma_b^2)^{-(q+\nu_b+2)/2} \\ &\times e^{\left( -\frac{1}{2\sigma_\epsilon^2} \left[ \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}_c' \mathbf{W}\mathbf{Y} + \nu_\epsilon s_\epsilon^2 + \nu_b s_b^2 \frac{\sigma_\epsilon^2}{\sigma_b^2} \right] \right)} |(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}^{-1})|^{1/2} \end{aligned}$$

No es posible ir más allá a nivel de marginalización ya que la f.d.p.c. involucra la razón entre las componentes de varianza, ambos en el exponente y en la matriz  $\boldsymbol{\Sigma}$  como parte del determinante.

## 3.7. Pruebas de hipótesis

En esta sección se discuten algunos casos relacionados con pruebas de hipótesis.

### 3.7.1. V conocida

Suponga que se tiene el modelo de Laird-Ware (Laird y Ware, 1982) especificado como  $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$  con  $VarCov(\mathbf{Y}_i) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{R}_i = \mathbf{V}_i$ , la cual se asumirá por el momento conocida.

Se desea verificar

$$H_0 : \mathbf{A}\boldsymbol{\beta} = \mathbf{d}, \quad (\mathbf{d} \text{ y } \mathbf{A} \text{ conocidas})$$

La prueba se basa en

$$\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d} \sim N\left(\mathbf{A}\boldsymbol{\beta} - \mathbf{d}; \mathbf{A}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{A}'\right)$$

Se puede usar como estadístico de prueba la forma cuadrática

$$\left(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d}\right)' \mathbf{P} \left(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d}\right)$$

Su distribución está respaldada por el siguiente resultado:

**Lema 3.7.1.** *Asuma que  $\mathbf{Z} \sim N(\boldsymbol{\mu}; \boldsymbol{\Sigma})$ . Entonces  $\mathbf{Z}'\mathbf{P}\mathbf{Z} \sim \chi_{\nu, \lambda}^2$  con  $\nu = \text{rango}(\mathbf{P})$ , los grados de libertad, y  $\lambda = \frac{1}{2}\boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu}$ , parámetro de no centralidad, si, y solo si,  $\mathbf{P}\boldsymbol{\Sigma}$  es idempotente ( $\mathbf{P}\boldsymbol{\Sigma} = (\mathbf{P}\boldsymbol{\Sigma})^2$ ).*

Haga  $\mathbf{Z} = \mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d}$  y  $\boldsymbol{\Sigma} = \mathbf{A}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{A}'$ . Escoja  $\mathbf{P}$  tal que  $\mathbf{P}\boldsymbol{\Sigma}$  sea idempotente. Por ejemplo  $\left(\mathbf{A}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{A}'\right)^{-1}$  cumple este requerimiento. Entonces

$$W = \left(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d}\right)' \left(\mathbf{A}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{A}'\right)^{-1} \left(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d}\right) \sim \chi_{\text{Rango}(\mathbf{A})}^2$$

Si  $\mathbf{A}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{A}'$  no es de rango completo escoja  $\mathbf{P}$ , una inversa generalizada.

### 3.7.2. V desconocida

Suponga que  $VarCov(\mathbf{Y}_i) = \sigma^2\mathbf{V}_i$  donde  $\sigma^2$  es desconocida y  $\mathbf{V}_i$  es conocida. Entonces

$$\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d} \sim N\left(\mathbf{A}\boldsymbol{\beta} - \mathbf{d}; \sigma^2\mathbf{A}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{A}'\right)$$

Si

$$\hat{\sigma}^2 = \frac{1}{T - \text{rango}(\mathbf{X})} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \left[\mathbf{A}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{A}'\right]^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

es un estimador de  $\sigma^2$ .

**Lema 3.7.2.** Usando los supuestos propuestos arriba

$$1. \quad \frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})' (\mathbf{A} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{A}')^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})}{\hat{\sigma}^2} \sim \chi_{\text{rango}(\mathbf{A})}^2$$

y

$$\frac{(T - \text{rango}(\mathbf{X})) \hat{\sigma}^2}{\sigma^2} \sim \chi_{T - \text{rango}(\mathbf{X})}^2$$

2.  $\hat{\sigma}^2$  y  $\hat{\boldsymbol{\beta}}$  son independientes.

**Prueba:**

1. Para probar este numeral considere la forma cuadrática  $(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})' \mathbf{P} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})$  donde  $\mathbf{P} = \frac{1}{\sigma^2} (\mathbf{A} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{A}')^{-1}$  con

$$\boldsymbol{\Sigma} = \text{VarCov} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d}) = \sigma^2 \mathbf{A} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{A}'$$

Se puede mostrar que  $\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Sigma} = \mathbf{I}$  y por lo tanto  $\mathbf{P}\boldsymbol{\Sigma}$ . Considere ahora la forma cuadrática  $(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{r}'\mathbf{V}^{-1}\mathbf{r}$ .

$$\mathbf{r} \sim N(\mathbf{0}; (\mathbf{I} - \mathbf{X}^*\mathbf{V}^{-1})' \mathbf{V} (\mathbf{I} - \mathbf{X}^*\mathbf{V}^{-1}))$$

donde  $\mathbf{X}^* = \mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'$ .

Se puede mostrar que  $\mathbf{X}^*\mathbf{V}^{-1}$  es idempotente:

$$\begin{aligned} \mathbf{X}^*\mathbf{V}^{-1}\mathbf{X}^*\mathbf{V}^{-1} &= \mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}' \\ &= \mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \\ &= \mathbf{X}^*\mathbf{V}^{-1} \end{aligned}$$

Ahora haga  $\mathbf{P} = \frac{1}{\sigma^2} \mathbf{V}^{-1}$ . Note que

$$\begin{aligned} \mathbf{P} \cdot \text{VarCov}(\mathbf{r}) &= \frac{1}{\sigma^2} (\mathbf{I} - \mathbf{X}^* \mathbf{V}^{-1} - \mathbf{X}^* \mathbf{V}^{-1} + \mathbf{V}^{-1} \mathbf{X}^* \mathbf{X}^* \mathbf{V}^{-1}) \\ &= \frac{1}{\sigma^2} (\mathbf{I} - 2\mathbf{X}^* \mathbf{V}^{-1} + \mathbf{X}^* \mathbf{V}^{-1}) \\ &= \frac{1}{\sigma^2} (\mathbf{I} - \mathbf{X}^* \mathbf{V}^{-1}) \leftarrow \text{Es idempotente ya que lo es } \mathbf{X}^* \mathbf{V}^{-1} \end{aligned}$$

Esto implica que

$$\frac{\mathbf{r}' \mathbf{V}^{-1} \mathbf{r}}{\sigma^2} \sim \chi_{\text{rango}(\mathbf{I} - \mathbf{X}^* \mathbf{V}^{-1})}^2$$

Pero  $\text{rango}(\mathbf{I} - \mathbf{X}^* \mathbf{V}^{-1}) = T - \text{rango}(\mathbf{X})$

Note que  $(T - \text{rango}(\mathbf{X})) \hat{\sigma}^2 = \mathbf{r}' \mathbf{V}^{-1} \mathbf{r}$

2. La independencia entre  $\hat{\sigma}^2$  y  $\hat{\boldsymbol{\beta}}$  sigue de la independencia entre  $\mathbf{r} = \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}$  y  $\hat{\boldsymbol{\beta}}$ . Ya que ambos se distribuyen normalmente, la independencia se puede verificar si su covarianza es cero.

$$\begin{aligned} \text{Cov} [\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}, \boldsymbol{\beta}] &= \text{Cov} [(\mathbf{I} - \mathbf{X}^* \mathbf{V}^{-1}) \mathbf{Y}, (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y}] \\ &= (\mathbf{I} - \mathbf{X}^* \mathbf{V}^{-1}) \mathbf{V} \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \\ &= \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} - \mathbf{X}^* \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \\ &= \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} - \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \\ &= \mathbf{0} \end{aligned}$$

Con esto se prueba el resultado.

Bajo las condiciones 1) y 2) se tiene que

$$\begin{aligned} F &= \frac{\left( \frac{(\mathbf{A} \hat{\boldsymbol{\beta}} - \mathbf{d})' [\mathbf{A} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{A}']^{-1} (\mathbf{A} \hat{\boldsymbol{\beta}} - \mathbf{d})}{\sigma^2 / \text{rango}(\mathbf{A})} \right)}{(T - \text{rango}(\mathbf{X})) \hat{\sigma}^2 / \sigma^2} \\ &= \frac{(\mathbf{A} \hat{\boldsymbol{\beta}} - \mathbf{d})' [\mathbf{A} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{A}']^{-1} (\mathbf{A} \hat{\boldsymbol{\beta}} - \mathbf{d})}{\text{rango}(\mathbf{A}) \hat{\sigma}^2} \sim F_{(\text{rango}(\mathbf{A}), T - \text{rango}(\mathbf{X}))} \end{aligned}$$

### 3.7.3. $VarCov(\mathbf{Y}_i)$ desconocida

Cuando esta matriz es desconocida se recomienda trabajar con  $\mathbf{V}(\hat{\boldsymbol{\theta}})$ , en lugar de trabajar con

$$W = (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})' \left[ \mathbf{A} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{A}' \right]^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d}) \sim \chi_{\text{rango}(\mathbf{A})}^2$$

que se usa en el caso de  $VarCov(\mathbf{Y}_i)$  conocida. Se trabajaría entonces con

$$W^* = (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})' \left[ \mathbf{A} (\mathbf{X}'\mathbf{V}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{X})^{-1} \mathbf{A}' \right]^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d}) \sim ?$$

Si  $VarCov(\mathbf{Y}_i) = \sigma^2\mathbf{V}_i$  ( $\mathbf{V}_i$  conocida) se tiene el estadístico de prueba

$$F = \frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})' \left[ \mathbf{A} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{A}' \right]^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})}{\text{rango}(\mathbf{A})\hat{\sigma}^2} \sim F_{(\text{rango}(\mathbf{A}), T - \text{rango}(\mathbf{X}))}$$

Pero si  $\mathbf{V}_i$  es desconocida se podría trabajar con

$$F^* = \frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})' \left[ \mathbf{A} (\mathbf{X}'\mathbf{V}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{X})^{-1} \mathbf{A}' \right]^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{d})}{\text{rango}(\mathbf{A})\hat{\sigma}^2} \sim ?$$

En la práctica se usan los valores críticos de la distribución  $\chi^2$  o de la  $F$  bajo el argumento asintótico que  $\mathbf{V}(\hat{\boldsymbol{\theta}}) \rightarrow \mathbf{V}(\boldsymbol{\theta})$  si  $T \rightarrow \infty$ , asumiendo que  $\hat{\boldsymbol{\theta}}$  sea un estimador consistente para  $\boldsymbol{\theta}$ .

La distribución asintótica de  $F^*$  no es una  $F$  ya que  $\hat{\sigma}^2 \rightarrow \sigma^2$  cuando  $T \rightarrow \infty$ , la distribución asintótica es una  $\chi^2$  escalada, aún así se usa la distribución  $F_{(\text{rango}(\mathbf{A}), T - \text{rango}(\mathbf{X}))}$  como aproximación.

## 3.8. Inferencia para la media

Recuerde que un estimador para  $\boldsymbol{\beta}$  es:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left( \sum_{i=1}^N \mathbf{X}_i' \mathbf{W}_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{W}_i \mathbf{Y}_i$$

donde  $\boldsymbol{\alpha}$  se reemplaza por su estimador ML o REML y  $\mathbf{W}_i = \mathbf{V}_i^{-1}$ . Condi-  
 cionando en  $\boldsymbol{\alpha}$ ,  $\widehat{\boldsymbol{\beta}}$  tiene media

$$\begin{aligned} E \left[ \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \right] &= \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i E[\mathbf{Y}_i] \\ &= \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i \mathbf{X}_i \boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned}$$

Aquí, a fin de que  $\widehat{\boldsymbol{\beta}}$  sea insesgado, es suficiente que la media de la respuesta  
 esté correctamente especificada.

Condicionando en  $\boldsymbol{\alpha}$ ,  $\widehat{\boldsymbol{\beta}}$  tiene covarianza:

$$Cov \left[ \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \right] = \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i Var(\mathbf{Y}_i) \mathbf{W}_i \mathbf{X}_i \right) \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i \mathbf{X}_i \right)^{-1}$$

Este estimador se conoce como *estimador sandwich* o *robusto* (las cantidades  
 $\left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i \mathbf{X}_i \right)^{-1}$  pueden pensarse como si fueran los panes y la cantidad  
 del medio como el jamón). Este resultado asume que la matriz de varianzas y  
 covarianzas  $Var(\mathbf{Y}_i)$  se modela de manera correcta como  $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i + \boldsymbol{\Sigma}_i$ .

El llamado *estimador naive* o *basado en el modelo* para la  $Cov(\widehat{\boldsymbol{\beta}})$ , el cual no  
 asume que la matriz de varianzas y covarianzas para  $\mathbf{Y}_i$  esté correctamente  
 especificada, se obtiene al reemplazar a  $Var(\mathbf{Y}_i)$  por

$$\left[ \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}} \right] \left[ \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}} \right]'$$

en vez de  $\mathbf{V}_i$ . La única condición para que  $\left[ \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}} \right] \left[ \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}} \right]'$  sea  
 insesgado para  $Var(\mathbf{Y}_i)$  es que la media esté especificada de manera correcta.

Esto sugiere que mientras el interés sea solo en inferencias relacionadas con  
 la estructura de la media, se debe prestar poca atención a la forma en que  
 se modela la estructura de covarianza, dado que el conjunto de datos sea  
 suficientemente grande.

El estimador sandwich puede obtenerse con el SAS usando la opción 'empi-  
 rical' del PROC MIXED.

### 3.8.1. Otra forma de obtener estos resultados

Sea  $\mathbf{X}'\mathbf{W}\mathbf{X} = \sum_{i=1}^N \mathbf{X}'_i \mathbf{W}_i \mathbf{X}_i$ , de esta forma

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

Entonces

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})] &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}E[\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{I}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned}$$

y

$$\begin{aligned} Cov[\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})] &= Cov[(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}Cov(\mathbf{Y})\mathbf{W}\mathbf{X} (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \quad (\mathbf{W} \text{ es simétrica}) \\ &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{W}^{-1}\mathbf{W}\mathbf{X} (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{I}\mathbf{W}\mathbf{X} (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{I} \\ &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \end{aligned}$$

## 3.9. Pruebas de hipótesis e intervalos de confianza

### 3.9.1. Método 1: Test aproximado de Wald

Para cualquier matriz conocida  $\mathbf{L}$ , se puede probar

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0} \quad \text{Vs.} \quad H_1 : \mathbf{L}\boldsymbol{\beta} \neq \mathbf{0}$$

usando el test estadístico

$$W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{L}' \left[ \mathbf{L} \left( \sum_{i=1}^m \mathbf{X}'_i \mathbf{V}_i^{-1}(\hat{\boldsymbol{\alpha}}) \mathbf{X}_i \right)^{-1} \mathbf{L}' \right]^{-1} \mathbf{L} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

Se puede probar que bajo la hipótesis nula  $W \sim \chi_{Rango(\mathbf{L})}^2$  si el tamaño muestral es grande. Sin embargo, Dempster, Rubin y Tsutakawa (1981) argumentan que el test tipo Wald está basado en errores estándar que subestiman la verdadera variabilidad en la estimación de  $\beta$  debido a que estos no tienen en cuenta la variabilidad introducida al estimar a  $\alpha$ . En la práctica este problema de subestimación de la varianza se resuelve frecuentemente usando estadísticos cuyas distribuciones son aproximaciones a la distribución F (o a la t).

### 3.9.2. Método 2: Test aproximados a la distribución F

El test alternativo está dado por:

$$F = \frac{W}{Rango(\mathbf{L})}$$

Bajo la hipótesis nula,  $F$  se distribuye de manera aproximada como una F central donde los grados de libertad del numerador están dados por  $Rango(\mathbf{L})$  y los del denominador están dados por  $\nu$  los cuales se estiman a partir de los datos. Si  $Rango(\mathbf{L}) = 1$  el test  $F$  es el cuadrado de una t aproximada.

Existen algunos métodos para estimar a  $\nu$ . A continuación se presentan los más comunes y la manera en que se implementan en SAS:

1. Método 'Residual': los grados de libertad residuales  $n - Rango(\mathbf{X})$  donde  $n$  es el número de observaciones. En realidad esta opción toma, para cada efecto fijo, los grados de libertad de la varianza de error residual. Este número típicamente sobreestima los grados de libertad óptimos del denominador. Por lo tanto esta opción no es muy recomendable. En SAS se implementa como DDFM=RESIDUAL.
2. Método 'Between/Within': Esta opción es el default cuando se usa la opción REPEATED sin efectos aleatorios (sin RANDOM). Se calcula al desagregar los grados de libertad residuales en porciones intra y entre sujetos. PROC MIXED entonces chequea si un efecto fijo cambia dentro de cualquier sujeto. Si cambia, asigna grados de libertad intra-sujeto al efecto; de otra forma, asigna los grados de libertad entre-sujetos al efecto. Si hay múltiples efectos intra-sujeto que contienen variables de clasificación los grados de libertad intra-sujeto se particionan en componentes correspondientes a las interacciones sujeto por efecto. En SAS se invoca como DDFM=BETWITHIN.

3. Método 'Containment': Este método es el default cuando se usa la opción RANDOM en SAS. Este método se basa en el *método de contención* para calcular los grados de libertad del denominador. El *Método de contención* se lleva a cabo de la siguiente forma: Denote el efecto fijo como A (A es un factor de efectos fijos para el cual los grados de libertad del denominador para probar la hipótesis de igualdad de las respuestas medias en todos los niveles se quiere determinar) y luego busque en la lista de efectos declarados como aleatorios (RANDOM) aquellos que contengan en su sintaxis a A (por ejemplo, B\*A o B(A) donde B es aleatorio). Por ejemplo, el efecto aleatorio B(A) contiene a A, pero otro efecto, por ejemplo C, no contiene a A aún si tiene los mismos niveles de B(A). Entre los efectos aleatorios que contienen a A, calcula su contribución al rango de la matriz  $(\mathbf{X}, \mathbf{Z})$ . Los grados de libertad asociados a A consisten en el mínimo de estas contribuciones al rango. Si no se encuentran efectos, los grados de libertad para el efecto fijo A se hacen iguales a los grados de libertad residuales. Esta elección de grados de libertad coincide con el test para diseños split-plot y se considera adecuada para diseños moderadamente desbalanceados. Los grados de libertad del denominador derivados de este método son correctos si el diseño es balanceado, pero usualmente en la práctica este no es el caso. Para datos desbalanceados, distintas componentes de varianza contribuyen a la variabilidad total. Por lo tanto, los grados de libertad asignados por este método no siempre son óptimos ya que corresponden a los grados de libertad asociados a un solo efecto aleatorio. Una mejor elección podría caer entre el mínimo y el máximo valor de grados de libertad asociados a los diferentes factores de efectos aleatorios. En SAS se implementa como DDFM=CONTAIN.
4. Método de Satterthwaite: Este procedimiento es ligeramente más costoso desde el punto de vista computacional. En SAS se implementa como DDFM=SATTERTH. Una de sus ventajas es que usa las matrices de los efectos fijos y aleatorios para determinar los grados de libertad del denominador, de manera que solo la matriz y no el nombre sintáctico importa para determinar los grados de libertad. Este método ejecuta una aproximación general de Satterthwaite para los grados de libertad del denominador de la F, como sigue. Sea  $\boldsymbol{\theta}$  un vector de parámetros en  $\mathbf{V}$  y defina  $\mathbf{C} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$  donde  $-$  denota la inversa generalizada de Moore-Penrose. Sean  $\hat{\mathbf{C}}$  y  $\hat{\boldsymbol{\theta}}$  las estimaciones correspondientes.

Considere primero el **caso unidimensional**, y sea  $\boldsymbol{\ell}$  un vector que define una combinación lineal estimable de  $\boldsymbol{\beta}$ . Los grados de libertad

de Satterthwaite para el estadístico

$$t = \frac{\ell\hat{\beta}}{\sqrt{\ell\hat{\mathbf{C}}\ell'}}$$

se calculan como

$$\nu = \frac{2(\ell\hat{\mathbf{C}}\ell')^2}{\mathbf{g}'\mathbf{A}\mathbf{g}}$$

donde  $\mathbf{g}$  es el gradiente de  $\ell\hat{\mathbf{C}}\ell'$  con respecto a  $\boldsymbol{\theta}$  evaluado en  $\hat{\boldsymbol{\theta}}$  y  $\mathbf{A}$  es la matriz de varianzas y covarianzas asintótica de  $\hat{\boldsymbol{\theta}}$  obtenida a partir de la Hessiana.

Para el **caso Multivariado**, sea  $\mathbf{L}$  una matriz de contrastes estimable y denote el rango de  $\mathbf{L}\hat{\mathbf{C}}\mathbf{L}'$  como  $q$  ( $q > 1$ ). Los grados de libertad para el denominador del estadístico  $F$  de acuerdo al método de Satterthwaite<sup>7</sup> se calculan primero llevando a cabo la descomposición espectral

$$\mathbf{L}\hat{\mathbf{C}}\mathbf{L}' = \mathbf{P}'\mathbf{D}\mathbf{P}$$

donde  $\mathbf{P}$  es una matriz ortogonal de vectores propios y  $\mathbf{D}$  es una matriz diagonal de valores propios (ambas de dimensión  $q \times q$ ). Defina a  $\ell_m$  la  $m$ -ésima fila de  $\mathbf{P}\mathbf{L}$  y sea

$$V_m = \frac{2(D_m)^2}{\mathbf{g}'_m\mathbf{A}\mathbf{g}_m}$$

Donde  $D_m$  es el  $m$ -ésimo elemento de la diagonal de  $\mathbf{D}$  y  $\mathbf{g}_m$  es el gradiente de  $\ell_m\mathbf{C}\ell'_m$  con respecto a  $\boldsymbol{\theta}$  evaluado en  $\hat{\boldsymbol{\theta}}$ . Luego haga

$$E = \sum_{m=1}^q \frac{V_m}{V_m - 2} I_{(V_m > 2)}$$

Los grados de libertad para el estadístico  $F$  se calculan como:

$$\nu = \begin{cases} \frac{2E}{E-q} & \text{Si } E > q \\ 0 & \text{Si } E \leq q \end{cases}$$

---

7

$$F = \frac{\hat{\beta}'\mathbf{L}'(\mathbf{L}\hat{\mathbf{C}}\mathbf{L}')^{-1}\mathbf{L}\hat{\beta}}{q}$$

**Caso especial**  $\text{Rango}(\mathbf{L}) = 1$ . Sea  $\mathbf{L} = \mathbf{h}$  un vector. Inspirado en ideas de formas cuadráticas (las medias cuadráticas típicamente provienen de formas cuadráticas) el método de Satterthwaite usa una distribución  $\chi_\nu^2$  como aproximación a la distribución de

$$S_h^2 = \frac{\nu \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h}}{E \left( \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h} \right)}$$

Note que  $E(S_h^2) = \nu$  ( $S_h^2$  tiene la misma media de una distribución  $\chi_\nu^2$ ). Para obtener los grados de libertad aproximados, se igualan las varianzas de las dos distribuciones:

$$\text{Var}(S_h^2) = \frac{\nu^2 \text{Var} \left( \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h} \right)}{\left( E \left( \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h} \right) \right)^2} \approx 2\nu$$

o

$$\nu \approx \frac{2 \left( E \left( \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h} \right) \right)^2}{\text{Var} \left( \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h} \right)}$$

Solo resta encontrar una aproximación para  $\text{Var} \left( \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h} \right)$ .

Note que la matriz  $\mathbf{Z}$  se puede particionar en submatrices  $\mathbf{Z}_1, \dots, \mathbf{Z}_{k-1}$  donde  $\mathbf{Z}_j$  es la matriz para el  $j$ -ésimo efecto aleatorio (y  $k-1$  es el número de efectos aleatorios). El error aleatorio también se puede ver como un efecto aleatorio. En realidad la ecuación del MLM

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{Z}_1 | \mathbf{Z}_2 | \dots | \mathbf{Z}_{k-1}) \mathbf{b} + \boldsymbol{\epsilon}$$

Puede escribirse como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{Z}_1 | \mathbf{Z}_2 | \dots | \mathbf{Z}_{k-1} | \mathbf{I}_N) \begin{pmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{pmatrix}$$

En términos de  $\mathbf{Z}_j$ ,  $j = 1, 2, \dots, k$  (con  $\mathbf{Z}_k = \mathbf{I}_N$ ) y usando una expan-

sión en series de Taylor, Giesbrecht y Burns (1985) muestran que

$$Var \left( \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h} \right) \approx \sum_{i=1}^k \sum_{j=1}^k Cov \left( \widehat{\sigma}_i^2, \widehat{\sigma}_j^2 \right) M_i M_j$$

donde

$$M_j = \mathbf{h}' \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{Z}_j \mathbf{Z}_j' \widehat{\mathbf{V}}^{-1} \mathbf{X} \left( \mathbf{X}' \widehat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-} \mathbf{h}$$

y  $Cov \left( \widehat{\sigma}_i^2, \widehat{\sigma}_j^2 \right)$  son las varianzas y covarianzas asintóticas obtenidas a partir de la Hessiana. El lado derecho de esta ecuación proporciona una aproximación para los grados de libertad  $\nu$

5. Método de Kenward-Rogers: La opción DDFM=KENWARDROGER permite obtener los grados de libertad de acuerdo a un método detallado en Kenward y Roger (1997). Esta aproximación consiste en inflar la matriz estimada de varianzas y covarianzas de efectos fijos y aleatorios usando un método propuesto por Prasad y Rao (1990) y Harville y Jeske (1992). Después de realizar este ajuste, se calculan los grados de libertad de acuerdo al método de Satterthwaite. Por defecto para los cálculos, se usa la matriz de información observada.

### 3.10. Inferencia para las componentes de varianza

Inferencia respecto a la estructura de la media usualmente es de primordial interés. Sin embargo, hay que tener en cuenta que cuando se trabaja con MLM, realizar inferencia para la estructura de covarianza también es de interés ya que puede justificar el uso de componentes aleatorias dentro del modelo. Otras razones incluyen:

1. Interpretación de la variación aleatoria en los datos.
2. Estructuras de covarianza sobrep parametrizadas conducen a inferencias ineficientes para la media.
3. Modelos demasiado restrictivos invalidan también las inferencia para la estructura de la media.

### 3.10.1. Test de Wald aproximados

En esta sección se discute algunos test tipo Wald aproximados basados en la normalidad asintótica de los estimadores de máxima verosimilitud y máxima verosimilitud restringida. **Recuerde:** “Inferencias que pueden ser plausibles para  $\beta$  no necesariamente lo son para  $\alpha$ ”. Sin embargo, el desempeño de la aproximación normal depende fuertemente del verdadero valor de  $\alpha$  y se necesitan muestras grandes cuando los valores de  $\alpha$  están relativamente cercanos al espacio paramétrico  $\Theta_\alpha$ . En caso de que  $\alpha$  sea un valor de la frontera, la aproximación normal falla. Esto acarrea importantes consecuencias para los test de significancia para las componentes de varianza: Los test tipo Wald pueden o no producir inferencias válidas. Cuando  $\mathbf{D}$  es definida positiva implica que  $H_0$  está en la frontera del espacio paramétrico  $\Theta_\alpha$  por lo tanto la distribución asintótica de  $\frac{\hat{d}_{ii}}{SE(\hat{d}_{ii})}$  no es normal bajo  $H_0$  de manera que el test de Wald no es directamente aplicable. Por otra parte, si se asume que la matriz de covarianzas para cada respuesta aportada por cada sujeto  $\mathbf{Y}_i$  se puede describir como  $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{n_i}$  sin asumir que esta matriz resulta de una estructura de efectos aleatorios subyacente,  $\mathbf{D}$  ya no está restringida a ser positiva definida.

### 3.10.2. Tests de razón de verosimilitud (LRT)

De manera similar a como se procede con efectos fijos, un LRT (Likelihood Ratio Test, siglas en inglés) se puede derivar para comparar modelos anidados con distintas estructuras de covarianza.

También se pueden obtener pruebas LRT válidas bajo estimación REML, en cuyo caso la función  $L_{ML}$  se reemplaza por  $L_{REML}$ . En ambos casos se tiene que el LRT está definido como:

$$-2 \log(\lambda_N) = -2 \log \left[ \frac{L_{ML}(\hat{\theta}_{ML,0})}{L_{ML}(\hat{\theta}_{ML})} \right]$$

y

$$-2 \log(\lambda_N) = -2 \log \left[ \frac{L_{REML}(\hat{\theta}_{REML,0})}{L_{REML}(\hat{\theta}_{REML})} \right]$$

Bajo ciertas condiciones de regularidad  $-2 \log(\lambda_N)$  se comporta asintóticamente como una  $\chi_{s-p}^2$  donde  $s - p$  corresponde a las diferencias de las

dimensiones de los espacios paramétricos  $\Theta_\alpha$  y  $\Theta_{\alpha,0}$ . Una de las condiciones de regularidad bajo las cuales la aproximación a una chi-cuadrado es válida es que  $H_0$  no esté en la frontera del espacio paramétrico  $\Theta_\alpha$ . Entonces el LRT sufre de exactamente los mismos problemas del test aproximado de Wald.

### 3.11. Pruebas marginales para la necesidad de efectos aleatorios

Cuando se trabaja con MLM es importante probar si los términos declarados como efectos aleatorios se necesitan dentro del modelo (por ejemplo pendientes aleatorias o términos cuadráticos aleatorios). En caso de que estos términos no sean necesarios, esto puede sugerir por ejemplo que la respuesta evoluciona linealmente en el tiempo y que cada unidad experimental no tiene ni su propio intercepto aleatorio ni término cuadrático aleatorio. Para ilustrar, suponga un modelo donde se declararon como efectos aleatorios el intercepto, la pendiente y un término cuadrático y se quiere evaluar si el término cuadrático es necesario dentro de dicho modelo. En este caso la matriz

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{pmatrix}$$

y la correspondiente hipótesis de interés es  $H_0 : d_{13} = d_{23} = d_{33} = 0$  (las entradas correspondientes a los efectos aleatorios). Esta hipótesis está en la frontera del espacio paramétrico  $\Theta_\alpha$  de manera que la teoría clásica basada en la verosimilitud no es recomendable.

Sea  $-2 \log(\lambda_N)$  el LRT. Stram y Lee (1994) discuten algunas situaciones específicas. Aunque sus resultados fueron derivados para el caso de estimación de máxima verosimilitud, los mismos resultados aplican para estimación REML tal y como se muestra en Morrell (1998). Se consideran algunos casos.

**Caso 1: Ningún efecto aleatorio versus un efecto aleatorio.** Para probar la hipótesis  $H_0 : D = 0$  vs  $H_a : D = d_{11}$  donde  $d_{11} > 0$  (un número no negativo), se tiene que la distribución nula asintótica de  $-2 \log(\lambda_N)$  es una mezcla de una  $\chi_1^2$  y una  $\chi_0^2$  con pesos de  $\frac{1}{2}$  respectivamente.  $\chi_0^2$  corresponde a una chi-cuadrado degenerada la cuál asigna una probabilidad de 1 al valor 0.

**Caso 2: Un efecto aleatorio versus dos efectos aleatorios.** Para probar

la hipótesis

$$H_0 : \mathbf{D} = \begin{pmatrix} \mathbf{d}_{11} & 0 \\ 0 & 0 \end{pmatrix}$$

donde  $d_{11} > 0$  versus la alternativa  $H_a$  que establece que  $\mathbf{D}$  es una matriz cuadrada de orden  $2 \times 2$  semidefinida positiva, se tiene que la distribución nula asintótica de  $-2 \log(\lambda_N)$  es una mezcla de una  $\chi_2^2$  y una  $\chi_1^2$  con pesos de  $\frac{1}{2}$ , respectivamente.

**Caso 3:  $q$  efectos aleatorios versus  $q+1$  efectos aleatorios.** Para probar la hipótesis

$$H_0 : \mathbf{D} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix}$$

(en la cual  $\mathbf{D}_{11}$  es una matriz cuadrada de orden  $q \times q$  definida positiva) versus  $H_a$  que establece que  $\mathbf{D}$  es una matriz cuadrada general de orden  $(q+1) \times (q+1)$  semidefinida positiva (el  $q+1$  es debido a que se incluyó un efecto aleatorio adicional en la hipótesis alternativa), el comportamiento asintótico de la distribución nula de  $-2 \log(\lambda_N)$  es una mezcla de una  $\chi_{q+1}^2$  y una  $\chi_q^2$  con pesos de  $\frac{1}{2}$  respectivamente.

**Caso 4:  $q$  efectos aleatorios versus  $q+k$  efectos aleatorios.** La distribución nula de  $-2 \log(\lambda_N)$  para probar la hipótesis

$$H_0 : \mathbf{D} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix}$$

vs

$$H_a : \mathbf{D} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}'_{12} & \mathbf{D}_{22} \end{pmatrix}$$

que establece que  $\mathbf{D}$  es una matriz cuadrada general de orden  $(q+k) \times (q+k)$  semidefinida positiva es una mezcla de variables aleatorias chi-cuadrado así como otros tipos de variables aleatorias formadas por las longitudes de las proyecciones de variables aleatorias normales multivariadas sobre superficies planas y curvas. A excepción de casos muy especiales, la distribución nula apropiada se debe obtener usando simulación.

### 3.12. Modelo autorregresivo de primer orden

Suponga que  $\{Y(y) : t = \dots, -1, 0, 1, \dots\}$  es una serie de tiempo discreta. Una serie de tiempo autorregresiva de primer orden, denotada por  $AR(1)$ ,

es generada por la ecuación

$$Y(t) = \rho Y(t-1) + \epsilon(t)$$

donde los  $\epsilon(t)$  son i.i.d. con media 0 y varianza  $\sigma_\epsilon^2$ . Los parámetros del proceso son  $\rho$  y  $\sigma_\epsilon^2$ . Aquí se presentan unos resultados bien conocidos

- $Cov[\epsilon(t), \epsilon(t-s)] = 0$  para todo  $s > 0$
- $Cov[Y(t-s), \epsilon(t)] = 0$  para todo  $s > 0$
- Si  $\rho = 0$  el proceso se llama “ruido blanco”
- Si  $\rho > 0$  el proceso oscila alrededor de  $E[Y(t)] = \mu$

Si el proceso comienza en  $Y(0) = 0$  entonces

$$\begin{aligned} Y(1) &= \epsilon(1) \\ Y(2) &= \epsilon(1) + \epsilon(2) \\ &\vdots \\ Y(t) &= \sum_{i=1}^t \epsilon(i) \end{aligned}$$

De aquí

$$Var[Y(t)] = t\sigma_\epsilon^2$$

Si se tiene el supuesto de estacionaridad de segundo orden implica que la varianza  $Var[Y(y)]$  es constante y no depende de  $t$ .

Asuma que  $E[Y(t)] = 0$ . La función de covarianza  $C(k)$ , para  $k > 0$  se define como

$$\begin{aligned} C(k) &= E[Y(t)Y(t-k)] \\ &= E[\{\rho Y(t-1) + \epsilon(t)\} Y(t-k)] \\ &= \rho E[Y(t-1)Y(t-k)] + E[\epsilon(t)\epsilon(t-k)] \\ &= \rho E[Y(t-1)Y(t-k)] + 0 \\ &= \rho C(k-1) \end{aligned}$$

De donde

$$C(k) = \rho^k C(0) = \rho^k Var[Y(t)]$$

Para hallar la  $Var [Y(t)]$  considere

$$\begin{aligned} C(0) &= E [Y(t)Y(t)] \\ &= E [\{\rho Y(t-1) + \epsilon(t)\} \{\rho Y(t-1) + \epsilon(t)\}] \\ &= \rho^2 E [Y(t-1)^2] + E [\epsilon(t)^2] \\ &= \rho^2 C(0) + \sigma_\epsilon^2 \end{aligned}$$

De donde

$$C(0) = \frac{\sigma_\epsilon^2}{1 - \rho^2}$$

La función de autocorrelación será

$$R(k) = Cor [Y(t), Y(t-k)] = \frac{C(k)}{C(0)} = \rho^k \frac{C(0)}{C(0)} = \rho^k$$

La matriz de varianzas y covarianzas de

$$\mathbf{Y}(t) = \begin{bmatrix} \vdots \\ Y(t-2) \\ Y(t-1) \\ Y(t) \end{bmatrix}$$

es

$$VarCov(\mathbf{Y}(t)) = \frac{\sigma_\epsilon^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{t-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{t-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{t-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{t-1} & \rho^{t-2} & \rho^{t-3} & \rho^{t-4} & \dots & 1 \end{bmatrix}$$

### 3.13. Pruebas de hipótesis sobre las varianzas de los efectos aleatorios

Esta sección está basada en Scheipl, Greven y Küchenhoff (2008). Verificar si la varianza de un efecto aleatorio es estadísticamente significativa es un problema importante en muchas aplicaciones. Por ejemplo, se puede tener interés en probar si hay heterogeneidad entre o inter-individuos en la media o la pendiente en estudios longitudinales. Además se puede buscar una reducción en la complejidad del modelo, lo cual amerita el desarrollo de pruebas

para esto. La prueba que se quiere estudiar es  $H_0 : \sigma_{b_i}^2 = 0$  contra la alternativa  $H_1 : \sigma_{b_i}^2 > 0$ . Este problema de pruebas de hipótesis es no estándar ya que el valor verificado por la hipótesis nula está en la frontera del espacio paramétrico.

Scheipl, Greven y Küchenhoff (2008) revisan muchas de las pruebas que hay propuestas para resolver este problema. Algunas de estas se estudian a continuación.

### 3.13.1. Prueba de la razón de la verosimilitud restringida

El estadístico de prueba es

$$RLRT = 2 \sup_{\theta \in H_1} l(\theta) - 2 \sup_{\theta \in H_0} l(\theta)$$

La distribución nula de este estadístico ha sido aproximada de diferentes maneras:

#### Aproximación de mezclas de $\chi^2$ 's

El estadístico de prueba es denotado como  $RLRT_{SL}$ , por sus autores Stram y Lee. La distribución nula es una chi-cuadrada mezclada  $0,5\chi_0^2 + 0,5\chi_1^2$ . ( $\chi_0^2$  denota una distribución degenerada en 0). Este modelo no es muy plausible para el caso de modelos mixtos, ya que se asume que el vector respuesta es representable como un conjunto muy grande de vectores independientes.

#### La aproximación de Crainiceanu y Ruppert cuando solo se tiene un efecto aleatorio

El estadístico es denotado como  $RLRT_{CR}$ . Estos autores en 2004 mostraron que

$$RLRT \stackrel{d}{=} \sup_{\lambda \geq 0} \left( (n-p) \log \left( 1 + \frac{N(\lambda)}{D(\lambda)} \right) - \sum_{s=1}^K \log(1 + \lambda \mu_s) \right)$$

donde  $\stackrel{d}{=}$  significa igualdad en distribución y

$$N(\lambda) = \sum_{s=1}^K \frac{\lambda \mu_s}{1 + \lambda \mu_s} w_s^2$$

$$D(\lambda) = \sum_{s=1}^K \frac{w_s^2}{1 + \lambda \mu_s} + \sum_{s=K+1}^{n-p} w_s^2$$

donde

- $K$  es el número de columnas de  $\mathbf{Z}$ ,
- $p$  es el número de columnas de  $\mathbf{X}$ ,
- $w_s$  son normales estándares independientes,
- $\mu_s$  son los valores propios de la matriz  $\mathbf{Z}' (\mathbf{I}_n - \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{Z}$

Ellos, vía simulación, implementan un algoritmo para realizar la prueba; este se encuentra en la librería de R llamada `RLRsim`.

### RLRT vía bootstrap

En este caso es posible calcular la distribución del estadístico de prueba y resolverlo vía remuestreo.

### Prueba de Wald para componentes de varianza

Para el modelo con un solo componente este estadístico está definido como

$$F_W = \frac{\frac{\mathbf{Y}'\mathbf{P}_b\mathbf{Y}}{\text{Traza}(\mathbf{P}_b)}}{\frac{\mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_b - \mathbf{P}_X)\mathbf{Y}}{n - \text{Traza}(\mathbf{P}_b) - \text{Traza}(\mathbf{P}_X)}}$$

bajo  $H_0 : \sigma_b^2 = 0$ , la distribución de  $F_W$  es  $F_{(\text{Traza}(\mathbf{P}_b), n - \text{traza}(\mathbf{P}_b) - \text{Traza}(\mathbf{P}_X))}$ , con matrices de proyección

- $\mathbf{P}_b = (\mathbf{X}|\mathbf{Z}) [(\mathbf{X}|\mathbf{Z})']^{-1} (\mathbf{X}|\mathbf{Z})'$  y
- $\mathbf{P}_X = \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$

donde  $\mathbf{A}^{-}$  denota la inversa generalizada de  $\mathbf{A}$  y  $(\mathbf{X}|\mathbf{Z})$  denota la concatenación de las matrices de diseño de efectos fijos y efectos aleatorios.

**Pruebas  $F$  aproximadas implementadas en SAS**

Si se considera el estadístico  $F$  para verificar una hipótesis lineal general de la forma

$$H_0 : \mathbf{L} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{pmatrix} = \mathbf{0}$$

$$F = \frac{1}{Rango(\mathbf{L})} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix}' \mathbf{L}' (\mathbf{L}\hat{\mathbf{C}}\mathbf{L}')^{-1} \mathbf{L} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} \underset{\sim}{\text{aprox.}} F_{(Rango(\mathbf{L}), \hat{\nu})}$$

con

$$\hat{\mathbf{C}} = \hat{\sigma}_\epsilon^2 \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \hat{\lambda}\mathbf{I}_K \end{bmatrix}^{-1}$$

simplemente se verifica la hipótesis  $H_0 : \mathbf{b}_l = \mathbf{0}$ , la cual es equivalente a la de interés.

**$R_{0,\lambda_0}$ : Propuesta de Cantoni y Hastie**

Estos autores adoptan una estrategia basada en splines suavizados (Cantoni y Hastie, 2002) que permite escoger entre dos alternativas lo cual podría permitir diferentes niveles de grados de libertad. Este método está adaptado para modelos lineales mixtos y produce un estimador insesgado. También proponen un estimador del tipo razón de verosimilitud cuya distribución exacta se exhibe en el artículo.

**$F_{0,\lambda}$ : Propuesta de Hastie y Tibshirani**

La propuesta de estos autores está basada en los llamados Modelos Aditivos Generalizados (Hastie y Tibshirani, 1990) los cuales constituyen una generalización de los modelos de regresión estándar, pero usando miembros de la familia exponencial. Estos modelos se estiman bajo restricciones de suavizamiento. Ellos muestran como, usando técnicas computacionales, es posible aproximar los grados de libertad de la prueba  $F$ .

**3.14. Recomendación para la construcción del modelo lineal mixto**

Robinson (2008) presenta recomendaciones de carácter general para la construcción de un modelo que incluya efectos aleatorios. Es importante tener

una estrategia clara para la construcción del modelo. El número de estrategias puede ser muy grande. Aquí se presenta una estrategia.

1. Escoja un conjunto mínimo de efectos fijos y aleatorios para el modelo.
  - a) Escoja el conjunto de efectos fijos que deben estar. Estos efectos deben ser tales que si ellos no están en el modelo, el modelo no tiene sentido.
  - b) Escoja los efectos aleatorios que deben estar. Estos efectos deben ser tales que si ellos no están en el modelo, el modelo no reflejará adecuadamente el diseño.

Este es el modelo base a partir del cual los otros modelos deben ser comparados.

2. Ajuste este modelo a los datos usando las técnicas vistas y realice los diagnósticos. Itere el proceso de mejoramiento de los efectos aleatorios incluyendo:
  - a) una estructura heteroscedástica de la varianza (con varios candidatos)
  - b) una estructura de correlación (con varios candidatos)
  - c) efectos aleatorios extras (por ejemplo, pendientes aleatorias)
3. Cuando los diagnósticos sugieran un modelo razonable, considere adicionar otros efectos fijos. En cada etapa reexamine los diagnósticos para asegurar que cualquier estimada que se utilice para determinar efectos fijos está basada en una buena concordancia entre datos, modelo y supuestos.

### 3.15. Medidas de bondad de ajuste

#### 3.15.1. Análisis de residuales

Como los residuales del modelo lineal mixto son correlacionados, ellos no pueden usarse directamente para chequear los supuestos del modelo. Sin embargo se pueden construir residuales no correlacionados que pueden obtenerse mediante transformaciones lineales, por ejemplo usando la descomposición de Cholesky de  $\mathbf{V}_i^{-1}$  (Jacquim-Gadda et al., 2007). Denote  $\mathbf{L}_i$  a la matriz

triangular obtenida mediante la descomposición de Cholesky de  $\mathbf{V}_i^{-1}$  tal que  $\mathbf{V}_i^{-1} = \mathbf{L}_i \mathbf{L}_i'$ , los llamados *Residuales Cholesky* están dados por

$$\mathbf{R}_i = \mathbf{L}_i' (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i)$$

Bajo el supuesto de que el modelo es correcto y despreciando la variabilidad de  $\hat{\boldsymbol{\beta}}$ ,  $\mathbf{R}_i \sim N(\mathbf{0}, \mathbf{I}_{n_i})$ . Jacquim-Gadda et al. (2007), usando resultados de estudios de simulación, sugieren que aún para tamaños muestrales tan pequeños como  $n = 50$  y 5 efectos fijos, la variabilidad de  $\hat{\boldsymbol{\beta}}$  tenía un efecto despreciable.

Se pueden usar métodos gráficos (gráficos Q-Q) o pruebas formales (Prueba de Shapiro-Wilks, etc.) para verificar normalidad. Puesto que estos residuales son combinaciones lineales de los residuales crudos, los cuales son específicos  $y_{ij} - \mathbf{X}_{ij} - \mathbf{X}'_{ij} \hat{\boldsymbol{\beta}}$ , ellos no pueden ser graficados contra los valores predichos o cualquier elemento de la matriz de diseño para explorar posibles malas especificaciones.

Park y Lewis (Jacquim-Gadda et al., 2007) proponen resumir todos los residuales para cada sujeto usando  $(\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i)' \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i)$  que bajo el supuesto que el modelo es correcto se distribuye  $\chi^2$  con  $n_i$  grados de libertad. Estos residuales son útiles para detectar sujetos influenciales o malas especificaciones de covariables asociadas a los sujetos, pero ellos son menos sensibles a violaciones del modelo y al supuesto de normalidad. Bajo el supuesto de un modelo homoscedástico, o sea  $\epsilon_{ij} \sim N(0, \sigma_e^2)$ , se puede chequear este supuesto usando los *residuales condicionales*  $R_{C_i}$  que se calculan como

$$\mathbf{R}_{C_i} = \mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i - \mathbf{Z}_i \hat{\mathbf{b}}_i$$

Usando las estimaciones empíricas de Bayes de los efectos aleatorios, se tiene que

$$\begin{aligned} \mathbf{R}_{C_i} &= \mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i - \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i \mathbf{V}_i' (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i) \\ &= \sigma_e^2 \mathbf{V}_i' (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i) \end{aligned}$$

y si se considera que la variabilidad de  $\hat{\boldsymbol{\beta}}$  es despreciable, la distribución de  $\mathbf{R}_{C_i}$  es  $N(\mathbf{0}, \sigma_e^4 \mathbf{V}_i^{-1})$ .

### 3.15.2. Estructuras de covarianza en el MLM

En el MLM la varianza del vector aleatorio de respuestas  $\mathbf{y}$  está dada por:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

Esta matriz  $\mathbf{V}$  puede modelarse seleccionando adecuadamente la matriz de efectos aleatorios  $\mathbf{Z}$  y especificando estructuras de covarianza para  $\mathbf{G}$  y  $\mathbf{R}$ . El modelo de efectos aleatorios simple es el caso más sencillo de este modelo donde estas estructuras están determinadas; en particular, la matriz  $\mathbf{Z}$  solo contiene variables dummies,  $\mathbf{G}$  es una matriz diagonal que contiene las componentes de varianza y  $\mathbf{R} = \sigma^2\mathbf{I}_n$ . El modelo lineal general se obtiene tomando  $\mathbf{Z} = \mathbf{0}$  y  $\mathbf{R} = \sigma^2\mathbf{I}_n$ .

En la práctica, la selección de estas estructuras de covarianza es crucial en la construcción de un MLM y una mala especificación puede tener consecuencias indeseables en las estimaciones de los efectos fijos. Lo que se acostumbra hacer es ajustar modelos con distintas especificaciones de estas estructuras y seleccionar el modelo que tenga asociado el menor AIC. Este proceso no implica que se seleccione el mejor modelo pero si garantiza que se selecciona el más óptimo entre los que se consideran. En esta sección se presentan algunas de las estructuras de covarianza para  $\mathbf{R}$  más usadas y que se encuentran disponibles en SAS. En todos los casos se muestran los elementos de la entrada  $(i, j)$  de la matriz de covarianza y el número de parámetros que se estiman.

#### Estructura Ante-dependiente

Elemento  $(i, j)$  de la matriz de covarianza:

$$\sigma_i\sigma_j \prod_{k=i}^{j-1} \rho_k$$

Número de parámetros de covarianza en la estructura:  $2t - 1$ , donde  $t$  es la dimensión de la matriz de covarianza,  $\sigma_i^2$  es el  $i$ -ésimo parámetro de varianza y  $\rho_k$  es el  $k$ -ésimo parámetro de autocorrelación que satisface  $|\rho_k| < 1$ . En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=ANTE(1). Esta estructura especifica una estructura de ante-dependencia de primer orden (Kenward 1987, Patel 1991, Macchiavelli y Arnold, 1994)

**Estructura autoregresiva de primer orden**

Elemento  $(i, j)$  de la matriz de covarianza:

$$\sigma^2 \rho^{|i-j|}$$

Número de parámetros de covarianza en la estructura: 2. En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=AR(1). PROC MIXED impone la restricción de estacionaridad  $|\rho| < 1$ .

**Estructura auto-regresiva heterogénea de primer orden**

Elemento  $(i, j)$  de la matriz de covarianza:

$$\sigma_i \sigma_j \rho^{|i-j|}$$

Número de parámetros de covarianza en la estructura:  $t + 1$ , donde  $t$  es la dimensión de la matriz de covarianza. En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=ARH(1). PROC MIXED impone la restricción de estacionaridad  $|\rho| < 1$ .

**Estructura auto-regresiva de medias móviles de primer orden**

Elemento  $(i, j)$  de la matriz de covarianza:

$$\sigma^2 \left[ \gamma \rho^{|i-j|-1} I_{(i \neq j)} + I_{(i=j)} \right]$$

Número de parámetros de covarianza en la estructura: 3.  $\rho$  es el parámetro auto-regresivo,  $\gamma$  modela una componente de medias móviles y  $\sigma^2$  es la varianza residual. En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=ARMA(1, 1). PROC MIXED impone las restricciones para estacionaridad  $|\rho| < 1$  y  $|\gamma| < 1$ , aunque para algunos valores de  $\rho$  y  $\gamma$  en esta región la matriz de varianzas y covarianzas resultante no es definida positiva. Cuando el valor estimado de  $\rho$  es negativo, la matriz de varianzas y covarianzas resultante se multiplica por  $\cos(\pi d_{ij})$  para tener en cuenta esta negatividad ( $d_{ij}$  es la distancia euclídea entre los vectores  $i$  y  $j$  que corresponden a las observaciones  $i$  y  $j$  en la base de datos).

**Estructura de simetría compuesta**

La estructura de simetría compuesta tiene varianza y covarianza constantes. Elemento  $(i, j)$  de la matriz de covarianza:

$$\sigma_1 + \sigma^2 I_{(i=j)}$$

Número de parámetros de covarianza en la estructura: 2. En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=CS. Matricialmente:

$$CS = \begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \cdots & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \cdots & \sigma_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_1 & \sigma_1 & \sigma_1 & \cdots & \sigma^2 + \sigma_1 \end{bmatrix}$$

Una fórmula matricial más precisa es:

$$\sigma^2 \mathbf{I} + \sigma_1 \mathbf{J}$$

donde  $\mathbf{I}$  es la matriz identidad y  $\mathbf{J}$  es una matriz de unos. Observe que  $\mathbf{J} = \mathbf{1}\mathbf{1}'$  donde  $\mathbf{1}$  es un vector de unos.

**Estructura de simetría compuesta heterogénea**

La estructura de simetría compuesta heterogénea tiene un parámetro de varianza distinto para cada elemento de la diagonal y usa la raíz cuadrada de estos parámetros en los elementos fuera de la diagonal.

Elemento  $(i, j)$  de la matriz de covarianza:

$$\sigma_i \sigma_j [\rho I_{(i \neq j)} + I_{(i=j)}]$$

donde  $\sigma_i^2$  es el parámetro correspondiente a la  $i$ -ésima varianza y  $\rho$  es el parámetro de correlación que satisface  $|\rho| < 1$ . Número de parámetros de covarianza en la estructura:  $t + 1$ . En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando

TYPE=CSH. Matricialmente:

$$CSH = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \cdots & \sigma_1\sigma_t\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \cdots & \sigma_2\sigma_t\rho \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_t\sigma_1\rho & \sigma_t\sigma_2\rho & \sigma_t\sigma_3\rho & \cdots & \sigma_t^2 \end{bmatrix}$$

( $t$  es la dimensión de la matriz de covarianza)

### Estructura de factor analítico

Elemento ( $i, j$ ) de la matriz de covarianza:

$$\sum_{k=1}^{\min(i,j,q)} \lambda_{ik}\lambda_{jk} + \sigma_i^2 I_{(i=j)}$$

donde los  $\lambda$  son elementos de la matriz  $\mathbf{\Lambda}$  que es una matriz rectangular de orden  $t \times q$  ( $q$  es el número de factores). Esta estructura es de la forma  $\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{D}$  donde  $\mathbf{\Lambda}$  es una matriz rectangular de orden  $t \times q$ ,  $\mathbf{D}$  es una matriz diagonal de orden  $t \times t$  con  $t$  parámetros distintos. Cuando  $q > 1$  los elementos de  $\mathbf{\Lambda}$  en su esquina superior derecha (es decir, los elementos en la  $i$ -ésima fila y  $j$ -ésima columna para  $j > i$ ) se definen como cero para ajustar la rotación de la estructura. Número de parámetros de covarianza en la estructura:  $\frac{q}{2}(2t - q + 1) + t$  ( $t$  es la dimensión de la matriz de covarianza). En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=FA( $q$ ).

### Estructura de factor analítico no diagonal

Elemento ( $i, j$ ) de la matriz de covarianza:

$$\sum_{k=1}^{\min(i,j,q)} \lambda_{ik}\lambda_{jk}$$

Es similar a la estructura anterior  $FA(q)$  excepto que no se incluye la matriz diagonal  $\mathbf{D}$ . Cuando  $q < t$  (esto es cuando el número de factores  $q$  es menor que la dimensión de la matriz) esta estructura es definida no negativa pero no de rango completo. En esta situación ella se puede usar para aproximar una matriz sin estructura usando la declaración RANDOM o combinando con la

opción LOCAL en la declaración REPEATED. Cuando  $q = 1$  se puede usar esta estructura para restringir  $\mathbf{G}$  a ser definida no negativa en la declaración RANDOM del PROC MIXED. El número de parámetros estimados es:  $\frac{q}{2}(2t - q + 1)$ . En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=FA0(q).

### Estructura de factor analítico con diagonal igual

Elemento  $(i, j)$  de la matriz de covarianza:

$$\sum_{k=1}^{\min(i,j,q)} \lambda_{ik}\lambda_{jk} + \sigma^2 I_{(i=j)}$$

donde los  $\lambda$  son elementos de la matriz  $\mathbf{\Lambda}$  que es una matriz rectangular de orden  $t \times q$ . Esta estructura es de la forma  $\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{D}$  donde  $\mathbf{\Lambda}$  es una matriz rectangular de orden  $t \times q$ ,  $\mathbf{D}$  es una matriz diagonal de orden  $t \times t$  con  $t$  parámetros iguales. Esta estructura es similar a la FA(q) pero los elementos de la diagonal en  $\mathbf{D}$  son iguales. Esto ofrece una alternativa más útil y parsimoniosa que la estructura completa de factor analítico. Número de parámetros de covarianza en la estructura:  $\frac{q}{2}(2t - q + 1) + 1$  ( $t$  es la dimensión de la matriz de covarianza). En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=FA1(q).

### Covarianza sin estructura

Elemento  $(i, j)$  de la matriz de covarianza:

$$\sigma_{ij}$$

Número de parámetros de covarianza en la estructura:  $t(t + 1)/2$  ( $t$  es la dimensión de la matriz de covarianza). Esta covarianza corresponde a una matriz de covarianzas completamente general (sin estructura) especificada directamente en términos de varianzas y covarianzas. Las varianzas se restringen a ser no negativas mientras que las covarianzas no tienen restricción. Esta covarianza no está restringida a ser definida no negativa para evitar restricciones lineales; sin embargo, se puede usar la estructura FA0 si se necesita

imponer una factorización de Cholesky. Matricialmente:

$$UN = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1t} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{t1} & \sigma_{t2} & \sigma_{t3} & \cdots & \sigma_t^2 \end{bmatrix}$$

En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=UN.

### Estructura Huynh-Feldt

Elemento  $(i, j)$  de la matriz de covarianza:

$$\frac{\sigma_i^2 + \sigma_j^2}{2} - \lambda I_{(i \neq j)}$$

Número de parámetros de covarianza en la estructura:  $t + 1$  ( $t$  es la dimensión de la matriz de covarianza). En el PROC MIXED del SAS se especifica como opción en las declaraciones RANDOM o REPEATED usando TYPE=HF. Matricialmente:

$$HF = \begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \cdots & \frac{\sigma_1^2 + \sigma_t^2}{2} - \lambda \\ \frac{\sigma_2^2 + \sigma_1^2}{2} - \lambda & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \cdots & \frac{\sigma_2^2 + \sigma_t^2}{2} - \lambda \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\sigma_t^2 + \sigma_1^2}{2} - \lambda & \frac{\sigma_t^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_t^2 + \sigma_3^2}{2} - \lambda & \cdots & \sigma_t^2 \end{bmatrix}$$

Esta estructura es similar a la estructura de simetría compuesta heterogénea (CSH) en el sentido de que tiene el mismo número de parámetros y heterogeneidad a lo largo de su diagonal principal. Sin embargo, los elementos fuera de la diagonal principal se obtienen al tomar promedios aritméticos en vez de promedios geométricos. Es posible obtener un LRT para evaluar las condiciones de Huynh-Feldt al ejecutar el PROC MIXED dos veces, una especificando TYPE=HF y otra especificando TYPE=UN y después sustrayendo los respectivos valores de -2 veces la verosimilitud obtenida. En caso de que el modelo no converja bajo el modelo de Huynh-Feldt, es posible especificar un conjunto de valores iniciales con la opción PARMS del SAS. El método default MIVQUE(0)<sup>8</sup> algunas veces proporciona conjuntos de valo-

<sup>8</sup>Minimum Variance Quadratic Unbiased Estimates with 0 a priors (Goodnight 1978)

res iniciales muy pobres para esta estructura. Una buena opción para valores iniciales consiste en seleccionar los parámetros estimados correspondientes al ajuste inicial usando TYPE=CS.

### Estructura de componentes de varianza

Elemento  $(i, j)$  de la matriz de covarianza:

$$\sigma_k^2 I_{(i=j)} \quad (i \text{ corresponde al } k\text{-ésimo efecto})$$

Número de parámetros de covarianza en la estructura:  $q$  (número de factores). Con esta estructura se especifican componentes de varianza estándar y es el default en las declaraciones RANDOM y REPEATED. En la declaración RANDOM, se especifica una componente de varianza distinta a cada efecto. En la declaración REPEATED, esta estructura frecuentemente se usa para especificar un modelo de varianza heterogéneo.

Jennrich y Schluchter (1986) proporcionan información general acerca del uso de estructuras de covarianza y Wolfinger (1996) presenta detalles acerca de muchas de las estructuras heterogéneas.

## 3.16. Predicción

### 3.16.1. Mejor predictor lineal insesgado

La estimación de las componentes de varianza es usualmente de interés. Una de las características que distinguen al modelo lineal mixto del modelo lineal clásico es su habilidad para estimar o más bien predecir realizaciones de variables aleatorias. Esto es posible usando los llamados *Mejores Predictores Lineales Insesgados*<sup>9</sup> (Best Linear Unbiased Predictor, BLUP en inglés). Una de las formas básica del MLM está dada por:

$$Y_j = \sum_i \beta_i X_{ji} + \sum_k b_k Z_{jk} + e_j$$

donde:

- $Y_j$  es la  $j$ -ésima observación.
- $\beta_i$  son parámetros de efectos fijos.

---

<sup>9</sup>Se denotarán según su abreviación en inglés, es decir BLUP

- $X_{ji}$  son constantes asociadas a los efectos fijos.
- $b_k$  son los efectos aleatorios.
- $Z_{jk}$  son constantes asociadas a los efectos aleatorios.
- $e_j$  es el  $j$ -ésimo error residual.

Este MLM puede escribirse de manera alternativa (más compacta) como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}$$

En esta notación, el valor esperado de una observación está dado por:

$$E[\mathbf{Y}] = E[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}] = \mathbf{X}\boldsymbol{\beta}$$

ya que los valores esperados para  $\mathbf{b}$  y  $\mathbf{e}$  son 0. Este valor esperado se denomina *esperanza incondicional* o la *Media de  $\mathbf{Y}$  promediada sobre todos los posibles valores de  $\mathbf{b}$* . En términos prácticos, significa que los niveles observados de los efectos aleatorios constituyen una muestra aleatoria de una población grande. En otras palabras, la esperanza incondicional es la media de  $\mathbf{Y}$  sobre toda la población. Por otra parte, la *esperanza condicional* de  $\mathbf{Y}$  dado  $\mathbf{b}$ , que se denota  $E[\mathbf{Y}|\mathbf{b}]$  es:

$$E[\mathbf{Y}|\mathbf{b}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$$

En términos prácticos, esta es la media de  $\mathbf{Y}$  para un conjunto específico de niveles del efecto aleatorio que en realidad se observó.

De esta manera, la media incondicional es un promedio basado en toda la población mientras la media condicional es un promedio específico a un conjunto observado de efectos aleatorios. Debido a que el conjunto de niveles observados de los efectos aleatorios no es un duplicado exacto de toda la población, las medias condicional e incondicional en general no son iguales.

La inferencia en el MLM se basa en combinaciones lineales de los efectos fijos. Estas combinaciones de los efectos fijos, que están dadas por  $\sum_i \mathbf{K}_i \boldsymbol{\beta}_i$  se denominan *funciones estimables* si se pueden construir a partir de una combinación lineal de medias incondicionales de las observaciones. Es decir, si

$$\begin{aligned} \mathbf{K}'\boldsymbol{\beta} &= \mathbf{T}'E[\mathbf{Y}] \\ &= \mathbf{T}'\mathbf{X}\boldsymbol{\beta} \quad \text{para algún } \mathbf{T} \end{aligned}$$

entonces es estimable. Cantidades tales como coeficientes de regresión, medias de tratamientos, diferencias entre tratamientos, contrastes son ejemplos de funciones estimables. Las funciones estimables tienen la característica de que no dependen de los efectos aleatorios.

En casos donde se consideran combinaciones lineales de  $\beta_j$  y  $\mathbf{b}_j$  se requiere una generalización de las funciones estimables. Combinaciones lineales de efectos fijos y aleatorios tales como  $\mathbf{K}'\beta + \mathbf{M}'\mathbf{b}$ , se pueden formar a partir de combinaciones de las medias condicionales. Estas combinaciones lineales se denominan *funciones predecibles*. Una función  $\mathbf{K}'\beta + \mathbf{M}'\mathbf{b}$  se dice que es una *Función predecible* si su componente  $\mathbf{K}'\beta$  es estimable.

Las ecuaciones del modelo mixto de Henderson se especifican como:

$$\begin{aligned} \begin{bmatrix} \tilde{\beta} \\ \tilde{\mathbf{b}} \end{bmatrix} &= \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{GZ}'\mathbf{V}^{-1} \left( \mathbf{y} - \mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \right) \end{bmatrix} \end{aligned}$$

Usando la solución para  $\beta$  en una función estimable produce lo que se denomina *mejor estimador lineal insesgado* (BLUE) de  $\mathbf{K}'\beta$ . Para funciones predecibles, las soluciones para  $\beta$  y  $\mathbf{b}$  proporcionan el denominado *mejor predictor lineal insesgado* (BLUP) de  $\mathbf{K}'\beta + \mathbf{M}'\mathbf{b}$ . El BLUP puede pensarse como una regresión hacia la media global basada en las componentes de varianza de los efectos del modelo. Estrictamente hablando, BLUP asume que las componentes de varianza son conocidas. Puesto que en la práctica se usan estimaciones de las componentes de varianza, el predictor resultante no es un verdadero BLUP. El predictor resultante se denomina *mejor predictor lineal insesgado estimado* (EBLUP).

### Observaciones futuras

A menos que se use inferencia robusta, un modelo basado en una estructura de covarianza apropiada es esencial para obtener inferencias válidas para los parámetros involucrados en la estructura de la media. Especificaciones demasiado restrictivas invalidan las inferencias cuando la estructura que se asume no es adecuada mientras que sobre-parametrizar la estructura de covarianza conduce a estimaciones ineficientes. Además la estructura de covarianza puede ser de interés para entender la variación aleatoria observada en los datos. Sin embargo, puesto que ella solo explica variabilidad no explicada por ten-

dencias sistemáticas, es altamente dependiente de la estructura especificada para la media (Verbeke y Molenberghs (2000)).

Una estructura de covarianza bien especificada produce mejores predicciones, por ejemplo, la predicción de una observación futura  $y_i^*$  para el sujeto  $i$ . Los valores predichos se forman usando las filas de la matriz  $(\mathbf{X} \ \mathbf{Z})$  como matrices de combinaciones lineales  $\mathbf{L}$ . Así, los valores predichos a partir de los datos originales son

$$\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{b}}$$

Sus respectivos errores estándar se obtiene a partir de  $\mathbf{L}$  y  $\hat{\mathbf{C}}^{10}$  donde

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix}$$

Esta es la *Matriz de varianzas y covarianzas estimada* de

$$\left( \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\mathbf{b}} - \mathbf{b} \right)$$

Muchas de las técnicas estudiadas en este capítulo se ilustran en el capítulo siguiente por medio de programas en R. Específicamente se usa la librería *nmls* la cual permite especificar distintas estructuras de covarianza, permite realizar pruebas de hipótesis sobre las componentes de varianza y sus algoritmos de estimación están basados en los desarrollos teóricos que se discutieron en este capítulo.

---

<sup>10</sup>Si  $\mathbf{G}$  y  $\mathbf{R}$  se conocen,  $\hat{\boldsymbol{\beta}}$  es el BLUE de  $\boldsymbol{\beta}$  y  $\hat{\mathbf{b}}$  es el BLUP de  $\mathbf{b}$

# Capítulo 4

## Ejemplos

En este capítulo se ilustran algunas de las técnicas estudiadas en capítulos anteriores. Los ajustes de los distintos modelos se hacen por medio de la librería *nmle* disponible en el software R.

### 4.1. Electroplateado

Este ejemplo muestra el caso del submuestreo (o subreplicación), situación que ocurre con frecuencia en problemas de diseño experimental<sup>1</sup>. También ocurre por ejemplo en problemas donde el investigador mide varias veces la misma unidad experimental, por ejemplo en antropometría, donde es costumbre tomar al menos tres mediciones de cada variable de interés al mismo sujeto. Otro ejemplo, incluye un estudio para evaluar la distribución de azúcar dentro de una bebida carbonatada; si esta está distribuida uniformemente dentro del envase y se toman 4 muestras, cada una de estas generará valores muy similares y por lo tanto se cuestionará si es necesario hacer las 4 mediciones y se podría optar por realizar solamente una medición.

Aquí, cada lámina es sometida a un proceso de electroplateado de entre tres posibles procesos (tratamientos). Luego, la lámina es medida en tres

---

<sup>1</sup>En algunos casos, las submuestras se tratan como si fueran independientes cuándo en realidad no lo son. Esto conlleva a inflar los test estadísticos y por ende disminuir los valores-p y así conducir a conclusiones erróneas. Por ejemplo, considere un experimento con peces. En un tanque se adiciona un químico y se monitorea el crecimiento de los peces dentro de él. La unidad experimental es el tanque y las unidades que se observan son los peces, pero no todos los peces dentro del tanque son independientes

puntos escogidos al azar. Aunque los datos tomados dentro de cada lámina son independientes, ellos no pueden ser tratados como si vinieran de láminas diferentes. Estas mediciones sirven para mejorar la estimación del grosor del electroplateado en una lámina particular. En R se pueden analizar estos datos. A continuación se muestra un código en R que permite trabajar con estos datos.

```
> # datos de electoplateado
>
> Y<-c(
+ 0.52,0.49,0.49, 0.53,0.51,0.50, 0.52,0.49,0.51,
+ 0.51,0.53,0.52, 0.54,0.52,0.50, 0.52,0.51,0.52,
+ 0.54,0.51,0.53, 0.54,0.56,0.54, 0.53,0.55,0.52,
+ 0.52,0.53,0.53, 0.54,0.54,0.53, 0.52,0.51,0.54,
+ 0.56,0.57,0.57, 0.58,0.55,0.57, 0.57,0.58,0.57,
+ 0.56,0.58,0.56, 0.55,0.57,0.58, 0.57,0.54,0.56)
>
> Trat<-c(rep(c('I','II','III'),each=18))
> Subject<-rep(1:18,each=3)
> X1<-ifelse(Trat=='I',1,0)
> X2<-ifelse(Trat=='II',1,0)
> X3<-ifelse(Trat=='III',1,0)
```

Se ajustarán dos modelos (que en el fondo son un solo modelo), pero que se diferencian en la parametrización que el programa realiza. Para el primer modelo se especifican tres variables dicótomas,  $X_1$ ,  $X_2$  y  $X_3$  para especificar los tratamientos. El modelo que se obtiene en este caso es

$$y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + b_{0i} + \epsilon_{ij}$$

El  $\beta_0$  corresponde al efecto del Tratamiento III. El otro modelo que se ajusta tiene un factor *Tratamiento*, y observe que no se le especifica coeficiente, ya que como es un factor, el número de coeficientes depende del número de niveles. En el resultado que muestra el programa, el valor asociado con el intercepto corresponde al nivel I del Tratamiento.

$$y_{ij} = \text{Tratamiento} + b_{0i} + \epsilon_{ij}$$

A fin de ajustar estos dos modelos se puede usar el siguiente código R (en 'modelo' aparece ajustado el modelo con las variable dicótomas y en 'modelo2' el ajustado con el factor '*Tratamiento*')

```
> library(nlme)
```

```

> Electro<-data.frame(Y=Y,Trat=Trat,X1=X1,X2=X2,X3=X3,Subject=Subject)
> #rm(Y,Trat,Subject)
> Electro.new <- groupedData( Y ~ X1+X2+X3 | Subject,
+                             data = as.data.frame( Electro ) )
>
> modelo<-lme(Y~X1+X2, data=Electro.new,random=~1)
> modelo2<-lme(Y~Trat,data=Electro.new,random=~1)
>
> summary(modelo)
Linear mixed-effects model fit by REML
Data: Electro.new
      AIC      BIC   logLik
-279.0863 -269.4272 144.5432

Random effects:
Formula: ~1 | Subject
      (Intercept)  Residual
StdDev: 1.869839e-06 0.01306094

```

Recordando los supuestos y la notación usada previamente, para este caso se tiene

$$\epsilon_i \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma_\epsilon^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

Esta es la estructura de la matriz de varianzas y covarianzas del error por defecto, además, para este problema  $VarCov(\mathbf{b}) = \mathbf{D}$  será un escalar  $\mathbf{D} = \sigma_{b_0}^2$ .

Hasta aquí se obtiene la información sobre  $\hat{\sigma}_\epsilon = 0,01306094$  y sobre  $\hat{\sigma}_{b_0} = 1,869839e - 06$ . En este paso se observa que se puede pensar en verificar una hipótesis del estilo  $H_0 : \sigma_b^2 = 0$ . Si no se rechaza esta hipótesis, el submuestreo aporta muy poco al proceso de estimación y se puede trabajar con un modelo de efectos fijos simplemente. De acuerdo al valor tan pequeño para  $\hat{\sigma}_{b_0}$  parece que no es necesario especificar un efecto aleatorio (no es necesario realizar el submuestreo). Este ha sido un tema importante en diseño experimental.

```

Fixed effects: Y ~ X1 + X2
      Value   Std.Error DF   t-value p-value
(Intercept) 0.5661111 0.003078494 36 183.89223    0
X1          -0.0533333 0.004353648 15 -12.25026    0
X2          -0.0338889 0.004353648 15  -7.78402    0

```

En la parte de arriba se muestra la información de los efectos fijos, sus desviaciones estándares, los grados de libertad y los valores  $t$  calculados junto con los valores  $p$  (observe la significancia de los tratamientos I y II que tienen asociados valores- $p < 0.001$ ). A continuación, el programa entrega la matriz de correlaciones entre los efectos fijos asociados con las variables en el modelo.

```
Correlation:
  (Intr) X1
X1 -0.707
X2 -0.707  0.500
```

Se observa una moderada asociación lineal entre  $X_1$  y  $X_2$ .

```
Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.9991750 -0.4678920  0.2977494  0.5954989  2.1267819
```

```
Number of Observations: 54
Number of Groups: 18
```

```
> random.effects(modelo)
  (Intercept)
1 -7.856623e-10
3 -3.757515e-10
6  2.391146e-10
2  3.415923e-11
4  4.440700e-10
10 -3.415923e-10
5  4.440700e-10
7 -3.415923e-10
11  2.732738e-10
12 -5.465477e-10
9  6.831846e-11
8  8.881400e-10
13  3.415923e-11
18 -5.807069e-10
14  3.415923e-11
15  4.440700e-10
16  3.415923e-11
17  3.415923e-11
```

Estos valores son prácticamente iguales a cero.

```
> fixed.effects(modelo)
```

```

(Intercept)          X1          X2
 0.56611111 -0.05333333 -0.03388889
> fitted(modelo)
      1      1      1      2      2      2      3      3
0.5127778 0.5127778 0.5127778 0.5127778 0.5127778 0.5127778 0.5127778 0.5127778
      3      4      4      4      5      5      5      6
0.5127778 0.5127778 0.5127778 0.5127778 0.5127778 0.5127778 0.5127778 0.5127778
      6      6      7      7      7      8      8      8
0.5127778 0.5127778 0.5322222 0.5322222 0.5322222 0.5322222 0.5322222 0.5322222
      9      9      9      10     10     10     11     11
0.5322222 0.5322222 0.5322222 0.5322222 0.5322222 0.5322222 0.5322222 0.5322222
     11     12     12     12     13     13     13     14
0.5322222 0.5322222 0.5322222 0.5322222 0.5661111 0.5661111 0.5661111 0.5661111
     14     14     15     15     15     16     16     16
0.5661111 0.5661111 0.5661111 0.5661111 0.5661111 0.5661111 0.5661111 0.5661111
     17     17     17     18     18     18
0.5661111 0.5661111 0.5661111 0.5661111 0.5661111 0.5661111
attr(,"label")
[1] "Fitted values"

```

Para calcular los residuales se usa la función `residuals()`. Por defecto se entregan los residuales de Pearson, que son los residuales estandarizados, pero se pueden calcular otros residuales como los residuales crudos usando la opción `type='response'` (este es el 'default'), o residuales normalizados si se usa la opción `type='normalized'`, estos últimos residuales se calculan como residuales estandarizados pre-multiplicados por la matriz que es la inversa de la raíz cuadrada de la factorización de la estimada de la matriz de correlación, en algunos programas se conocen como *Residuales Cholesky*.

```

> residuals(modelo)
      1      1      1      2      2      2
0.007222223 -0.022777777 -0.022777777 0.017222222 -0.002777778 -0.012777778
      3      3      3      4      4      4
0.007222223 -0.022777777 -0.002777777 -0.002777778 0.017222222 0.007222222
      5      5      5      6      6      6
0.027222222 0.007222222 -0.012777778 0.007222222 -0.002777778 0.007222222
      7      7      7      8      8      8
0.007777778 -0.022222222 -0.002222222 0.007777777 0.027777777 0.007777777
      9      9      9     10     10     10
-0.002222222 0.017777778 -0.012222222 -0.012222222 -0.002222222 -0.002222222
     11     11     11     12     12     12
0.007777778 0.007777778 -0.002222222 -0.012222222 -0.022222222 0.007777778
     13     13     13     14     14     14
-0.006111111 0.003888889 0.003888889 0.013888889 -0.016111111 0.003888889

```

```

      15      15      15      16      16      16
0.003888888 0.013888888 0.003888888 -0.006111111 0.013888889 -0.006111111
      17      17      17      18      18      18
-0.016111111 0.003888889 0.013888889 0.003888889 -0.026111111 -0.006111111
attr(,"label")
[1] "Residuals"
> residuals(modelo,type='response')
      1      1      1      2      2      2
0.007222223 -0.022777777 -0.022777777 0.017222222 -0.002777778 -0.012777778
      3      3      3      4      4      4
0.007222223 -0.022777777 -0.002777777 -0.002777778 0.017222222 0.007222222
      5      5      5      6      6      6
0.027222222 0.007222222 -0.012777778 0.007222222 -0.002777778 0.007222222
      7      7      7      8      8      8
0.007777778 -0.022222222 -0.002222222 0.007777777 0.027777777 0.007777777
      9      9      9     10     10     10
-0.002222222 0.017777778 -0.012222222 -0.012222222 -0.002222222 -0.002222222
     11     11     11     12     12     12
0.007777778 0.007777778 -0.002222222 -0.012222222 -0.022222222 0.007777778
     13     13     13     14     14     14
-0.006111111 0.003888889 0.003888889 0.013888889 -0.016111111 0.003888889
     15     15     15     16     16     16
0.003888888 0.013888888 0.003888888 -0.006111111 0.013888889 -0.006111111
     17     17     17     18     18     18
-0.016111111 0.003888889 0.013888889 0.003888889 -0.026111111 -0.006111111
attr(,"label")
[1] "Residuals"
> residuals(modelo,type='normalized')
      1      1      1      2      2      2      3
0.5529634 -1.7439611 -1.7439611 1.3186048 -0.2126782 -0.9783197 0.5529633
      3      3      4      4      4      5      5
-1.7439612 -0.2126782 -0.2126782 1.3186048 0.5529633 2.0842463 0.5529633
      5      6      6      6      7      7      7
-0.9783197 0.5529633 -0.2126782 0.5529633 0.5954990 -1.7014255 -0.1701425
      8      8      8      9      9      9     10
0.5954989 2.1267819 0.5954989 -0.1701426 1.3611404 -0.9357841 -0.9357840
     10     10     11     11     11     12     12
-0.1701425 -0.1701425 0.5954989 0.5954989 -0.1701426 -0.9357840 -1.7014255
     12     13     13     13     14     14     14
0.5954990 -0.4678920 0.2977495 0.2977495 1.0633910 -1.2335335 0.2977495
     15     15     15     16     16     16     17
0.2977494 1.0633909 0.2977494 -0.4678920 1.0633910 -0.4678920 -1.2335335
     17     17     18     18     18
0.2977495 1.0633910 0.2977495 -1.9991750 -0.4678920
attr(,"label")

```

```
[1] "Standardized residuals"
>
```

Para obtener intervalos de confianza para los parámetros de interés en el modelo se dispone de la función `intervals()`. Esta función se aplica al objeto con el que se ha venido trabajando.

```
> intervals(modelo)
Approximate 95% confidence intervals

Fixed effects:
      lower      est.      upper
(Intercept) 0.55986764 0.56611111 0.57235459
X1          -0.06261291 -0.05333333 -0.04405375
X2          -0.04316847 -0.03388889 -0.02460931
attr("label")
[1] "Fixed effects:"

Random Effects:
Level: Subject
      lower      est.      upper
sd((Intercept)) 4.36919e-171 1.869839e-06 8.002164e+158
```

El límite superior tan enorme es debido al valor tan bajo para  $\hat{\sigma}_{b_0}$  y esto dice que no se necesita especificar el efecto aleatorio dentro del modelo y se confirma la no necesidad del submuestreo.

```
Within-group standard error:
      lower      est.      upper
0.01075704 0.01306094 0.01585828
> summary(modelo2)
Linear mixed-effects model fit by REML
Data: Electro.new
      AIC      BIC  logLik
-279.0863 -269.4272 144.5432

Random effects:
Formula: ~1 | Subject
      (Intercept)  Residual
StdDev: 1.869870e-06 0.01306094

Fixed effects: Y ~ Trat
      Value  Std.Error DF  t-value p-value
(Intercept) 0.5127778 0.003078494 36 166.56774 0e+00
TratII      0.0194444 0.004353648 15  4.46624 5e-04
```

```

TratIII      0.0533333 0.004353648 15  12.25026  0e+00
Correlation:
  (Intr) TratII
TratII      -0.707
TratIII     -0.707  0.500

```

Se observa la significancia estadística de TratII y TratIII (valores- $p < 0.001$ ) y la moderada asociación lineal entre estos tratamientos. R no entrega la estimación para TratI ya que este es el nivel de referencia.

```

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.9991750 -0.4678920  0.2977494  0.5954989  2.1267819

```

```

Number of Observations: 54
Number of Groups: 18

```

Otra forma de solucionar este problema es por medio del método GEE (Generalized Estimating Equations) propuesto por Liang y Zeger(1986)<sup>2</sup>

```

>
> library(gee)
> modelo3<-gee(Y~as.factor(Trat), Subject)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
      (Intercept) as.factor(Trat)II as.factor(Trat)III
      0.51277778      0.01944444      0.05333333
> summary(modelo3)

```

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA

```

Model:
Link: Identity
Variance to Mean Relation: Gaussian
Correlation Structure: Independent

```

```

Call:
gee(formula = Y ~ as.factor(Trat), id = Subject)

```

```

Summary of Residuals:
      Min      1Q      Median      3Q      Max

```

---

<sup>2</sup>Este método se usa con frecuencia para estimar los parámetros de un modelo lineal generalizado cuando hay una posible estructura de correlación desconocida entre los resultados.

```
-0.026111111 -0.006111111 0.003888889 0.007777778 0.027777778
```

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	0.51277778	0.003078494	166.567747	0.002983149	171.891450
as.factor(Trat)II	0.01944444	0.004353648	4.466242	0.004380325	4.439042
as.factor(Trat)III	0.05333333	0.004353648	12.250264	0.003586096	14.872256

Estimated Scale Parameter: 0.0001706

Number of Iterations: 1

Working Correlation

	[,1]	[,2]	[,3]
[1,]	1	0	0
[2,]	0	1	0
[3,]	0	0	1

>

A diferencia de los modelos anteriores, aquí el efecto de TratIII tiene asociado una significancia de borde (valor-p=0.0533). Este método entrega valores-p mayores al método anterior, pero es más recomendable cuando las respuestas son categóricas y no existe un interés directo en efectos aleatorios sino de inferencia a nivel de la población.

## 4.2. Toros

Este ejemplo corresponde al porcentaje de inseminaciones artificiales exitosas usando semen de 6 toros tomados al azar de una cierta población (hay 35 datos). De cada eyaculación de un toro se producen múltiples pajillas que posteriormente se usan para fertilizar vacas. Se detalla el programa en R que se usó para ajustar un modelo a estos datos (ver apéndice). Puesto que cada toro tiene su propia distribución, es de interés estimar el valor promedio de la población y esto se logra con un modelo de solo intercepto, de esta manera se podría predecir el porcentaje de inseminaciones artificiales asociadas a cada toro. A continuación se detalla la forma en que este modelo se puede ajustar usando el software R.

```
# Ejemplo de la proporción de inseminaciones artificiales exitosas por
# eyaculación de 6 toros tomados al azar de una población
```

```
y=c(.46,.31,.37,.62,.30,
```

```

.70, .59,
.52, .44, .57, .40, .67, .64, .70,
.47, .21, .70, .46, .14,
.42, .64, .50, .69, .77, .81, .87,
.35, .68, .59, .38, .57, .76, .57, .29, .60)
nro.med<-c(5,2,7,5,7,9)

# Función que repite cada elemento un número diferente de veces
rep.1<-function(x,r){
  if(length(x)!=length(r))stop('Vectores deben tener el mismo tamaño.')
  res<-NULL
  for(i in 1:length(x)) res<-c(res,rep(x[i],r[i]))
  res
}

ident<-rep.1(1:6,nro.med)

Toros<-data.frame(y=y,ident=ident)

library(nlme)

Toros.new <- groupedData( y ~ 1 | ident,
                          data = as.data.frame( Toros ) )

modelo<-lme(y~1, data=Toros.new,random=~1)
summary(modelo)
Linear mixed-effects model fit by REML
Data: Toros.new
      AIC      BIC   logLik
-14.64709 -10.06801 10.32355

Random effects:
Formula: ~1 | ident
      (Intercept) Residual
StdDev:  0.08764421 0.1577036

A diferencia del ejemplo anterior, el error estándar asociado al intercepto aleatorio
ya no es tan pequeño y esto de alguna manera confirma la necesidad de especificar
esta componente en la formulación del modelo.

Fixed effects: y ~ 1
              Value Std.Error DF t-value p-value
(Intercept) 0.5331792 0.04557203 29 11.6997      0

```

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-1.9651811	-0.8792900	0.1048080	0.7629851	1.5857839

Number of Observations: 35

Number of Groups: 6

```
> plot(modelo)
```

```
> qqnorm(modelo)
```

```
> plot(modelo, ident ~ resid(.))
```

```
> plot(modelo, resid(., type = "p") ~ fitted(.) | ident, abline = 0,col='red')
```

```
> plot(modelo, y ~ fitted(.) | ident, abline = c(0,1))
```

Estas intrucciones generan los siguientes gráficos que sirven para ver el ajuste del modelo vía residuales

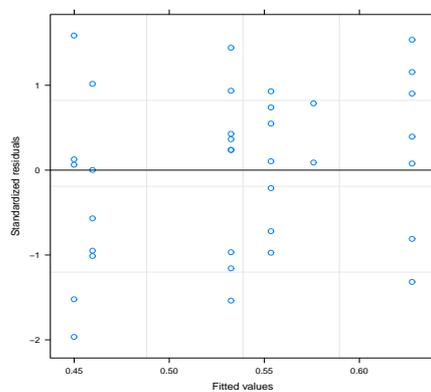


Figura 4.1. Gráfico de residuales estandarizados datos toros.

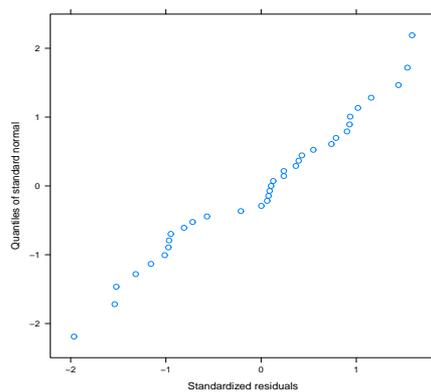


Figura 4.2. Gráfico de normalidad datos toros.

Un box-plot de estos datos se muestra a continuación:

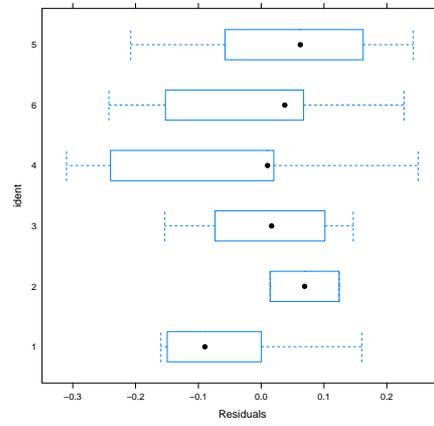


Figura 4.3. Box-plot residuales datos toros.

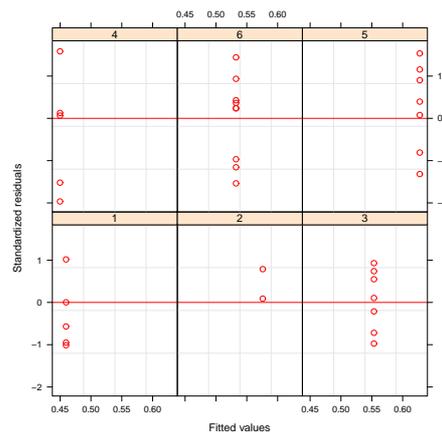


Figura 4.4. Perfiles individuales datos toros.

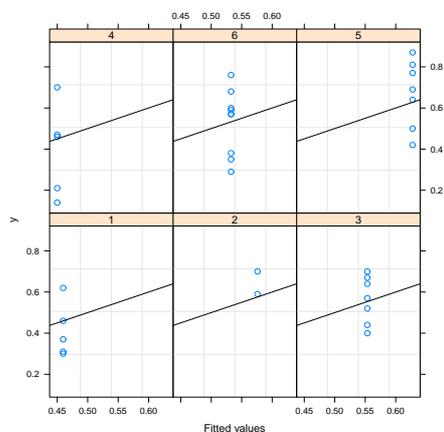


Figura 4.5. Perfiles individuales con recta ajustada datos toros.

Por ejemplo, el gráfico 4.3. basado en boxplots muestra la dispersión alrededor de cada toro. Esta dispersión se asume que es igual para cada toro, pero este gráfico muestra que este supuesto podría no ser razonable en este caso.

```
> coef(modelo)
(Intercept)
1  0.4596276
2  0.5758776
3  0.5534714
4  0.4499162
6  0.5324754
5  0.6277068

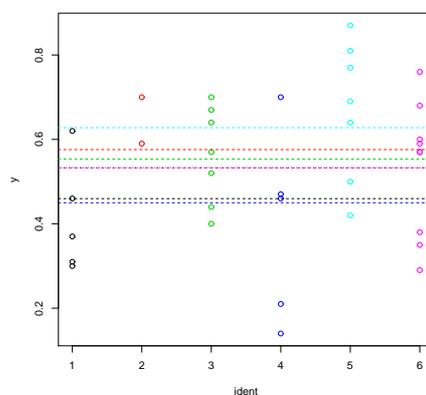
> resid(modelo,type='response') # Residuales crudos
      1      1      1      1      1      2      2
0.00037240 -0.14962760 -0.08962760  0.16037240 -0.15962760  0.12412240  0.01412240
      3      3      3      3      3      3      3
-0.03347140 -0.11347140  0.01652860 -0.15347140  0.11652860  0.08652860  0.14652860
      4      4      4      4      4      5      5
0.02008385 -0.23991615  0.25008385  0.01008385 -0.30991615 -0.20770681  0.01229319
      5      5      5      5      5      6      6
-0.12770681  0.06229319  0.14229319  0.18229319  0.24229319 -0.18247540  0.14752460
      6      6      6      6      6      6      6
0.05752460 -0.15247540  0.03752460  0.22752460  0.03752460 -0.24247540  0.06752460
attr(,"label")
[1] "Residuals"
> resid(modelo) # Residuales Pearson
      1      1      1      1      1      2      2
0.00037240 -0.14962760 -0.08962760  0.16037240 -0.15962760  0.12412240  0.01412240
      3      3      3      3      3      3      3
```

```

-0.03347140 -0.11347140 0.01652860 -0.15347140 0.11652860 0.08652860 0.14652860
      4      4      4      4      4      5      5
0.02008385 -0.23991615 0.25008385 0.01008385 -0.30991615 -0.20770681 0.01229319
      5      5      5      5      5      6      6
-0.12770681 0.06229319 0.14229319 0.18229319 0.24229319 -0.18247540 0.14752460
      6      6      6      6      6      6      6
0.05752460 -0.15247540 0.03752460 0.22752460 0.03752460 -0.24247540 0.06752460
attr(,"label")
[1] "Residuals"
> resid(modelo,type='normalized') # Residuales normalizados
      1      1      1      1      1      2      2
0.002361392 -0.948789941 -0.568329408 1.016922813 -1.012200029 0.787061250 0.089550273
      3      3      3      3      3      3      3
-0.212242433 -0.719523144 0.104808011 -0.973163499 0.738908899 0.548678632 0.929139165
      4      4      4      4      4      5      5
0.127351850 -1.521310459 1.585783893 0.063941762 -1.965181080 -1.317070732 0.077951222
      5      5      5      5      5      6      6
-0.809790021 0.395001666 0.902282376 1.155922732 1.536383265 -1.157078113 0.935454818
      6      6      6      6      6      6      6
0.364764018 -0.966847847 0.237943841 1.442735528 0.237943841 -1.537538646 0.428174107
attr(,"label")
[1] "Standardized residuals"
> fitted(modelo)
      1      1      1      1      1      2      2      3      3
0.4596276 0.4596276 0.4596276 0.4596276 0.4596276 0.5758776 0.5758776 0.5534714 0.5534714
      3      3      3      3      3      4      4      4      4
0.5534714 0.5534714 0.5534714 0.5534714 0.5534714 0.4499162 0.4499162 0.4499162 0.4499162
      4      5      5      5      5      5      5      5      6
0.4499162 0.6277068 0.6277068 0.6277068 0.6277068 0.6277068 0.6277068 0.6277068 0.5324754
      6      6      6      6      6      6      6      6      6
0.5324754 0.5324754 0.5324754 0.5324754 0.5324754 0.5324754 0.5324754 0.5324754
attr(,"label")
[1] "Fitted values"
> fixed.effects(modelo)
(Intercept)
0.5331792
> random.effects(modelo)
(Intercept)
1 -0.0735515596
2 0.0426984390
3 0.0202922385
4 -0.0832630064
6 -0.0007037626
5 0.0945276511
> plot(ident,y,col=ident)
> abline(h=unique( fitted(modelo)),col=1:6,lty=2)
> abline(h=unique( fixed.effects(modelo)),lty=3)

```

Estas líneas producen el gráfico con el ajuste del modelo para cada toro.



**Figura 4.6.** Gráfico modelos individuales y modelo de efectos fijos ajustados datos toros.

Este tipo de gráficos demuestran la gran habilidad de los modelos lineales mixtos para describir comportamientos individuales. Por ejemplo, en este caso, se tiene en cuenta la variabilidad intra-toro en el ajuste.

### 4.3. Hueso ramus

A una cohorte de niños se les registró la altura (en m.m.) del hueso ramus (ubicado en la mandíbula<sup>3</sup>). Estas mediciones se realizaron a los 8, 8 y medio, 9 y 9 y medio años (20 datos). El objetivo era establecer una curva de crecimiento normal para ser usada por ortodoncistas. Los datos aparecen en el apéndice. Se ajustará un MLM a estos datos usando R.

```
HuesoRamus<-scan()
478 488 490 497
464 473 477 484
463 468 478 485
451 453 461 472
476 485 489 493
525 532 533 537
512 530 543 545
498 500 503 527
481 508 523 544
450 470 473 483
```

<sup>3</sup>Alves, N. et al. (2011) Peripheral Osteoma in the Ramus of Mandible: Report of Case *International journal of odontostomatology*, 5(3):215-219

```

512 514 516 519
485 492 530 555
521 528 537 550
482 489 493 498
496 504 512 518
507 517 527 533
472 477 484 495
533 546 551 553
462 475 481 484
463 476 513 518

```

```

HuesoRamus<-HuesoRamus/10

edad<-rep(c(8,8.5,9,9.5),20)
individuo<-rep(1:20,each=4)

library(nlme)
Hueso<-data.frame(HuesoRamus=HuesoRamus,edad=edad,individuo=individuo)
Hueso.n <- groupedData( HuesoRamus ~ edad | individuo,
                        data = as.data.frame( Hueso ) )

```

**Salida del programa explicada** Definición del modelo usando la función del R llamada *lme()*. En este modelo se especifica la edad como variable independiente o explicativa.

```

modelo<-lme(HuesoRamus ~ edad, data=Hueso.n,random=~1+edad,
            corAR1(0.5),method='ML') #modelo de intercepto y pendiente aleatorios
summary(modelo)
Linear mixed-effects model fit by maximum likelihood
Data: Hueso.n
      AIC      BIC    logLik
248.0334 264.7076 -117.0167

Random effects:
Formula: ~1 + edad | individuo
Structure: General positive-definite, Log-Cholesky parametrization
      StdDev      Corr
(Intercept) 5.796287e-07 (Intr) <-sigma.gorro(b0)
edad        1.610902e-01 0.114 <-sigma.gorro(b1) y correlación(b0,b1)
Residual    2.158711e+00 <- sigma.gorro del modelo

```

Los resultados anteriores son:

- Desviación típica del intercepto aleatorio  $b_0$ :  $\hat{\sigma}_{b_0} = 5,796287e - 07 = 0,0000005796287$  (parece no ser necesario especificarlo en el modelo por estar tan cercano a cero)
- Desviación típica de la pendiente aleatoria  $b_1$ :  $\hat{\sigma}_{b_1} = 1,610902e - 01 = 0,1610902$  (sí parece importante dentro del modelo)
- Correlación estimada entre  $b_0$  y  $b_1$ :  $\hat{\rho}_{b_0, b_1} = 0,114$
- Desviación típica estimada del modelo:  $\hat{\sigma}_e = 2,158711e + 00$ .

```

Correlation Structure: AR(1)
Formula: ~1 | individuo
Parameter estimate(s):
  Phi
0.9348663 <- Coeficiente de correlación del cluster

Fixed effects: HuesoRamus ~ edad
              Value Std.Error DF  t-value p-value
(Intercept) 33.75093 1.7867548 59 18.88951    0
edad         1.86335 0.2004835 59  9.29429    0
Correlation:
  (Intr)
edad -0.949

```

De acuerdo a estos resultados, la edad es altamente importante en este modelo (valor- $p < 0.0001$ ). Los resultados referentes al modelo poblacional son presentados. Estos son

- Estimación del intercepto  $\hat{\beta}_0$ :  $\hat{\beta}_0 = 33,75093$
- Desviación típica del intercepto estimado  $\hat{\beta}_0$ :  $\hat{\sigma}_{\hat{\beta}_0} = 1,7867548$
- Estimación de la pendiente  $\hat{\beta}_1$ :  $\hat{\beta}_1 = 1,86335$ . Esto implica que a medida que se incrementa la edad hay un aumento en la medición relacionada con el hueso ramus.
- Desviación típica de la pendiente estimada  $\hat{\beta}_1$ :  $\hat{\sigma}_{\hat{\beta}_1} = 0,2004835$
- $\hat{\rho}_{\hat{\beta}_0, \hat{\beta}_1} = -0,949$

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-1.4190638	-0.6826500	-0.1152897	0.7184711	1.7676155

Number of Observations: 80

Number of Groups: 20

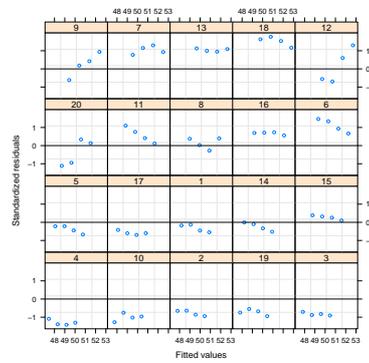
>

# Gráficos

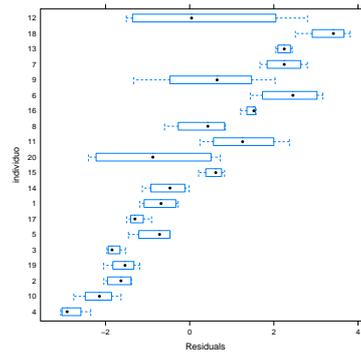
```
plot(modelo, resid(., type = "p") ~ fitted(.) | individuo, abline = 0)
```

```
plot(modelo, individuo ~ resid(.))
```

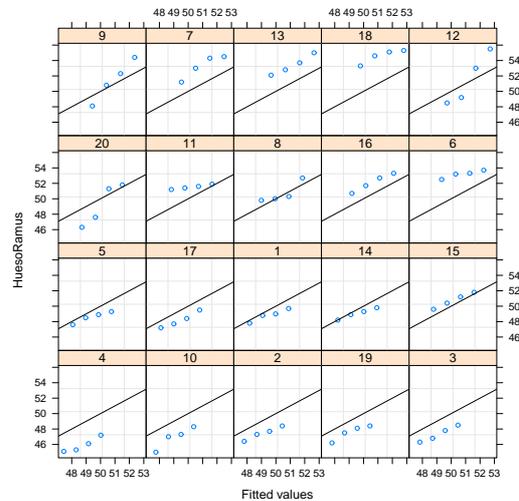
```
plot(modelo, HuesoRamus ~ fitted(.) | individuo, abline = c(0,1))
```



**Figura 4.7.** Gráfico de residuales del modelo datos hueso ramus. No se observa un patrón de crecimiento homogéneo sino más bien una tendencia individual muy propia de cada sujeto.



**Figura 4.8.** Individuo versus residuales. Se observa la alta heterogeneidad entre algunos sujetos.



**Figura 4.9.** Modelos individuales ajustados. Estos modelos ajustados tienen en cuenta la variabilidad intra-individual.

En la figura 4.8 se puede apreciar una falta de ajuste para algunos sujetos, por ejemplo el 11 y el 14 ya que el modelo predice una tendencia lineal negativa. Por esta razón es importante usar este tipo de gráfico para evaluar de alguna manera el desempeño del modelo, sobre todo a nivel individual.

¿Qué pasa si se especifica un modelo de intercepto aleatorio? (para, de esta forma, tener en cuenta la variabilidad intra-individual) ¿Será que mejora el ajuste? En R, este modelo se puede especificar así,

```
# Otro modelo: intercepto aleatorio (random=~1).
> modelo1<-lme(HuesoRamus ~ edad, data=Hueso.n,random=~1,
               corAR1(0.5),method='ML')
> summary(modelo1)
Linear mixed-effects model fit by maximum likelihood
Data: Hueso.n
      AIC      BIC    logLik
244.4313 256.3414 -117.2156

Random effects:
Formula: ~1 | individuo
      (Intercept) Residual
StdDev: 0.0005995177 2.556453

Correlation Structure: AR(1)
Formula: ~1 | individuo
```

```

Parameter estimate(s):
  Phi
0.9525837
Fixed effects: HuesoRamus ~ edad
              Value Std.Error DF   t-value p-value
(Intercept) 33.75029 1.8454047 59 18.288826    0
edad         1.86334 0.2009975 59  9.270478    0
Correlation:
  (Intr)
edad -0.953
    
```

Este modelo detecta el efecto positivo de la edad (a mayor edad hay más incremento en la medición asociada al hueso ramus)

```

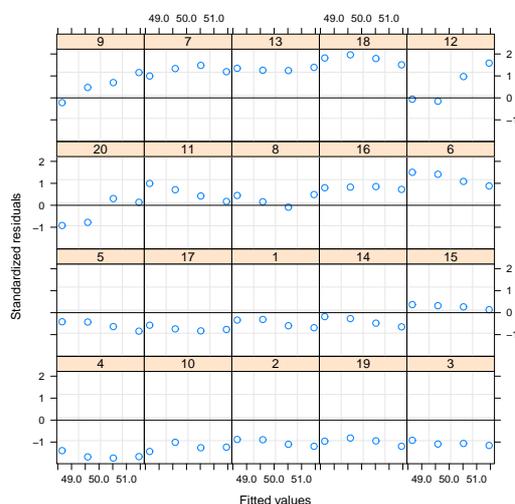
Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.7291042 -0.8520756 -0.1191249  0.8592802  1.9602548
    
```

```

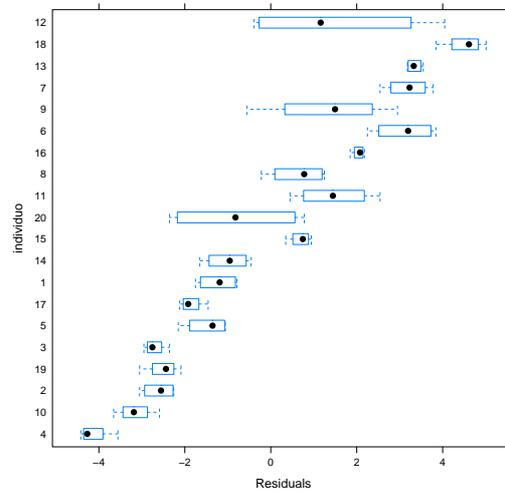
Number of Observations: 80
Number of Groups: 20
    
```

```

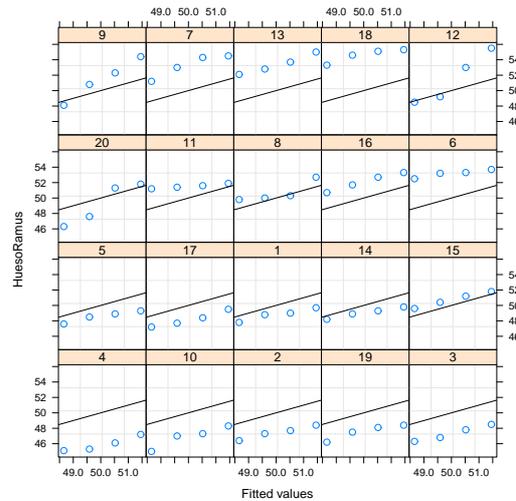
plot(modelo1, resid(., type = "p") ~ fitted(.) | individuo, abline = 0)
plot(modelo1, individuo ~ resid(.))
plot(modelo1, HuesoRamus ~ fitted(.) | individuo, abline = c(0,1))
    
```



**Figura 4.10.** Gráfico de residuales del Modelo1 datos hueso ramus. No se observa un patrón de crecimiento homogéneo sino más bien una tendencia individual muy propia de cada sujeto.



**Figura 4.11.** Individuo versus residuales (Modelo1). Se observa la alta heterogeneidad entre algunos sujetos.



**Figura 4.12.** Modelos individuales ajustados (Modelo1). Estos modelos ajustados tienen en cuenta la variabilidad intra-individual.

De acuerdo al gráfico 4.10, para los sujetos de referencia 11 y 14, se observa que este modelo de intercepto aleatorio captura mejor la tendencia de crecimiento de la medida asociada al hueso ramus. Ahora, se especifica un modelo cuadrático con intercepto aleatorio con el fin de ver su efecto en el ajuste.

```

#Modelo 2: Modelo cuadrático con intercepto aleatorio
> modelo2<-lme(HuesoRamus ~ edad+I(edad^2), data=Hueso.n,
               random=~1,corAR1(0.5),method='ML')
> summary(modelo2)
Linear mixed-effects model fit by maximum likelihood
Data: Hueso.n
      AIC      BIC    logLik
246.3033 260.5955 -117.1517

Random effects:
Formula: ~1 | individuo
      (Intercept) Residual
StdDev: 0.0006759438 2.556428

Correlation Structure: AR(1)
Formula: ~1 | individuo
Parameter estimate(s):
      Phi
0.952686
Fixed effects: HuesoRamus ~ edad + I(edad^2)
              Value Std.Error DF   t-value p-value
(Intercept) 26.908248 19.574269 58   1.3746745 0.1745
edad         3.438343  4.490151 58   0.7657521 0.4469
I(edad^2)   -0.090000  0.256320 58  -0.3511235 0.7268
Correlation:
      (Intr) edad
edad   -0.999
I(edad^2) 0.995 -0.999

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.7459268 -0.8551972 -0.1359319  0.8468834  1.9434681

Number of Observations: 80
Number of Groups: 20

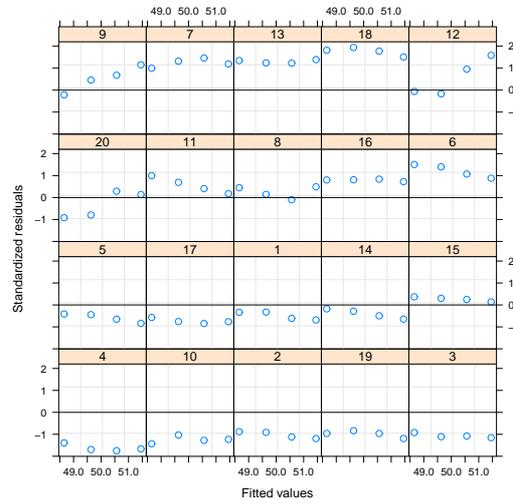
La estimación del efecto de la edad es positiva (3.438343) y del término cuadrático
es negativa, indicando una tendencia a decrecer con la edad aunque muy leve.

plot(modelo2, resid(., type = "p") ~ fitted(.) | individuo, abline = 0)
plot(modelo2, individuo ~ resid(.))
plot(modelo2, HuesoRamus ~ fitted(.) | individuo, abline = c(0,1))

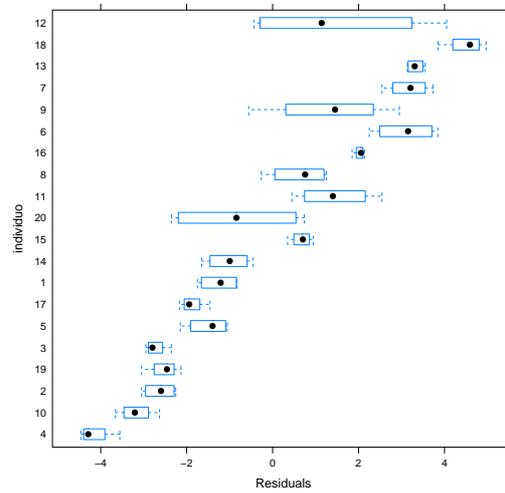
> anova(modelo2)
      numDF denDF  F-value p-value

```

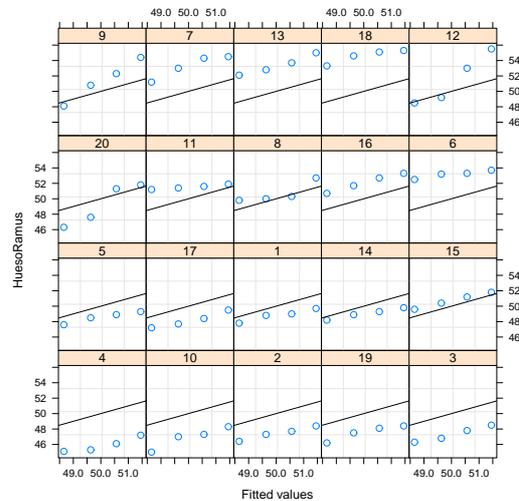
(Intercept)	1	58	7916.351	<.0001
edad	1	58	85.016	<.0001
I(edad^2)	1	58	0.123	0.7268



**Figura 4.13.** Gráfico de residuales del Modelo2 datos hueso ramus. No se observa un patrón de crecimiento homogéneo sino más bien una tendencia individual muy propia de cada sujeto.



**Figura 4.14.** Individuo versus residuales (Modelo2). Se observa la alta heterogeneidad entre algunos sujetos.



**Figura 4.15.** Modelos individuales ajustados (Modelo2). Estos modelos ajustados tienen en cuenta la variabilidad intra-individual.

Los comportamientos capturados por los residuales son muy similares a los del Modelo 1. Finalmente, se especificará un modelo cuadrático con intercepto y pendiente aleatorios.

```
#Modelo cuadrático de intercepto y pendiente aleatorios
modelo3<-lme(HuesoRamus ~ edad+I(edad^2), data=Hueso.n,
             random=~1+edad,corAR1(0.5),method='ML')
```

```
summary(modelo3)
Linear mixed-effects model fit by maximum likelihood
Data: Hueso.n
      AIC   BIC  logLik
249.9038 268.96 -116.9519
```

```
Random effects:
Formula: ~1 + edad | individuo
Structure: General positive-definite, Log-Cholesky parametrization
      StdDev      Corr
(Intercept) 8.436195e-07 (Intr)
edad        1.611203e-01 0.11
Residual    2.158702e+00
```

```
Correlation Structure: AR(1)
Formula: ~1 | individuo
Parameter estimate(s):
```

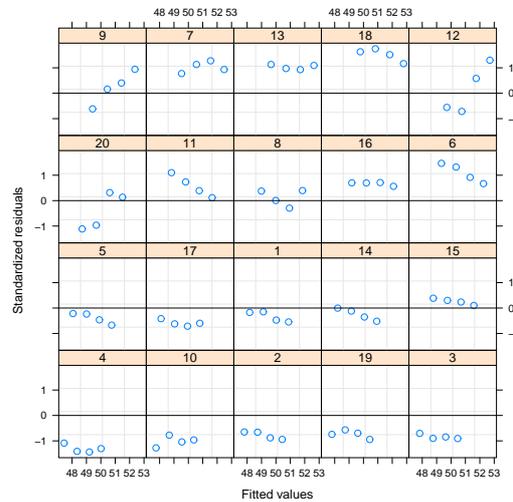
```

Phi
0.9350134
Fixed effects: HuesoRamus ~ edad + I(edad^2)
              Value Std.Error DF   t-value p-value
(Intercept) 26.908174 19.443214 58  1.3839365 0.1717
edad         3.438352  4.460671 58  0.7708149 0.4439
I(edad^2)   -0.090000  0.254635 58 -0.3534470 0.7250
Correlation:
      (Intr) edad
edad   -0.999
I(edad^2) 0.996 -0.999
Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.4385111 -0.6999002 -0.1347011  0.7024288  1.7481394
Number of Observations: 80
Number of Groups: 20
    
```

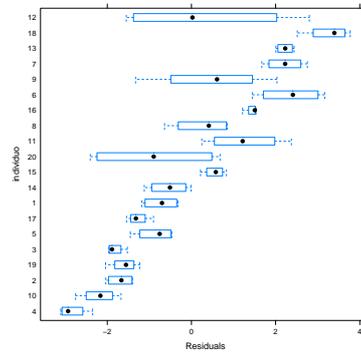
Al igual que todos los modelos anteriores, la estimación del efecto de la edad es positiva (3.438352, valor muy cercano al obtenido con el modelo anterior) y del término cuadrático es negativa, indicando una tendencia a decrecer con la edad aunque muy leve.

```

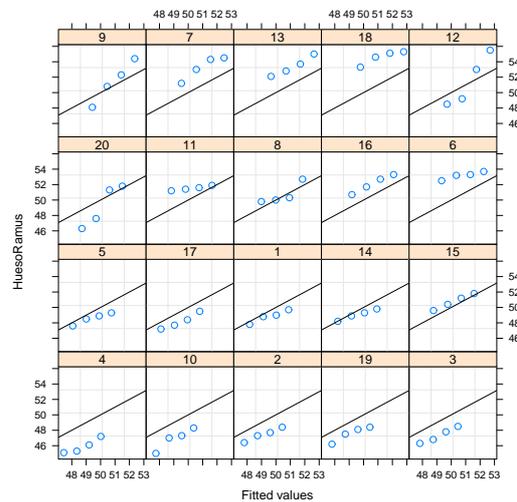
plot(modelo3, resid(., type = "p") ~ fitted(.) | individuo, abline = 0)
plot(modelo3, individuo ~ resid(.))
plot(modelo3, HuesoRamus ~ fitted(.) | individuo, abline = c(0,1))
    
```



**Figura 4.16.** Gráfico de residuales del Modelo3 datos hueso ramus. No se observa un patrón de crecimiento homogéneo sino más bien una tendencia individual muy propia de cada sujeto.



**Figura 4.17.** Individuo versus residuales (Modelo2). Se observa la alta heterogeneidad entre algunos sujetos.



**Figura 4.18.** Modelos individuales ajustados (Modelo2). Estos modelos ajustados tienen en cuenta la variabilidad intra-individual.

Los resultados de este modelo son muy similares a los obtenidos con los modelos 1 y 2 lo que indica que la complejidad extra incluida en el modelo a través de la inclusión de un término cuadrático aleatorio no es necesaria.

```
anova(modelo2,modelo3)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
modelo2    1  6 246.3033 260.5955 -117.1517
modelo3    2  8 249.9038 268.9600 -116.9519 1 vs 2 0.3995455 0.8189
```

```

anova(modelo3)
      numDF denDF F-value p-value
(Intercept)      1    58 7667.581 <.0001
edad             1    58  85.450 <.0001
I(edad^2)        1    58   0.125  0.725

modelo3$varFix
      (Intercept)      edad  I(edad^2)
(Intercept) 363.862142 -83.366227  4.74487886
edad        -83.366227  19.151425 -1.09213235
I(edad^2)    4.744879  -1.092132  0.06240756
modelo3$sigma
[1] 2.158702
modelo3$fixDF
$X
      (Intercept)      edad  I(edad^2)
      58            58            58

$terms
      (Intercept)      edad  I(edad^2)
      58            58            58

attr("assign")
attr("assign")$'(Intercept)'  

[1] 1

attr("assign")$edad
[1] 2

attr("assign")$'I(edad^2)'  

[1] 3

attr("varFixFact")
      [,1]      [,2]      [,3]
[1,] -0.5547538  0.0000000  0.0000000
[2,]  1.6729700 -0.1977600  0.0000000
[3,] 18.9935663 -4.3717635  0.2498151

coef(modelo3)
      (Intercept)      edad  I(edad^2)
4      26.90817  3.287460   -0.09
10     26.90817  3.325602   -0.09
2      26.90817  3.331724   -0.09
19     26.90817  3.332075   -0.09
3      26.90817  3.334349   -0.09

```

5	26.90817	3.364952	-0.09
17	26.90817	3.369320	-0.09
1	26.90817	3.378473	-0.09
14	26.90817	3.382728	-0.09
15	26.90817	3.452521	-0.09
20	26.90817	3.444579	-0.09
11	26.90817	3.459691	-0.09
8	26.90817	3.480162	-0.09
16	26.90817	3.505211	-0.09
6	26.90817	3.523031	-0.09
9	26.90817	3.534954	-0.09
7	26.90817	3.547500	-0.09
13	26.90817	3.563985	-0.09
18	26.90817	3.578896	-0.09
12	26.90817	3.569818	-0.09

Después de comparar los AIC para los 4 modelos (248.0334 para el modelo lineal con intercepto y pendiente aleatorios ('Modelo'), 244.4313 para el modelo lineal de solo intercepto aleatorio ('Modelo1'), 246.3033 para el modelo cuadrático con intercepto aleatorio ('Modelo2') y 249.9038 para el modelo cuadrático con intercepto y pendiente aleatorios ('Modelo3')) se selecciona el que tiene asociado un menor valor de AIC, que en este caso particular es el Modelo1 lineal de intercepto y pendiente aleatorios.

#### 4.4. Puntajes

Estos datos son simulados. Se registró el puntaje obtenido en una prueba a un grupo de 18 personas pertenecientes a 4 familias diferentes (la dependencia la induce la pertenencia a una familia específica). Se busca comparar estos puntajes en términos del género. Se usará R para ajustar dos modelos lineales mixtos a estos datos: uno con intercepto aleatorio y otro con intercepto y pendiente aleatorios.

```
score<-c(67, 66, 64, 71, 72, 63, 63, 67, 69, 68, 70, 63, 64,
67, 66, 67, 67, 69)
flia<-rep(1:4,c(5,6,2,5))
sexo<-rep(c('F','M','F','M','F','M','F','M'),
c(3,2,3,3,1,1,2,3))
```

El primer modelo que se especificará será un modelo que relaciona el score con el sexo y tiene un intercepto aleatorio.

```

library(nlme)
ejem<-data.frame(score=score,flia=flia,sexo=sexo)
ejem.n <- groupedData( score ~ sexo| flia, data = as.data.frame( ejem ) )
modelo<-lme(score ~ sexo, data=ejem.n,random=~1,method='REML')
summary(modelo)
Linear mixed-effects model fit by REML
Data: ejem.n
      AIC      BIC   logLik
80.2952 83.38556 -36.1476

Random effects:
Formula: ~1 | flia
      (Intercept) Residual
StdDev:    1.592043 1.759580

Fixed effects: score ~ sexo
              Value Std.Error DF   t-value p-value
(Intercept) 64.72450 1.0018533 13 64.60476 0e+00
sexoM        3.57981 0.8369816 13  4.27705 9e-04
Correlation:
      (Intr)
sexoM -0.418

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.3826629 -0.8908950  0.1646147  0.6613020  1.2583009

Number of Observations: 18
Number of Groups: 4

```

De las estimaciones de los parámetros de este modelo se observa que el efecto del sexo es estadísticamente significativo (valor  $p < 0.001$ ) y el ser hombre (sexo=M) incrementa el score (efecto para sexo estimado 3.57981). El hecho de que la desviación estándar estimada para el intercepto aleatorio esté lejos del cero (1.592043) indica la necesidad de especificar la variabilidad intra-sujeto en el modelo. Ahora se especificará un modelo con intercepto y pendiente aleatorios.

```

modelo<-lme(score ~ sexo, data=ejem.n,random=~1+sexo,method='REML')
summary(modelo)

```

Linear mixed-effects model fit by REML

Data: ejem.n

	AIC	BIC	logLik
	81.25271	85.88825	-34.62636

Random effects:

Formula: ~1 + sexo | flia

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	0.9451345	(Intr)
sexoM	2.1455126	0.617
Residual	1.4559151	

Fixed effects: score ~ sexo

	Value	Std.Error	DF	t-value	p-value
(Intercept)	64.9133	0.6853736	13	94.71227	0.0000
sexoM	3.2080	1.2879873	13	2.49071	0.0271

Correlation:

(Intr)

sexoM 0.084

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-1.27647312	-0.61725621	0.03662495	0.71652710	1.47093988

Number of Observations: 18

Number of Groups: 4

De las estimaciones de los parámetros de este modelo se observa que el efecto del sexo es estadísticamente significativo (valor- $p=0.0271$ ) y el ser hombre incrementa el score (efecto para sexo estimado 3.2080). El hecho de que la desviación estándar estimada para el intercepto aleatorio y la pendiente estén lejos del cero (0.9451345 y 2.1455126, respectivamente) indican la necesidad de especificar la variabilidad intra-sujeto en el modelo a través de estas componentes aleatorias.

# Capítulo 5

## Diagnósticos en el modelo mixto

Los modelos lineales mixtos son muy sensibles a respuestas extrañas y a puntos extremos en los espacios de diseño de los efectos fijos y efectos aleatorios (Zewotir y Galpin, 2005). Se sabe que no todas las observaciones en los datos juegan un papel igual en la determinación de las estimadas, pruebas y otros estadísticos en el modelo. Algunas veces las estimaciones del modelo pueden ser determinadas por unas pocas observaciones mientras que la mayoría de las observaciones son básicamente ignoradas. Es importante que el analista esté consciente de las observaciones que tienen una gran influencia inusual en los resultados. Tales casos detectados pueden considerarse adecuados y retenidos en el análisis, o pueden representar datos inapropiados y deben por tanto ser removidos del análisis, o pueden sugerir la necesidad de recoger datos adicionales, o podrían sugerir que el esquema de modelación es inapropiado, o también podrían indicar errores en el registro o entrada de datos.

### 5.1. Diagnósticos vía eliminación de observaciones

Zewotir y Galpin (2005) presentan medidas de diagnóstico eliminando una observación a la vez y mirando los cambios en diferentes partes del modelo. Recuerde la formulación del modelo lineal mixto,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$$

donde

- $\mathbf{X}$  es una matriz de diseño de los efectos fijos de dimensión  $n \times p$ .
- $\mathbf{Z} = [\mathbf{Z}_1 | \mathbf{Z}_2 | \cdots | \mathbf{Z}_r]$  donde  $\mathbf{Z}_i$  es una matriz de dimensión  $n \times q_i$ .
- $\mathbf{b}_i$  es un vector aleatorio de dimensión  $q_i \times 1$ ,  $\mathbf{u}_i \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I}_i)$ , con  $i = 1, \dots, r$ .
- $\boldsymbol{\epsilon}$  es un vector aleatorio de dimensión  $n \times 1$ .  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I})$ .
- $\mathbf{b}_i$  y  $\boldsymbol{\epsilon}$  son independientes.
- $\mathbf{b} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{D})$ , donde  $\mathbf{D}$  es una matriz diagonal de bloques  $\gamma_i \mathbf{I}$  con  $\gamma_i = \sigma_i^2 / \sigma_e^2$ .
- $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_e^2 \mathbf{H})$  con  $\mathbf{H} = \mathbf{I} + \sum_{i=1}^r \gamma_i \mathbf{Z}_i' \mathbf{Z}_i$ .

Si los  $\gamma_i$ 's son conocidos (y por lo tanto lo es  $\mathbf{H}$ ), los m.l.e. de  $\boldsymbol{\beta}$ ,  $\sigma_e^2$  y el valor realizado de  $\mathbf{b}$  están dados por,

- $\hat{\boldsymbol{\beta}} = W_H(\mathbf{X}, \mathbf{X})^{-1} W_H(\mathbf{X}, \mathbf{Y})$  donde  $W_H(\mathbf{A}, \mathbf{b}) = \mathbf{A}' \mathbf{H}^{-1} \mathbf{b}$  se conoce como el operador de Hemmerle y Hartley (Hammerle y Hartley, 1973). Este estimador  $\hat{\boldsymbol{\beta}}$  es un Best Linear Unbiased Estimator, BLUE.
- $\hat{\sigma}_e^2 = \frac{1}{n} W_H(\mathbf{d}, \mathbf{d})$ , donde  $\mathbf{d} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ .
- $\tilde{\mathbf{b}} = \mathbf{DZ}' \mathbf{H}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{D} W_H(\mathbf{Z}, \mathbf{d})$ .

Si los  $\gamma_i$ 's son desconocidos los m.l.e. son sustituidos en  $\mathbf{D}$  (y/o  $\mathbf{H}$ ) para obtener  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\sigma}_e^2$  y  $\tilde{\mathbf{b}}$ .

**Algo de notación:**

- Valores ajustados de  $\mathbf{Y}$ : Son los  $\hat{\mathbf{Y}}$ . Estos se obtienen como  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \tilde{\mathbf{Z}}\tilde{\mathbf{b}}$ .
- Residuales:  $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ .
- $\mathbf{R} = \mathbf{H}^{-1} - \mathbf{H}^{-1} \mathbf{X} W_H(\mathbf{X}, \mathbf{X})^{-1} \mathbf{X}' \mathbf{H}^{-1}$ . Esta es una matriz simétrica.
- $\mathbf{C} = \mathbf{H}^{-1}$ .
- $\mathbf{c}_i$ :  $i$ -ésima columna de  $\mathbf{H}^{-1}$ .
- $c_{ii}$ :  $i$ -ésimo elemento de la diagonal de  $\mathbf{H}^{-1}$ .

- $\mathbf{e} = \mathbf{R}\mathbf{Y}$ .

Si los  $\gamma_i$ 's son conocidos, y por lo tanto  $\mathbf{H}$ , entonces  $\mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{R})$  y la correlación entre  $e_i$  y  $e_j$  está determinada completamente por los elementos de  $\mathbf{R}$ :

$$\text{Cor}(e_i, e_j) = \frac{r_{ij}}{\sqrt{r_{ii}r_{jj}}}$$

Si los  $\gamma_i$ 's son desconocidos, sus estimadas son útiles solo como una aproximación. A partir de los resultados anteriores se pueden definir los *residuales studentizados*

$$t_i = \frac{e_i}{\hat{\sigma}_e \sqrt{r_{ii}}}$$

Se pueden hallar otras cantidades de interés como por ejemplo:

- $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_i = W_H(\mathbf{X}, \mathbf{X})^{-1} \mathbf{X}' \mathbf{c}_i \frac{e_i}{r_{ii}}$
- $\hat{\sigma}_{e(i)}^2 = \frac{n}{n-1} \hat{\sigma}_e^2 - \frac{e_i^2}{(n-1)r_{ii}} = \hat{\sigma}_e^2 \left( \frac{n}{n-1} - \frac{t_i^2}{n-1} \right)$
- $\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)} = \mathbf{DZ}' \mathbf{R}_i \frac{e_i}{r_{ii}}$

### 5.1.1. Influencia en los estimadores de los efectos fijos

Análogo de la distancia de Cook

$$\begin{aligned} CD_i(\boldsymbol{\beta}) &= \frac{(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})' W_H(\mathbf{X}, \mathbf{X}) (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})}{p \hat{\sigma}_e^2} \\ &= \frac{\mathbf{c}_i' \mathbf{X} W_H(\mathbf{X}, \mathbf{X})^{-1} \mathbf{c}_i e_i^2}{p r_{ii} \hat{\sigma}_e^2} \\ &= \frac{(c_{ii} - r_{ii}) t_i^2}{p r_{ii}} \end{aligned}$$

Valores grandes de  $CD_i(\boldsymbol{\beta})$  señalan puntos para ser considerados posteriormente.

### Análogo a la razón de determinantes de matrices de varianzas

La razón de los determinantes de las matrices de varianzas mide el cambio del determinante de la matriz de varianzas y covarianzas de  $\hat{\boldsymbol{\beta}}$  cuando la  $i$ -ésima observación es eliminada.

$$\begin{aligned} VR_i(\boldsymbol{\beta}) &= \frac{|VarCov(\hat{\boldsymbol{\beta}}_{(i)})|}{|VarCov(\hat{\boldsymbol{\beta}})|} \\ &= \left(\frac{n-t_i^2}{n-1}\right) \frac{c_{ii}}{r_{ii}} \end{aligned}$$

Si este número es cercano a 1 indica que el punto no es influyente. Otro indicador que puede ser usado como medida es  $|VR_i(\boldsymbol{\beta}) - 1|$ .

Se puede también utilizar la traza en lugar del determinante en la medida anterior:

$$\begin{aligned} VR_i(\boldsymbol{\beta}) &= \text{Traza} \left\{ [VarCov(\hat{\boldsymbol{\beta}})]^{-1} VarCov(\hat{\boldsymbol{\beta}}_{(i)}) \right\} \\ &= \frac{n-t_i^2}{n-1} \left( \frac{c_{ii}}{r_{ii}} + p - 1 \right) \end{aligned}$$

Ya que  $VR_i(\boldsymbol{\beta})$  estará cercano a  $p$  para observaciones no influyentes, entonces puede usarse  $|VR_i(\boldsymbol{\beta}) - p|$ .

### Análogo al estadístico de Cook-Weisberg

Bajo el supuesto  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_e^2 \mathbf{H})$  tenemos que  $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma_e^2 W_H(\mathbf{X}, \mathbf{X})^{-1})$ .

El elipsoide del  $100(1-\alpha)100\%$  de confianza para  $\boldsymbol{\beta}$  está dado por

$$E = \left\{ \boldsymbol{\beta} \left| \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' W_H(\mathbf{X}, \mathbf{X}) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{p \sigma_e^2} \leq F_{(\alpha; p, n-p)} \right. \right\}$$

Cuando la  $i$ -ésima observación es omitida se tiene

$$E_{(i)} = \left\{ \boldsymbol{\beta} \left| \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{(i)})' W_{(i)}(\mathbf{X}, \mathbf{X}) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{(i)})}{p \sigma_{e(i)}^2} \leq F_{(\alpha; p, n-p-1)} \right. \right\}$$

Cook y Weisberg (1982) propusieron el log de la razón de  $E_{(i)}$  contra  $E$  como una medida de influencia

$$\begin{aligned}
CW_i &= \log \left( \frac{Vol(E_{(i)})}{Vol(E)} \right) \\
&= \frac{1}{2} \log \left( \frac{c_{ii}}{r_{ii}} \right) + \frac{p}{2} \log \left( \frac{n-t_i}{n-1} \right) + \frac{p}{2} \log \left( \frac{F_{(\alpha;p,n-p)}}{F_{(\alpha;p,n-p-1)}} \right)
\end{aligned}$$

Aquí  $Vol(E)$  se refiere al volumen del hiper-elipsoide  $E$ . Si el  $CW_i$  es negativo (positivo) entonces el volumen del elipsoide de confianza es decreciente (creciente) en precisión después de eliminar la  $i$ -ésima observación. Sin considerar el signo de los  $CW_i$ 's, un valor grande de  $|CW_i|$  indica una fuerte influencia sobre  $\hat{\beta}$  y/o  $VarCov(\hat{\beta})$ .

Hay una relación entre el  $CW_i$  y el  $VR_i$ :

$$CW_i = \frac{1}{2} \log(VR_i(\beta)) + \frac{p}{2} \log \left( \frac{F_{(\alpha;p,n-p)}}{F_{(\alpha;p,n-p-1)}} \right)$$

### Análogo al estadístico de Andrews-Pregibon

Este estadístico (Andrews y Pregibon, 1973), de naturaleza exploratoria, se usa para examinar la posible presencia de uno a más outliers importantes

$$AP_i = \frac{1}{2} \log \left( \frac{\mathbf{e}'_{(i)} \mathbf{e}_{(i)} |W_{(i)}(\mathbf{X}, \mathbf{X})|}{\mathbf{e}' \mathbf{e} |W_H(\mathbf{X}, \mathbf{X})|} \right)$$

### 5.1.2. Influencia en los efectos aleatorios

La medida de diagnóstico propuesta para medir la posible influencia en los efectos aleatorios, examina la distancia al cuadrado del predictor de efectos aleatorios con los datos completos con el predictor cuando se ha eliminado la  $i$ -ésima observación, relativa a la varianza de efectos aleatorios  $VarCov(\mathbf{b}) = \sigma_e^2 \mathbf{D}$ . Esto es análogo a la distancia de Cook y puede escribirse como

$$\begin{aligned}
CD_i(\mathbf{b}) &= \frac{(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)})' \mathbf{D}^{-1} (\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)})}{\hat{\sigma}_e^2} \\
&= \frac{t_i^2}{r_{ii}} \mathbf{R}'_i \mathbf{Z} \mathbf{D}' \mathbf{D} \mathbf{Z}' \mathbf{R}_i \\
&= \frac{t_i^2}{r_{ii}} \mathbf{R}'_i (\mathbf{H} - \mathbf{I}) \mathbf{R}_i
\end{aligned}$$

Un  $CD_i(\mathbf{b})$  grande indica que la  $i$ -ésima observación es influyente en la predicción de los efectos aleatorios.

### 5.1.3. Influencia sobre la función de verosimilitud

Debido a las dificultades computacionales en la estimación de los cocientes de las componentes de varianzas, se considera la estimación de  $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \sigma_e^2)$ , asumiendo  $\gamma$  fijo. Para  $\hat{\boldsymbol{\theta}}(\gamma)$  y  $\hat{\boldsymbol{\theta}}_{(i)}(\gamma)$ , los e.m.v. de  $\boldsymbol{\theta}$  basados en  $n$  y  $n - 1$  observaciones, respectivamente. Sean  $\ell(\hat{\boldsymbol{\theta}}(\gamma))$  y  $\ell(\hat{\boldsymbol{\theta}}_{(i)}(\gamma))$  las funciones de log-verosimilitud evaluadas en  $\hat{\boldsymbol{\theta}}(\gamma)$  y  $\hat{\boldsymbol{\theta}}_{(i)}(\gamma)$ .

La distancia entre las funciones de log-verosimilitud mide la influencia de la  $i$ -ésima observación en la función de verosimilitud, definida la distancia como,

$$LD_i = 2 \left\{ \ell(\hat{\boldsymbol{\theta}}(\gamma)) - \ell(\hat{\boldsymbol{\theta}}_{(i)}(\gamma)) \right\}$$

donde

$$\ell(\hat{\boldsymbol{\theta}}(\gamma)) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\hat{\sigma}_e^2) - \frac{1}{2} \log(|\mathbf{H}|) - \frac{n}{2}$$

y

$$\begin{aligned}
\ell(\hat{\boldsymbol{\theta}}_{(i)}(\gamma)) &= -\frac{n-1}{2} \log(2\pi) - \frac{n-1}{2} \log(\hat{\sigma}_{e(i)}^2) - \frac{1}{2} \log(|\mathbf{H}|) - \\
&\quad \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)})' \mathbf{H}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)})}{2\hat{\sigma}_{e(i)}^2}
\end{aligned}$$

y ya que

$$\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)}\right)' \mathbf{H}^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)}\right) = n\hat{\sigma}_e^2 + \frac{\sigma_i^2}{r_{ii}^2} (c_{ii} - r_{ii})$$

se obtiene

$$LD_i = n \log \left( \frac{n - t_i^2}{n - 1} \right) + \frac{[n + t_i^2 (c_{ii} - r_{ii})] (n - 1)}{r_{ii} (n - t_i^2)} - n$$

Si  $LD_i$  es grande entonces la  $i$ -ésima observación es influyente en la función de verosimilitud.

## 5.2. Verificación de la normalidad de los efectos aleatorios

Lange y Ryan (1989) presentan una metodología para determinar si el supuesto de normalidad de los efectos aleatorios puede mantenerse. Su propuesta es una generalización de los gráficos normales. Determinar la bondad del ajuste para el modelo de efectos aleatorios involucra chequear si los supuestos que conciernen con

- la componente determinística  $\mathbf{X}_i\boldsymbol{\beta}$
- la componente aleatoria  $\mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$  para cada unidad de muestreo

se cumplen.

Considere el caso más simple:

$$y_{ij} = \mu + b_i + \epsilon_{ij}$$

para  $j = 1, \dots, n_i$  e  $i = 1, \dots, n$ . Como supuestos se tienen  $\epsilon_{ij} \sim_{iid} N(0, \sigma^2)$  y  $b_i \sim_{iid} N(0, \delta)$ . Si los  $b_i$  fueran observados, entonces las aproximaciones tradicionales pudieran utilizarse para verificar su normalidad. Pero, aún sin observar los  $b_i$ , se pueden “estimar” estas variables aleatorias no observables mediante el uso de la técnica conocida como Bayes empírico. Para el modelo sencillo que se está considerando los estimadores estandarizados de los  $b_i$  pueden escribirse como

$$z_i = \frac{\hat{b}_i}{\sigma_{b_i}} = \frac{\bar{y}_i - \mu}{\sqrt{\frac{\sigma^2}{n_i} + \delta}}$$

donde

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

Si se asume que  $\mu$ ,  $\sigma^2$  y  $\delta$  son conocidos, entonces los  $z_i$ 's son una muestra aleatoria de una  $N(0, 1)$  y se pueden utilizar las técnicas tradicionales para bondad de ajuste.

Una extensión para casos más complejos se puede realizar obteniendo un valor centrado  $\bar{y}_i$  y estandarizado para cada individuo, y bajo los supuestos del modelo

$$z_i = \frac{\mathbf{1}'(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})}{(\mathbf{1}'\mathbf{V}_i\mathbf{1})^{1/2}}$$

para  $i = 1, 2, \dots, n$ . Un gráfico normal de los  $\bar{y}_i$  hallados arriba pueden dar una idea de lo adecuado del modelo.

El estimador Bayes empírico de  $\mathbf{b}_i$  es

$$\begin{aligned} \hat{\mathbf{b}}_i &= E(\mathbf{b}_i | \mathbf{Y}_i) \\ &= \mathbf{D}\mathbf{Z}'_i\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \end{aligned}$$

Asumiendo que  $\boldsymbol{\beta}$  es conocido y que la covarianza marginal de  $\hat{\mathbf{b}}_i$  está dado por

$$Cov(\hat{\mathbf{b}}_i) = \mathbf{D}\mathbf{Z}'_i\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{D}$$

Varias aproximaciones para elaborar Q-Q-plots se pueden realizar para diferentes combinaciones lineales, lo cual permite definir

$$z_i = \frac{\mathbf{c}'\hat{\mathbf{b}}_i}{(\mathbf{c}'Cov(\hat{\mathbf{b}}_i)\mathbf{c})^{1/2}}$$

para  $i = 1, 2, \dots, n$ . Mediante la elección de diferentes formas del vector  $\mathbf{c}$ , el gráfico puede hacerse sensible a diferentes tipos de desviaciones del modelo. Si el modelo asume dos efectos aleatorios para cada individuo (pendiente aleatoria, intercepto aleatorio) se pueden generar dos gráficos Q-Q marginales haciendo  $\mathbf{c} = (1, 0)'$  y  $\mathbf{c} = (0, 1)'$ . En general se puede tener  $\mathbf{c}_u = (u, 1 - u)'$ , para  $0 \leq u \leq 1$ .

# Capítulo 6

## Modelo lineal generalizado mixto

En este capítulo se discute material relacionado al modelamiento de respuestas que no son necesariamente continuas pero que también se recolectan de manera longitudinal o por medio de mediciones repetidas. Suponga que  $\mathbf{Y}$  es el vector de datos observado. Asuma que la distribución de  $\mathbf{Y}$  condicionado en  $\mathbf{b}$  son independientes y obtenidos de una distribución perteneciente a la familia exponencial. Entonces

$$f_{y_i|\mathbf{b}}(y_i|\mathbf{b}, \boldsymbol{\beta}, \phi) = e^{\left\{ \frac{y_i \eta_i + c(\eta_i)}{a(\phi)} + d(y_i, \phi) \right\}} \quad (1)$$

$$\mathbf{b} \sim f_b(\mathbf{b}|\mathbf{D})$$

El modelo lineal generalizado mixto MLMG (o GLIMMIX en inglés) está formulado como:

$$\eta_i = \mathbf{X}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{b}$$

con  $\mathbf{X}'_i$  la  $i$ -ésima fila de  $\mathbf{X}$  y con  $\mathbf{z}'_i$  la  $i$ -ésima fila de  $\mathbf{Z}$ . las distribuciones normal, binomial y Poisson son miembros de esta familia y cada una de ellas tiene asociado un importante modelo de regresión.

### 6.1. Estimación

La verosimilitud para (1) está dada por

$$L(\boldsymbol{\beta}, \phi, \mathbf{D} | \mathbf{Y}) = \int \prod_{i=1}^n f_{y_i | \mathbf{b}}(y_i | \mathbf{b}, \boldsymbol{\beta}, \phi) f_{\mathbf{b}}(\mathbf{b} | \mathbf{D}) d\mathbf{b} \quad (2)$$

Esta función no puede evaluarse en forma cerrada y requiere de métodos numéricos. El método de la cuadratura de Gauss pueden ser una opción adecuada para esta evaluación.

Inicialmente se presenta la aproximación presentada por McGilchrist (1994). Asuma el modelo lineal mixto tradicional con la siguiente estructura

$$\begin{aligned} \mathbf{Y} &= \boldsymbol{\eta} + \boldsymbol{\epsilon} \\ &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \end{aligned}$$

donde  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{D})$ , con  $\mathbf{D}$  conocida. El vector de respuesta media  $\boldsymbol{\eta}$  depende de una componente fija  $\mathbf{X}\boldsymbol{\beta}$ , con  $\mathbf{X}$  la matriz de diseño con constantes conocidas y  $\boldsymbol{\beta}$  el vector de parámetros poblacionales desconocidos. Además asuma que  $\mathbf{Z}\mathbf{b}$  puede ser particionado como

$$\begin{aligned} \mathbf{Z} &= (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k) \\ \mathbf{b}' &= (\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_k) \end{aligned}$$

donde  $\mathbf{b}_j$  tiene  $v_j$  componentes y  $\mathbf{b}_j \sim N(\mathbf{0}, \sigma_j^2 \mathbf{A}_j)$  y además es independiente de las otras componentes  $\mathbf{b}$ . Sea  $\sigma_j^2 = \sigma^2 \theta_j$ , entonces

$$\mathbf{A} = \begin{bmatrix} \theta_1 \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \theta_2 \mathbf{A}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \theta_k \mathbf{A}_k \end{bmatrix}$$

El procedimiento BLUP consiste en la maximización de la suma de dos componentes de log-verosimilitud. Sea

1. Sea  $l_1$  la log-verosimilitud para  $\mathbf{Y}$  dado  $\mathbf{b}$  condicionalmente fijo,
2. sea  $l_2$  la log-verosimilitud para  $\mathbf{b}$  y
3.  $l = l_1 + l_2$

Así,  $l$  representa la “log-verosimilitud” basada en la distribución conjunta de  $\mathbf{Y}$  y  $\mathbf{b}$ .

El procedimiento BLUP selecciona estimadas o predicciones de  $\boldsymbol{\beta}$ ,  $\mathbf{b}$ ,  $\sigma^2$  y los parámetros de  $\mathbf{A}$  que maximicen  $l$ . Resolviendo el sistema de ecuaciones se obtiene

$$\begin{bmatrix} \mathbf{X}'\mathbf{D}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{D}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{D}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{D}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{D}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{D}^{-1}\mathbf{Y} \end{bmatrix}$$

### 6.1.1. Algoritmo para la estimación

Schall (1991) presenta un algoritmo que permite la estimación de las componentes en el MLMG. La función de linkeo  $g(\cdot)$  aplicada a los datos es linealizada mediante una aproximación de Taylor de primer orden.

$$g(\mathbf{Y}) \doteq g(\mu) + (\mathbf{Y} - \mu)g'(\mu) = \mathbf{w}$$

$\mathbf{w}$  es ahora la variable dependiente ajustada. Entonces el modelo se puede re-escribir así

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}g'(\mu)$$

Aquí

$$\begin{aligned} E(\mathbf{w}) &= \mathbf{X}\boldsymbol{\beta} \\ Cov(\mathbf{b}) &= \mathbf{D} \\ Cov(\boldsymbol{\epsilon}g'(\mu)) &= V\left(\frac{\partial\eta}{\partial\mu}\right)^2 = \mathbf{W}^{-1} \end{aligned}$$

Por tanto

$$Cov(\mathbf{w}) = \mathbf{W}^{-1} + \mathbf{UDU}'$$

El modelo  $\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}g'(\mu)$  es un modelo lineal de efectos aleatorios. Con el siguiente algoritmo es posible estimar las componentes del modelo:

**Algoritmo**

1. Dados  $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_c^2$ , estimaciones de  $\sigma_1^2, \sigma_2^2, \dots, \sigma_c^2$ , calcule los estimadores mínimo-cuadrados  $\hat{\beta}$  y  $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_c$  para  $\beta$  y  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_c$  como soluciones para el conjunto de ecuaciones lineales:

$$\mathbf{C} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{1/2}\mathbf{X} & \mathbf{W}^{1/2}\mathbf{U} \\ \mathbf{0} & \mathbf{D}^{1/2} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{1/2}\mathbf{w} \\ \mathbf{0} \end{bmatrix}$$

donde  $\mathbf{D}$ ,  $\mathbf{w}$  y  $\mathbf{W}$  fueron definidos arriba y todas las cantidades desconocidas son evaluadas en sus valores estimados actuales.

2. Sea  $\mathbf{T}^*$  la inversa de la matriz formada por las últimas  $q$  filas y columnas de  $\mathbf{C}'\mathbf{C}$ , particionada de tal manera que sea conforme con  $\mathbf{D}$

$$\begin{bmatrix} \mathbf{T}_{11}^* & \cdots & \mathbf{T}_{1c}^* \\ \vdots & & \vdots \\ \mathbf{T}_{c1}^* & \cdots & \mathbf{T}_{cc}^* \end{bmatrix}$$

Dadas las estimadas  $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_c$  para  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_c$  calcule las estimadas  $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_c^2$ , de  $\sigma_1^2, \sigma_2^2, \dots, \sigma_c^2$ , como

$$\hat{\sigma}_i^2 = \frac{\hat{\mathbf{b}}_i' \hat{\mathbf{b}}_i}{q_i - v_i^*}$$

donde

$$v_i^* = \frac{\text{Traza}(\mathbf{T}_{ii}^*)}{\sigma_i^2}$$

es evaluada en la estimada actual de  $\sigma_i^2$ .

Para obtener un algoritmo para RMLE reemplace el Paso 2 del algoritmo anterior por lo siguiente:

- Sea  $\mathbf{T}$  la matriz formada por las últimas  $q$  filas y columnas de la matriz inversa de  $\mathbf{C}'\mathbf{C}$ , particionada de tal manera que sea conforme con  $\mathbf{D}$

$$\begin{bmatrix} \mathbf{T}_{11} & \cdots & \mathbf{T}_{1c} \\ \vdots & & \vdots \\ \mathbf{T}_{c1} & \cdots & \mathbf{T}_{cc} \end{bmatrix}$$

Dada las estimadas  $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_c$  para  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_c$  calcule las estimadas  $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_c^2$ , de  $\sigma_1^2, \sigma_2^2, \dots, \sigma_c^2$ , como

$$\hat{\sigma}_i^2 = \frac{\hat{\mathbf{b}}_i' \hat{\mathbf{b}}_i}{q_i - v_i}$$

donde

$$v_i = \frac{\text{Traza}(\mathbf{T}_{ii})}{\sigma_i^2}$$

es evaluada en la estimada actual de  $\sigma_i^2$ .

Algunas veces  $\text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{V}$ , esto es, la función varianza de  $\mathbf{Y}$  es una función de un parámetro desconocido de molestia (nuisance). En ese caso la matriz de pesos es  $\sigma^{-2} \mathbf{W}$ . La componente extra de dispersión puede ser estimada por

$$\hat{\sigma}^2 = \frac{(\mathbf{w} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}})' \mathbf{W} (\mathbf{w} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}})}{n - \sum_{i=1}^c (q_i - v_i^*)}$$

en el paso 2 del algoritmo o por

$$\hat{\sigma}^2 = \frac{(\mathbf{w} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}})' \mathbf{W} (\mathbf{w} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}})}{n - p - \sum_{i=1}^c (q_i - v_i)}$$

en el caso de REML.

En el Paso 1, entonces, la matriz  $\mathbf{W}$  debe ser reemplazada por  $\hat{\sigma}^{-2} \mathbf{W}$ .

## 6.2. Otras formas de estimación usando métodos Monte Carlo

Entre las muchas posibilidades de obtener los estimadores se presentan los algoritmos para su estimación expuestos por McCulloch (1997).

### 6.2.1. Algoritmo EM Monte Carlo (MCEM)

El método EM (Expectation-Maximization) fue propuesto por Dempster et al. (1977). Este método tiende a tener una velocidad de convergencia menor cuando los estimadores de las componentes de varianza están cercanas a la frontera del espacio de parámetros. Considere que los efectos aleatorios  $\mathbf{b}$  son los datos faltantes. Denote por  $\mathbf{W}$  los datos completos, esto es  $\mathbf{W} = (\mathbf{Y}, \mathbf{b})$ . La función de verosimilitud para datos completos está dada por

$$\log(L_W) = \sum_i \log(f_{y_i|b}(y_i | \mathbf{b}, \boldsymbol{\beta}, \phi)) + \log(f_b(\mathbf{b} | \mathbf{D})) \quad (3)$$

Esta selección de datos faltantes tiene dos ventajas:

1. Conociendo los  $\mathbf{b}$ , los  $y_i$ 's son independientes.
2. El paso M del algoritmo EM maximiza (3) con respecto a  $\boldsymbol{\beta}$ ,  $\phi$  y  $\mathbf{D}$ .

Debido a que  $\boldsymbol{\beta}$  y  $\phi$  solo entran en el primer término, el paso M con respecto a  $\boldsymbol{\beta}$  y  $\phi$  solo usa  $f_{y|b}$  y por lo tanto es similar a los cálculos en el modelo lineal generalizado estándar cuando los valores de  $\mathbf{b}$  se asumen conocidos.

Maximizar con respecto a  $\mathbf{D}$  es simplemente máxima verosimilitud usando la distribución de  $\mathbf{b}$  después de reemplazar los estadísticos suficientes con sus valores esperados condicionales.

El algoritmo EM toma la siguiente forma:

1. Escoja valores iniciales  $\boldsymbol{\beta}^{(0)}$ ,  $\phi^{(0)}$  y  $\mathbf{D}^{(0)}$ . Haga  $m = 0$ .
2. Calcule (con las esperanzas evaluadas bajo  $\boldsymbol{\beta}^{(m)}$ ,  $\phi^{(m)}$  y  $\mathbf{D}^{(m)}$ .)

- a)  $\boldsymbol{\beta}^{(m+1)}$ ,  $\phi^{(m+1)}$

$$E[\log(f_{y|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\beta}, \phi)) | \mathbf{Y}]$$

- b)  $\mathbf{D}^{(m+1)}$  que maximiza

$$E[\log(f_b(\mathbf{b} | \mathbf{D})) | \mathbf{Y}]$$

- c) Haga  $m = m + 1$

3. Si se obtiene convergencia declare  $\boldsymbol{\beta}^{(m+1)}$ ,  $\phi^{(m+1)}$  y  $\mathbf{D}^{(m+1)}$  como los estimadores de máxima verosimilitud, en otro caso vaya a 2) y repita hasta alcanzar convergencia.

En general, ninguna de las esperanzas en 2a) o en 2b) puede calcularse en forma cerrada para el modelo (1). Esto se debe a que la distribución condicional  $\mathbf{b} | \mathbf{Y}$  involucra  $f_y$ . Usando el algoritmo Metropolis es posible sacar muestras de la distribución condicional de  $\mathbf{b} | \mathbf{Y}$  que no requiere la especificación de  $f_y$ . Usualmente se utilizan aproximaciones Monte Carlo para las esperanzas. El algoritmo Metropolis opera así:

- Especifique una distribución generadora de puntos candidatos  $h_b(\mathbf{b})$ . Denote por  $\mathbf{b}$  el punto actual del proceso de muestreo y por  $\mathbf{b}^*$  el punto candidato generado por la distribución generadora. Sea

$$A(\mathbf{b}, \mathbf{b}^*) = \min \left\{ 1, \frac{f_{b|y}(\mathbf{b}^* | \mathbf{Y}, \boldsymbol{\beta}, \phi, \mathbf{D}) h_b(\mathbf{b})}{f_{b|y}(\mathbf{b} | \mathbf{Y}, \boldsymbol{\beta}, \phi, \mathbf{D}) h_b(\mathbf{b}^*)} \right\}$$

Si se escoge  $h_b = f_b$  entonces se tiene

$$\begin{aligned} \frac{f_{b|y}(\mathbf{b}^* | \mathbf{Y}, \boldsymbol{\beta}, \phi, \mathbf{D}) h_b(\mathbf{b})}{f_{b|y}(\mathbf{b} | \mathbf{Y}, \boldsymbol{\beta}, \phi, \mathbf{D}) h_b(\mathbf{b}^*)} &= \frac{\prod_{i=1}^n f_y(y_i | \mathbf{b}^*, \boldsymbol{\beta}, \phi) f_b(\mathbf{b}^* | \mathbf{D}) f_b(\mathbf{b} | \mathbf{D})}{\prod_{i=1}^n f_y(y_i | \mathbf{b}, \boldsymbol{\beta}, \phi) f_b(\mathbf{b} | \mathbf{D}) f_b(\mathbf{b}^* | \mathbf{D})} \\ &= \frac{\prod_{i=1}^n f_y(y_i | \mathbf{b}^*, \boldsymbol{\beta}, \phi)}{\prod_{i=1}^n f_y(y_i | \mathbf{b}, \boldsymbol{\beta}, \phi)} \end{aligned}$$

Incorporando el paso Metropolis en el algoritmo EM se obtiene el algoritmo MCEM. El algoritmo MCEM toma la siguiente forma:

1. Escoja valores iniciales  $\boldsymbol{\beta}^{(0)}$ ,  $\phi^{(0)}$  y  $\mathbf{D}^{(0)}$ . Haga  $m = 0$ .
2. Genere  $N$  valores  $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(N)}$  de  $f_{b|y}(\mathbf{b}^* | \mathbf{Y}, \boldsymbol{\beta}^{(m)}, \phi^{(m)}, \mathbf{D}^{(m)})$  usando el Algoritmo Metropolis.
  - a) Escoja  $\boldsymbol{\beta}^{(m+1)}$  y  $\phi^{(m+1)}$  que maximicen una estimada Monte Carlo de

$$E[\log(f_{y|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\beta}, \phi)) | \mathbf{Y}]$$

esto es, maximice

$$\frac{1}{N} \sum_{i=1}^N \log(f_{y|b}(\mathbf{Y} | \mathbf{b}^{(i)}, \boldsymbol{\beta}, \phi)) \quad (6)$$

- b) Escoja  $\mathbf{D}^{(m+1)}$  que maximiza

$$\frac{1}{N} \sum_{i=1}^N \log(f_b(\mathbf{b}^{(i)} | \mathbf{D}))$$

- c) Haga  $m = m + 1$

3. Si se obtiene convergencia declare  $\boldsymbol{\beta}^{(m+1)}$ ,  $\phi^{(m+1)}$  y  $\mathbf{D}^{(m+1)}$  como los estimadores de máxima verosimilitud, en otro caso vaya a 2)

### 6.2.2. Algoritmo Monte Carlo Newton-Raphson (MCNR)

Los g.l.m. son usualmente ajustados con un algoritmo Newton-Raphson o con el algoritmo scoring. Tiene sentido desarrollar una simulación análoga al Newton-Raphson para ajustar MLMG (o GLMM en inglés).

Note que cuando la densidad marginal de  $Y$  está formada como una mezcla como en (2) con parámetros separados para  $f_{y|b}$  y  $f_b$ , entonces las ecuaciones de máxima verosimilitud para  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi)$  y  $\mathbf{D}$  toman la forma

$$E \left[ \frac{\partial \log (f_{y|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \middle| \mathbf{Y} \right] \quad (7a)$$

y

$$E \left[ \frac{\partial \log (f_b(\mathbf{b} | \mathbf{D}))}{\partial \mathbf{D}} \middle| \mathbf{Y} \right] \quad (7b)$$

(7b) solo involucra la distribución de  $\mathbf{b}$  y a menudo es fácil de resolver; por ejemplo cuando la distribución es normal, (7a) se puede resolver con Newton-Raphson o con el *scoring* como en los m.l.g. estándar. Expandiendo

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

como una función de  $\boldsymbol{\beta}$  alrededor de  $\boldsymbol{\beta}_0$

$$\begin{aligned} \frac{\partial \log (f_{y|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\theta}))}{\partial \boldsymbol{\beta}} &\approx \frac{\partial \log (f_{y|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\theta}))}{\partial \boldsymbol{\beta}} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &+ \frac{\partial^2 \log (f_{y|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\theta}))}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \end{aligned} \quad (8)$$

El algoritmo scoring se convierte en

$$\begin{aligned} \frac{\partial \log (f_{y|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\theta}))}{\partial \boldsymbol{\beta}} &\approx \mathbf{X}' \mathbf{W}(\boldsymbol{\theta}_0, \mathbf{b}) \frac{\partial \eta}{\partial \boldsymbol{\mu}} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\theta}, \mathbf{b})) \frac{1}{\phi} \\ &- \mathbf{X}' \mathbf{W}(\boldsymbol{\theta}_0, \mathbf{b}) \mathbf{X} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \frac{1}{\phi} \end{aligned} \quad (9)$$

donde

$$\begin{aligned}\mu_i(\boldsymbol{\theta}, \mathbf{b}) &= E[y_i | \mathbf{b}] \\ \mathbf{W}(\boldsymbol{\theta}, \mathbf{b})^{-1} &= \text{Diag} \left\{ \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 \text{Var}(y_i | \mathbf{b}) \right\} \\ \frac{\partial \eta}{\partial \mu} &= \text{Diag} \left\{ \frac{\partial \eta_i}{\partial \mu_i} \right\}\end{aligned}$$

Usar esta aproximación en (7a) conduce a la ecuación de la iteración

$$\begin{aligned}\boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} + E \left[ \mathbf{X}' \mathbf{W}(\boldsymbol{\theta}^{(m)}, \mathbf{b}) \mathbf{X} | \mathbf{Y} \right]^{-1} \mathbf{X}' \\ &\quad \times E \left[ \mathbf{X}' \mathbf{W}(\boldsymbol{\theta}^{(m)}, \mathbf{b}) \frac{\partial \eta}{\partial \mu} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(m)}} (\mathbf{Y} - \mu(\boldsymbol{\beta}^{(m)}, \mathbf{b})) \Big| \mathbf{Y} \right]\end{aligned}$$

Este análogo del scoring procede iterativamente resolviendo (7b), (9) y una ecuación para  $\phi$ . En algunos casos  $\phi$  podría ser igual a uno y por lo tanto no requería ser estimado.

Las esperanzas no pueden ser evaluadas en forma cerrada, lo que lleva al MCNR:

1. Escoja valores iniciales  $\boldsymbol{\beta}^{(0)}$ ,  $\phi^{(0)}$  y  $\mathbf{D}^{(0)}$ . Haga  $m = 0$ .
2. Genere  $N$  valores  $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(N)}$  de  $f_{b|y}(\mathbf{b}^* | \mathbf{Y}, \boldsymbol{\beta}^{(m)}, \phi^{(m)}, \mathbf{D}^{(m)})$  usando el Algoritmo Metropolis y úselos para formar las estimadas Monte Carlo de las esperanzas (denotadas  $\hat{E}[\cdot]$ ).

a) Calcule

$$\begin{aligned}\boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} + E \left[ \mathbf{X}' \mathbf{W}(\boldsymbol{\theta}^{(m)}, \mathbf{b}) \mathbf{X} | \mathbf{Y} \right]^{-1} \mathbf{X}' \\ &\quad \times E \left[ \mathbf{X}' \mathbf{W}(\boldsymbol{\theta}^{(m)}, \mathbf{b}) \frac{\partial \eta}{\partial \mu} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(m)}} (\mathbf{Y} - \mu(\boldsymbol{\beta}^{(m)}, \mathbf{b})) \Big| \mathbf{Y} \right]\end{aligned}$$

b) Calcule  $\phi^{(m+1)}$  para resolver

$$E \left[ \frac{\partial \log(f_{\theta|b}(\mathbf{Y} | \mathbf{b}, \boldsymbol{\theta}))}{\partial \phi} \Big| \mathbf{Y} \right] = 0$$

c) Escoja  $\mathbf{D}^{(m+1)}$  que maximiza

$$\frac{1}{N} \sum_{i=1}^N \log \left( f_b \left( \mathbf{b}^{(i)} \mid \mathbf{D} \right) \right)$$

d) Haga  $m = m + 1$

3. Si se obtiene convergencia declare  $\beta^{(m+1)}$ ,  $\phi^{(m+1)}$  y  $\mathbf{D}^{(m+1)}$  como los estimadores de máxima verosimilitud, en otro caso vaya a 2)

### 6.2.3. Máxima verosimilitud simulada (SML)

$$\begin{aligned} L(\beta, \phi, \mathbf{D} \mid \mathbf{Y}) &= \int f_{y|b}(\mathbf{Y} \mid \mathbf{b}, \beta, \phi) f_b(\mathbf{b} \mid \mathbf{D}) d\mathbf{b} \\ &= \int \frac{f_{y|b}(\mathbf{Y} \mid \mathbf{b}, \beta, \phi) f_b(\mathbf{b} \mid \mathbf{D})}{h_b(\mathbf{b})} h_b(\mathbf{b}) d\mathbf{b} \\ &\approx \frac{1}{N} \sum_{i=1}^N \frac{f_{y|b}(\mathbf{Y} \mid \mathbf{b}^{(i)}, \beta, \phi) f_b(\mathbf{b}^{(i)} \mid \mathbf{D})}{h_b(\mathbf{b}^{(i)})} \end{aligned}$$

Esto produce una estimada insesgada de la verosimilitud. La verosimilitud simulada es entonces maximizada numéricamente.

## 6.3. Ejemplos

### 6.3.1. Genética porcina (respuesta dicótoma)

Al registrarse múltiples observaciones de tipo binomial por sujeto surge un problema de modelamiento que no se debe enfrentar con técnicas estadísticas estándar tales como la regresión logística. Por esto es necesario recurrir a métodos más adecuados para modelar este tipo de datos como por ejemplo los modelos lineales mixtos generalizados (MLMG) que tienen en cuenta no solo la variabilidad entre sujetos sino también la variabilidad intrasujeto a la vez que permite realizar comparaciones múltiples de los niveles de los factores de interés. Un ejemplo donde surgen con frecuencia este tipo de datos es el área de Producción Porcina donde es de fundamental importancia estudiar no solo el tamaño de las camadas sino también la calidad de las mismas. Uno de

los objetivos en este tipo de problemas está relacionado con evaluar el tamaño y la calidad de las camadas producidas por las cerdas criadas de manera controlada en algunas granjas especializadas de un área rural del departamento de Antioquia. Este tamaño y calidad de las camadas está afectado de manera directa e indirecta por múltiples factores ambientales y genéticos. Algunos de estos factores son difíciles de controlar tales como el ambiente, el personal de la granja, las instalaciones, condiciones climáticas, equipos entre otras. Sin embargo, la productividad está muy relacionada con estos factores y por lo tanto es importante identificar de que manera la afectan a fin de adoptar medidas que la puedan incrementar. Por otro lado, algunos factores se controlan día a día en cada uno de los procedimientos como por ejemplo el tipo de semen que se aplica a las cerdas (ya sea el producido en la misma granja o el semen producido en el centro de inseminación); así mismo, otros factores que se espera afecten este proceso están dados por la granja en la cual se realiza la inseminación y el estado fisiológico de la cerda, es decir, si esta es una cerda primeriza, repetidora, desteta o abierta (estos términos se definen más adelante en este mismo ejemplo).

Se ajustó un modelo lineal mixto generalizado para modelar una respuesta dicótoma longitudinal a fin de evaluar si existen diferencias estadísticas significativas entre estos diferentes factores teniendo en cuenta la variabilidad inherente en cada proceso y así poder determinar cómo interactúan unos factores con otros para producir una camada con numerosas crías.

Las preguntas de investigación se listan a continuación:

1. ¿El número de nacidos totales en la camada se ve afectado por el tipo de semen que se le aplique a la cerda?
2. ¿El número de nacidos totales en la camada se ve afectado por la granja en que se realice el procedimiento de inseminación artificial de la cerda?
3. ¿El número de nacidos totales en la camada se ve afectado por el estado fisiológico en que se encuentre la cerda (destetas (D), abiertas (A), primerizas (P) o repetidoras (R))?

### Los datos

Se consideraron 417 cerdas provenientes de dos granjas de un área rural del departamento de Antioquia y a cada una de ellas se les registró el número de partos y el número de crías en cada parto. Estas crías pueden clasificarse

como Nacidas Vivas, Nacidas Muertas o Nacidas Momia. La tabla 1 muestra un descriptivo del número de partos.

Variable	N	Promedio	Des.std	min	max
# de partos	417	1.95	0.57	1	4

**Tabla 6.1.** Análisis descriptivo para el número de partos por cerda.

En la anterior tabla se observa un total de 470 cerdas con un promedio de cerca de 2 partos por cerda; además, se presentaron como mínimo 1 parto y como máximo 4 partos.

### Variable respuesta

Puesto que una cerda puede parir cerdos vivos, muertos o momias, se consideró conveniente crear una categoría con los muertos y las momias y una con los nacidos vivos. De esta manera se asocia un vector de ceros y unos a cada cerda en cada parto donde el valor de 1 corresponde a un nacido vivo y 0 a un nacido muerto. Estas dos categorías son los niveles de la respuesta de interés que se especifica como una proporción: Nacidos Vivos/Total Camada. Estas proporciones, que se denotan  $\pi_{ijkt}$ , se pueden estimar usando la siguiente fracción:

$$\hat{\pi}_{ijkt} = \frac{Y_{ijkt}}{n_{ijkt}}$$

Donde:

$Y_{ijkt}$ : número de nacidos vivos en la combinación de factores  $i, j, k$  al  $t$ -ésimo parto.

$n_{ijkt}$ : total de la camada en la combinación de factores  $i, j, k$  al  $t$ -ésimo parto.

Por lo tanto,  $\hat{\pi}_{ijkt}$  es una estimación de la probabilidad de nacer vivo en la combinación de factores  $i, j, k$  al  $t$ -ésimo parto.

Así,  $\text{logit}(\hat{\pi}_{ijkt})$  es una estimación del log del odds de nacer vivo en la combinación  $i, j, k$  al  $t$ -ésimo parto vs nacer muerto o momia en la combinación  $i, j, k$  al  $t$ -ésimo parto.

Desde el punto de vista poblacional, la distribución de la variable respuesta es una binomial:

$$Y_{ijkt} \sim \text{bin}(n_{ijkt}, \pi_{ijkt})$$

Con:

$$\pi_{ijkt} = P(Y_{ijkt} = 1)$$

### Posibles factores que afectan la variable respuesta

**Tipo de semen:** Esta variable cualitativa tiene dos niveles que son:

- Semen procesado en el centro de inseminación artificial
- Semen procesado en la granja.

**Tipo de cerda:** Variable cualitativa que presenta 4 niveles:

- Cerda primeriza (P): Es una cerda joven que nunca ha tenido partos.
- Cerda desteta (D): Es la cerda que después de tener el parto y cumplir un periodo de lactancia es alejada de sus lechones, para ser inseminada (servida) nuevamente a los más en un periodo de 5 días.
- Cerda abierta(A): Cerda que después de haber cumplido su periodo de lactancia y haber sido desteta no entra nuevamente en celo en menos de 7 días.
- Cerda repetidora(R): Cerda que después de haber sido servida (inseminada), vuelve a entrar en celo, es decir, esta cerda no queda gestante y repite nuevamente su ciclo.

**Granja:** Esta variable cualitativa tiene dos niveles que son:

- Granja A
- Granja B

De los boxplots de adelante, se observan las siguientes características:

- En promedio, las cerdas Primerizas tienen una camada con aproximadamente un total de nacidos vivos de 12 crías, mientras que las cerdas Repetidoras, en promedio, tienen un total de 5 nacidos vivos; por otro lado, en un nivel intermedio se encuentran las cerdas Abiertas y Destetas que en promedio tienen totales de 10 crías nacidas vivas por camada.
- El semen de la granja parece producir un mayor número de total nacidos, ya que en promedio nacen un total de 13 crías aproximadamente, mientras que el laboratorio produce en promedio 10 crías por camada.
- Parece ser que la Granja B, en promedio, produce un mayor número de nacidos vivos por camada aproximadamente de 13 crías; mientras que para la granja A, en promedio, se obtiene un total de 10 nacidos vivos.

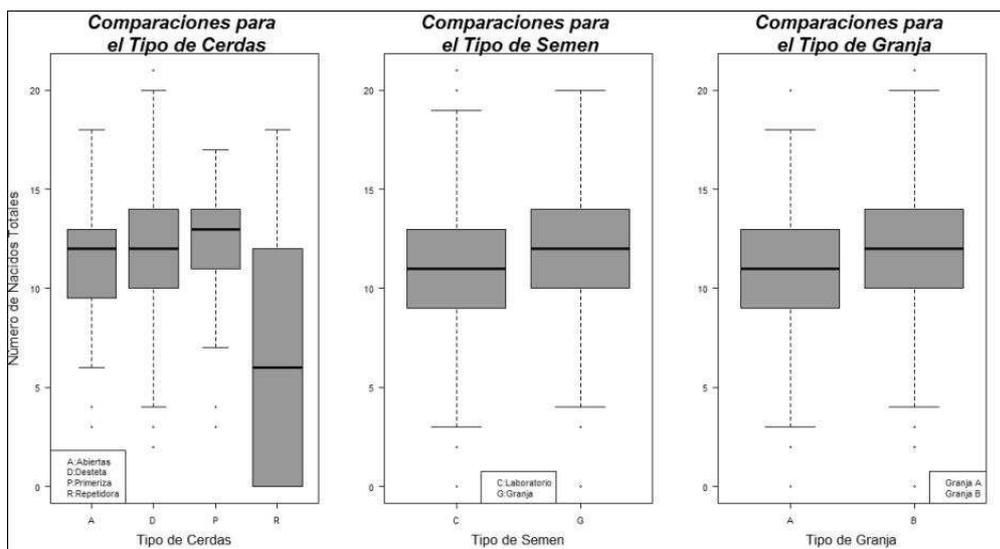


Figura 6.1. Boxplot tipo de cerdas, semen y granja

## Metodología

Se consideraron tres factores que fueron: Tipo de cerda (4 niveles, Primerizas, Destetas, Repetidoras y Abiertas), Tipo de Granja (2 niveles, A y B) y Tipo de Semen (2 niveles, C y G). Puesto que una cerda puede parir cerdos vivos, muertos o momias, se consideró conveniente crear una categoría con

los muertos y las momias y una con los nacidos vivos. De esta manera se asocia un vector de ceros y unos a cada cerda en cada parto donde el valor de 1 corresponde a un nacido vivo y 0 a un nacido muerto o momia. Estas dos categorías son los niveles de la respuesta de interés (Nacidos Vivos/Total Camada). Se especificó un intercepto aleatorio ya que las trayectorias asociadas a cada cerda en cuanto a tamaño de las camadas parece tener su propio intercepto. Para modelar esta respuesta dicótoma longitudinal se ajustó un modelo lineal mixto generalizado donde la función de vínculo es la binomial y se especificó un intercepto aleatorio. Con este modelo se investigaron los efectos fijos y las interacciones. Para seleccionar un modelo razonable se usó un procedimiento Backward basado en el valor-p. Las comparaciones de los niveles de los factores se realizaron por medio de medias de mínimos cuadrados de Fisher usando estadísticos t de Student basados en la aproximación de Satterthwaite. Este modelo se ajustó usando los datos disponibles con la ayuda del PROC GLIMMIX disponible en SAS V. 9.2<sup>©</sup>)

Específicamente se consideró el siguiente modelo:

$$\log \left( \frac{\pi_{ijkt}}{1 - \pi_{ijkt}} \right) = \mu + C_{it} + G_{jt} + S_{kt} + (CG)_{ijt} + (CS)_{ikt} + (GS)_{jkt} + (CSG)_{ijkt}$$

Donde:

$\mu$ : Intercepto

$C_{it}$ : Efecto  $i$ -ésima Cerda ( $i = D, A, P, R$ )

$G_{jt}$ : Efecto  $j$ -ésima Granja ( $j = A, B$ )

$S_{kt}$ : Efecto  $k$ -ésimo Tipo de Semen ( $k = G, C$ )

$(CG)_{ijt}$ : Interacción  $i$ -ésima cerda,  $j$ -ésima Granja

$(CS)_{ikt}$ : Interacción  $i$ -ésima cerda,  $k$ -ésimo Tipo de Semen

$(GS)_{jkt}$ : Interacción  $j$ -ésima Granja,  $k$ -ésimo Tipo de Semen

$(CSG)_{ijkt}$ : Interacción  $j$ -ésima Granja,  $k$ -ésimo Tipo de Semen,  $i$ -ésima Cerda

$t$ : Indica el número de veces que se registra una misma cerda. (medidas repetidas)

$\pi_{ijkt}$ : Probabilidad de nacer vivo en la combinación  $i,j,k$  al  $t$ -ésimo parto

Si una interacción se conserva en el modelo, se deben retener también los efectos fijos involucrados en esa interacción específica.

### Selección Backward

Con un nivel de significancia del 5%, se tiene que al correr el modelo por primera vez, la interacción triple es la menos significativa con  $valor - p = 0,9176 > 0,05$ , por lo que se procede a eliminarla del modelo. Luego se corre el modelo nuevamente y se elimina la nueva interacción que resulte ser menos significativa siguiendo un procedimiento Backward que permita que el modelo solo quede, de ser posible, con solo variables significativas. En la siguiente tabla se muestra el orden de eliminación.

Orden de Eliminación	Variable/interacción	Valor P
Paso 1	Granja * Semen *Cerde	0.9176
Paso 2	Semen* Granja	0.9464
Paso 3	Semen* Cerda	0.6853
Paso 4	Semen	0.4026

**Tabla 6.2.** Orden de eliminación de variables en el proceso de selección Backward

Finalmente el modelo termina seleccionando las siguientes variables:

Factor	Valor P
Tipo Cerda	0.0759
Granja	0.0204
Tipo Cerda * Granja	0.0029

**Tabla 6.3.** Estimaciones modelo final.

### Resultados

Se encontró una asociación de borde entre el factor Cerda y el logaritmo del odds de nacer vivo ( $F=2.30$  basada en 856 g.l,  $valor-p=0.0759$ ). También se observó una diferencia significativa entre el factor tipo de Granja y el logaritmo del odds de nacer vivo ( $F=5.39$  basada en 847 g.l,  $valor-p=0.0204$ ).

La interacción entre Tipo de Cerda y Granja resultó significativa ( $F=4.70$  basada en 871 g.l, valor- $p=0.0029$ ).

Al realizar comparaciones múltiples teniendo en cuenta la interacción, se observaron algunas diferencias significativas:

1. Entre las cerdas Abiertas y Primerizas de la granja B ( $T=2.43$ , 753 g.l, valor- $p=0.0155$ )
2. Entre las cerdas Destetas y Primerizas de la granja B ( $T=4.70$ , 636 g.l, valor- $p<0.0001$ )
3. Entre las Primerizas de la Granjas A y B ( $T=3.67$ , 862, valor- $p=0.0003$ )
4. Entre Abiertas de la Granja A y Primerizas de la Granja B ( $T=2.50$ , 820 g.l., valor- $p=0.0124$ )
5. Entre Destetas de la Granja A y Primerizas de la Granja B ( $T=4.39$ , 621 g.l., valor- $p<0.0001$ )
6. Entre Primerizas de la Granja B y Repetidoras de la Granja A ( $T=-2.43$ , 877 g.l., valor- $p=0.0152$ )

### Conclusiones

- Se observó una diferencia estadística significativa entre el estado fisiológico de la cerda, donde las cerdas Primerizas son las que producen una mayor camada, le siguen las cerdas Destetas y Abiertas y las que menos crías producen por parto son las cerdas Repetidoras. Todo esto con un nivel de confianza del 95 %.
- Se encontró que independientemente del tipo de semen que se le aplique a la cerda, ya sea el procesado en granja o el procesado en laboratorio, este no afecta el número de nacidos totales en la camada de la cerda.
- Se identificó una diferencia estadística significativa entre la Granja, es decir, la granja en que se realice el proceso de inseminación afecta el número de nacidos en la camada.
- Solo la interacción o combinación del Tipo de Cerda con Granja resulta ser significativa. Así, se pueden establecer diferencias al realizar combinaciones entre Tipo de Cerda ya sea Primeriza, Desteta o Abierta combinadas con alguna de las Granjas, y esto influirá en el número de nacidos totales de la camada.

Adicionalmente y de una manera general, se puede argumentar que los MLMG permiten trabajar con la información disponible y no requieren balanceo a la vez que admite respuestas categóricas. Para este tipo de datos binomiales con repeticiones se recomiendan los modelos MLMG. Finalmente, los MLGM son efectivos y prácticos para analizar respuesta binomial con repeticiones y permiten hacer comparaciones múltiples de los niveles de factores.

### Código SAS

```
proc glimmix data=pigs2;
class id TipoCerde Granja;
model NV/Total = TipoCerde Granja TipoCerde*Granja / solution ddfm=satterth
oddsratio;
random intercept / subject=id;
lsmeans Granja TipoCerde*Granja / cl diff;
run;
```

### 6.3.2. Modelo lineal mixto tipo Poisson

Una de las ideas con los modelos lineales mixtos generalizados es especificar el modelo en términos de la distribución de los datos, de la función de la media y expresar esta media en función de alguna transformación de  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$ . La *distribución* se toma como un miembro de la familia exponencial (ver Anexo C para más detalles) y la relación entre la media  $\mu = E(\mathbf{Y})$  y el *predictor lineal*  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$  debe ser una transformación monótona. Es común denotar este predictor lineal como:

$$\eta = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$$

y la transformación de la media como  $g(\mu)$ . A esta última función  $g(\mu)$  se le conoce como *Función de vínculo* o *Función Link* ya que vincula el modelo mixto con la media de  $\mathbf{Y}$ . Puesto que la función de vínculo es monótona, la relación entre  $\mu$  y  $\eta$  (y así entre  $\mu$  y  $\boldsymbol{\beta}$ ) se puede expresar en términos de la inversa de la función de vínculo:

$$\begin{aligned} g(\mu) &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} = \eta \\ \mu &= g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) \end{aligned}$$

En el caso de que se tengan respuestas relacionadas con conteos, se aconseja usar como distribución la Poisson, cuya media condicional está dada por  $\lambda$ , la función de vínculo  $\eta = \log(\lambda) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$  y la inversa de la función de vínculo  $g^{-1}(\eta) = e^\eta$ .

**Ejemplo.** Número de errores cometidos por una persona en una tarea bajo diferentes condiciones.

En este ejemplo se consideran datos sobre el número de errores cometidos por una persona en una tarea bajo diferentes condiciones. En realidad, esta respuesta puede modelarse con un modelo con respuesta repetida tipo Poisson. Se usará entonces para ilustrar, un modelo lineal mixto generalizado con respuesta repetida tipo Poisson (conteo). A continuación se presenta el programa en código R junto con los resultados del ajuste.

**Programa en R:**

```
Y <- c(1,3,5, 2,1,4, 2,2,3,0,1,1, 3,2,3, 1,1,1, 2,3,2, 2,3,3, 1,4,3, 2,2,4, 3,
2,2, 2,4,4, 1,0,3, 2,3,6, 2,3,5, 0,0,2, 3,2,4, 1,3,3)
```

```
Trat <- c(rep(c('I','II','III'),18))
```

```
Subject <- rep(1:18,each=3)
```

```
library(MASS)
```

```
Errores <- data.frame(Y=Y,Trat=Trat,Subject=Subject)
```

```
res <- glmmPQL(Y ~ Trat , random = 1 | Subject,family = poisson, data =
Errores)
```

```
summary(res)
```

**Resultados**

Linear mixed-effects model fit by maximum likelihood

Data: Errores

AIC BIC logLik

NA NA NA

Random effects:

Formula: ~ 1 | Subject

(Intercept) Residual

StdDev: 0.3313524 0.6125762

Variance function:

Structure: fixed weights

Formula:  $\sim$  invwt

Fixed effects:  $Y \sim$  Trat

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.4775209	0.1408649	34	3.389922	0.0018
TratII	0.2623643	0.1530742	34	1.713968	0.0956
TratIII	0.6592456	0.1417545	34	4.650614	0.0000

Correlation:

	(Intr)	TratII
TratII	-0.614	
TratIII	-0.663	0.610

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-2.0298440	-0.4907547	-0.1472716	0.4592057	1.7706241

Number of Observations: 54

Number of Groups: 18

De estos resultados se observa un efecto importante del TratIII (valor- $p < 0.0001$ ). Ambos factores, TratII y TratIII, incrementan el logaritmo del número promedio de errores cometidos en la realización de la tarea específica.

# Capítulo 7

## Modelo no lineal mixto

Los modelos no lineales son importantes en muchas disciplinas donde dada la naturaleza de los fenómenos estudiados, existen modelos que son justificados teóricamente; por ejemplo en procesos físicos, biológicos o bioquímicos, donde se obtienen modelos a partir de soluciones a ecuaciones diferenciales. En otros casos hay modelos no lineales propuestos empíricamente, como ocurre en ciertas aplicaciones en economía o mercadeo. Se menciona también que la aplicación de ciertas transformaciones, estilo Box-Cox, lleva a modelos no lineales, pero en la práctica se tiende a no realizar la estimación simultánea de este parámetro. Por lo tanto se hace una breve introducción y se presentan las recomendaciones, que son muy importantes y fueron dadas por Bates y Watts (1988).

### 7.1. El modelo no lineal tradicional

Recuerde que un modelo es lineal si es lineal en los parámetros, en otro caso es no lineal. Así, se tiene que un modelo  $E(y|x) = \beta_0 + \beta_1x + \beta_2x^2$  es lineal y no lo es  $E(y|x) = 1/(\beta_0 + \beta_1x)$ . Otro problema es la presencia del error en el modelo, ya que la modelación del error:  $y = \beta_0 + \beta_1x + \beta_2x^2 + \epsilon$  conduce a un modelo lineal, pero  $y = (\beta_0 + \beta_1x + \beta_2x^2)\epsilon$  no lo es. Esta última situación ilustra la dificultad que representa algunas representaciones del error.

Bates y Watts (1988) dan la siguiente definición de un modelo no lineal y es la que se va a adoptar en este texto:

**Definición:** Se dice que un modelo es no lineal si al menos una de las derivadas de la función esperanza con respecto a los parámetros depende de al menos uno de los parámetros.

La forma que se asume para los modelos no lineales se representa como,

$$y_i = f(\mathbf{X}_n, \boldsymbol{\theta}) + \epsilon_i$$

donde  $f$  es una función no lineal de  $\mathbf{X}_n$  y de los efectos fijos  $\boldsymbol{\theta}$ . Se asume que el error entra en forma aditiva en el modelo. Además, se asume en muchas situaciones que  $\epsilon_i \sim N(0, \sigma^2)$ . El supuesto de varianza constante puede no ser muy realista en muchos caso, pero la modelación de esta varianza puede ser difícil y por lo tanto en muchos caso se sobre-simplifica asumiendo homoscedasticidad, lo cual pudiera ser equivocado.

## 7.2. Modelos mixtos no lineales

No todos los datos se pueden caracterizar de manera óptima por medio de modelos lineales o modelos lineales generalizados mixtos. Existen situaciones donde es más conveniente usar modelos no lineales; por ejemplo, modelos de crecimiento. Estos modelos no lineales (es decir aquellos cuyos parámetros entran al modelo de manera individual y no lineal).

Tradicionalmente, los modelos mixtos no lineales tienen la forma general,

$$\mathbf{Y} = f(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}) + \mathbf{e}$$

donde  $f$  es una función no lineal de las constantes conocidas  $\mathbf{X}$ , los efectos aleatorios  $\mathbf{Z}$  y los parámetros  $\boldsymbol{\beta}$  y  $\boldsymbol{\gamma}$ . Los efectos aleatorios permiten tener en cuenta la variabilidad intra-individual. Algunos modelos no lineales, se pueden linealizar y otros no. Los primeros, al dejarse linealizar, se pueden trabajar como si fueran lineales, los segundos requieren de estrategias diferentes que tengan en cuenta la naturaleza no lineal de los datos.

### 7.2.1. Modelos no lineales que son linealizables

Un modelo se dice que es “linealizable” si existe una función de la media que es una función lineal de los parámetros del modelo. Es decir, existe una transformación  $g(\cdot)$  de  $E(Y)$  o de  $E(Y|u)$  que hace que  $g(E(Y))$  sea lineal.

Por ejemplo,

$$\mu_Y = \frac{1}{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$$

es linealizable, ya que existe una función  $g(t) = \frac{1}{t}$  tal que

$$g(\mu_Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Algunos modelos mixtos no lineales se pueden linealizar y trabajar con ellos como si se tratara de un modelo lineal mixto. Usualmente esto se logra por medio de transformaciones monótonas invertibles. Para ilustrar, considere un modelo individual no lineal de intercepto aleatorio de la forma

$$\mathbf{Y}_i = e^{\beta_0 + \beta_1^T \mathbf{x}_i + b_i + \epsilon_i}$$

Observe que al tomar logaritmos naturales a ambos lados de esta igualdad se obtiene,

$$\mathbf{Y}_i^* = \log(\mathbf{Y}_i) = \beta_0 + \beta_1 x_1 + b_i + \epsilon_i$$

El cual tiene la forma de un modelo lineal.

Sin embargo, este podría no ser siempre el caso. Considere por ejemplo el siguiente modelo mixto,

$$\mathbf{Y}_i = e^{\beta_0 + \beta_1^T \mathbf{x}_i + b_i} + \epsilon$$

Este modelo, cuyo término de error es aditivo, no se puede linealizar.

Otro ejemplo de un modelo que no se puede linealizar es el modelo logístico con efectos aleatorios:

$$Y_{ij} = \frac{\beta_1 + b_{i1}}{1 + (\beta_2 + b_{i2})e^{-x_{ij}/\beta_3} + \epsilon_{ij}}$$

Donde:

- $i = 1, 2, \dots, n$  (número de sujetos) y  $j = 1, 2, \dots, m$  (número de observaciones tomadas al sujeto  $i$ )
- $\beta_1$ : Asíntota superior poblacional
- $\beta_2$ : Parámetro de forma
- $\beta_3$ : Parámetro de escala

- $b_{i1}$  y  $b_{i2}$ : Parámetros de efectos aleatorios. Estos efectos aleatorios permiten que cada sujeto tenga su propia asíntota y forma centradas a  $\beta_1$  y  $\beta_2$ .

### 7.2.2. Parámetros condicionalmente linealizables

En farmacocinética, el modelo de Michaelis-Menten (Nelson y Cox, 2000) relaciona la velocidad inicial de una reacción enzimática con la concentración de sustrato ( $x$ ) por medio de la ecuación:

$$f(x, \boldsymbol{\theta}) = \frac{\theta_1 x}{\theta_2 + x}$$

Este modelo sólo es válido cuando la concentración del sustrato es mayor que la concentración de la enzima, y para condiciones de estado estacionario, es decir, cuando la concentración del complejo enzima-sustrato es constante. Al calcular las derivadas parciales de  $f$  con respecto a  $\theta_1$  y  $\theta_2$  se obtienen expresiones que dependen de al menos uno de los parámetros  $\theta_1$  y  $\theta_2$  este modelo se considera no lineal. En efecto,

$$\frac{\partial f}{\partial \theta_1} = \frac{x}{\theta_2 + x}$$

$$\frac{\partial f}{\partial \theta_2} = \frac{-\theta_1 x}{(\theta_2 + x)^2}$$

Este modelo también es un ejemplo de un modelo en donde hay un *parámetro condicionalmente lineal*, que en este caso corresponde al parámetro  $\theta_1$ .  $\theta_1$  es condicionalmente lineal ya que la derivada de la esperanza con respecto a  $\theta_1$  no involucra a  $\theta_1$ . De esta manera, se puede estimar a  $\theta_1$ , condicionando en  $\theta_2$  por medio de una regresión lineal de la velocidad sobre  $\frac{x}{\theta_2 + x}$  (Bates y Watts, 1988).

Por ejemplo, en el contexto de modelos no-lineales mixtos, un modelo de la forma

$$y_i = b_i + \alpha \cos(\beta x_i) + \epsilon_i$$

es condicionalmente linealizable respecto al parámetro  $\alpha$ .

El SAS y el R ofrecen distintas estrategias de optimización de las verosimilitudes que se basan en el algoritmo de Newton-Raphson (Kendall, 1989), en el Simplex de Nelder-Mead (Nelder y Mead, 1965) y en métodos de gradiente conjugado (Bulirsch y Stoer, 1991). Es de resaltar que el modelamiento basado en modelos mixtos no lineales puede implicar retos adicionales en

comparación a los modelos lineales mixtos y en general pueden ser más inestables numéricamente que estos últimos.

### 7.3. Extensión al caso de efectos aleatorios

Lindstrom y Bates (1990) extienden la metodología de modelos mixtos al caso no lineal. Se comienza por considerar el modelo

$$y_{ij} = f(\phi_i, \mathbf{X}_i) + e_{ij}$$

donde

- $y_{ij}$  es la  $j$ -ésima respuesta del  $i$ -ésimo individuo,
- $f$  es una función no lineal,
- $\phi_i$  es un vector de parámetros de longitud  $r$ ,
- $e_{ij}$  es el término de error que se asume normalmente distribuido.

Entonces se tiene que,

- $\phi_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i$
- $\mathbf{b}_i \sim (\mathbf{0}, \sigma^2\mathbf{D})$
- $\boldsymbol{\beta}$  es un  $p$ -vector de parámetros poblacionales fijos,
- $\mathbf{b}_i$  es un  $q$ -vector de efectos aleatorios asociados con el individuo  $i$ ,
- $\mathbf{A}_i$  una matriz de diseño de dimensión  $r \times p$  para los efectos fijos,
- $\mathbf{B}_i$  una matriz de diseño de dimensión  $r \times q$  para los efectos aleatorios,
- $\sigma^2\mathbf{D}$  es una matriz de covarianzas.

Entonces la respuesta  $y_{ij}$  se puede expresar como

$$y_{ij} = f(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_{ij})$$

Para el individuo  $i$ , su información se puede expresar en forma matricial como

$$\mathbf{Y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix} \quad \mathbf{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{in_i} \end{bmatrix} \quad \eta_i(\phi_i) = \begin{bmatrix} f(\phi_i, \mathbf{X}_{i1}) \\ f(\phi_i, \mathbf{X}_{i2}) \\ \vdots \\ f(\phi_i, \mathbf{X}_{in_i}) \end{bmatrix}$$

Entonces

$$\begin{aligned} \mathbf{Y}_i &= \eta_i(\phi_i) + \mathbf{e}_i \\ \mathbf{e}_i &\sim N(\mathbf{0}, \sigma^2 \mathbf{\Delta}_i) \end{aligned}$$

donde  $\mathbf{\Delta}_i$  depende de  $i$  solo a través de su dimensión. En muchas situaciones se asume  $\mathbf{\Delta}_i = \mathbf{I}$ , pero en muchas ocasiones se pueden asumir estructuras más complejas, por ejemplo un modelo que se asume con frecuencia cuando se trabaja con datos temporales es asumir un modelo AR(1).

Todas las observaciones se representan en forma matricial así

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{bmatrix} \quad \phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_M \end{bmatrix} \quad \eta(\phi) = \begin{bmatrix} \eta_1(\phi_1) \\ \eta_2(\phi_2) \\ \vdots \\ \eta_M(\phi_M) \end{bmatrix}$$

$$\begin{aligned} \tilde{\mathbf{D}} &= \mathbf{I}_M \otimes \mathbf{D} \\ \mathbf{\Delta} &= \text{Diag}(\mathbf{\Delta}_1, \mathbf{\Delta}_2, \dots, \mathbf{\Delta}_M) \end{aligned}$$

El modelo completo es

$$\begin{aligned} \mathbf{Y} | \mathbf{B} &\sim N(\eta(\phi), \sigma^2 \mathbf{\Delta}) \\ \phi &= \mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b} \\ \mathbf{B} &\sim N(\mathbf{0}, \sigma^2 \mathbf{D}) \end{aligned}$$

donde  $\mathbf{B} = \text{Diag}(\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_M)$

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_M \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_M \end{bmatrix}$$

#### 7.4. Estimación de $\beta$ y $\mathbf{b}$

Cuando los componentes de varianza  $\Delta$  y  $\mathbf{D}$  son conocidos y  $\eta$  es una función lineal de  $\beta$  y  $\mathbf{b}$ , esto es,  $\eta(\mathbf{A}_i\beta + \mathbf{B}_i\mathbf{b}_i) = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i$ , entonces si se define

$$\begin{aligned} \mathbf{X} &= [\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_M]' \\ \mathbf{Z} &= \text{Diag}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_M) \end{aligned}$$

se tiene que los estimadores corrientes de  $\beta$  y  $\mathbf{B}$  corresponden al estimador de mínimos cuadrados generalizados (GLM)

$$\hat{\beta}_{lin} = \hat{\beta}_{lin}(\theta) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

y la media posterior

$$\hat{\mathbf{b}}_{lin} = \hat{\mathbf{b}}_{lin}(\theta) = \tilde{\mathbf{D}}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta}_{lin}(\theta))$$

donde

$$\mathbf{V} = \Delta + \mathbf{Z}\tilde{\mathbf{D}}\mathbf{Z}'$$

y  $\theta$  contiene los elementos únicos de  $\mathbf{D}$  y los de  $\Delta$ .

El vector de parámetros  $\theta$  es una conveniencia que permite escribir todos los parámetros en  $\Delta$  (excepto  $\sigma$ ) en un vector sin necesidad de especificar cómo está parametrizado  $\Delta$ . Note entonces que  $\mathbf{D}$  y  $\Delta$  dependen de  $\theta$  por definición.

Las estimadas de  $\hat{\beta}_{lin}$  y  $\hat{\mathbf{b}}_{lin}$  maximizan conjuntamente la función

$$\begin{aligned} g_{lin}(\beta, \mathbf{b}|\mathbf{Y}) &= -\frac{1}{\sigma^2}(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b}) \\ &\quad -\frac{1}{\sigma^2}\mathbf{b}'\tilde{\mathbf{D}}^{-1}\mathbf{b} \quad (*) \end{aligned}$$

que para  $\beta$  fijo es el logaritmo de la densidad posterior de  $\mathbf{b}$  (hasta una constante) y para  $\mathbf{b}$  fijo es la log-verosimilitud para  $\beta$  (hasta una constante). Los dos términos de (\*) son una suma de cuadrados y un término cuadrático en  $\mathbf{b}$ . Transformando el término cuadrático en  $\mathbf{b}$  a una suma de cuadrados equivalente, se puede tratar la optimización simplemente como un problema de mínimos cuadrados. Es entonces relativamente fácil trasladar esto al caso no lineal.

Se crea este problema de mínimos cuadrados aumentando el vector de datos con “pseudo-datos” como

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\beta + \tilde{\mathbf{Z}}\mathbf{b} + \tilde{\mathbf{e}}$$

donde

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \Delta^{-1/2}\mathbf{Y} \\ \mathbf{0} \end{bmatrix} \quad \tilde{\mathbf{X}} = \begin{bmatrix} \Delta^{-1/2}\mathbf{X} \\ \mathbf{0} \end{bmatrix} \quad \tilde{\mathbf{Z}} = \begin{bmatrix} \Delta^{-1/2}\mathbf{Z} \\ \tilde{\mathbf{D}}^{-1/2} \end{bmatrix}$$

$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

y donde

$$\tilde{\mathbf{D}}^{-1/2} = (\mathbf{L}^{-1/2}, \mathbf{L}^{-1/2}, \dots, \mathbf{L}^{-1/2})$$

siendo  $\mathbf{L}$  el *factor de Cholesky* de  $\Delta_i$ .

En un modelo de efectos mixtos no lineal el e.m.v.  $\hat{\beta}(\theta)$  y la moda posterior  $\hat{\mathbf{B}}(\theta)$  maximizan la función

$$g(\beta, \mathbf{b} | \mathbf{Y}) = -\frac{1}{2\sigma^2} (\mathbf{Y} - \eta(\mathbf{A}\beta + \mathbf{B}\mathbf{b}))' \Delta^{-1} (\mathbf{Y} - \eta(\mathbf{A}\beta + \mathbf{B}\mathbf{b})) - \frac{1}{2\sigma^2} \mathbf{b}' \tilde{\mathbf{D}}^{-1} \mathbf{b}$$

Para  $\beta$  fijo, esta función es el logaritmo de la densidad posterior de  $\mathbf{b}$  (hasta una constante). Se puede probar entonces que el  $\mathbf{b}$  que maximiza  $g$  para un valor dado de  $\beta$  es la moda posterior.  $\hat{\beta}$  es el E.M.V. relativo a una distribución marginal aproximada de  $\mathbf{Y}$ .

Como en el caso lineal estos estimadores pueden calcularse como la solución a un problema (no lineal) de mínimos cuadrados formado por el vector de

datos aumentado con los pseudo-datos como:

$$\tilde{\mathbf{Y}} = \tilde{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}) + \tilde{\mathbf{e}}$$

donde

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \boldsymbol{\Delta}^{-1/2}\mathbf{Y} \\ \mathbf{0} \end{bmatrix} \quad \tilde{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}) = \begin{bmatrix} \boldsymbol{\Delta}^{-1/2}\eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}) \\ \tilde{\mathbf{D}}^{-1/2}\mathbf{b} \end{bmatrix}$$

y con  $\tilde{\mathbf{e}} \sim (\mathbf{0}, \sigma^2\mathbf{I})$

#### 7.4.1. Vía máxima verosimilitud

Se quiere definir el E.M.V. para  $\boldsymbol{\theta}$  con respecto a la densidad marginal de  $\mathbf{Y}$

$$p(\mathbf{Y}) = \int p(\mathbf{Y}|\mathbf{b})p(\mathbf{b})d\mathbf{b}$$

pero debido a que la función esperanza  $\eta$  es una función no lineal de  $\mathbf{b}$ , no existe una expresión cerrada para esta densidad y los cálculos de estas estimadas es difícil. En su lugar se puede aproximar la distribución condicional de  $\mathbf{Y}$  para  $\mathbf{b}$  cerca a  $\hat{\mathbf{b}}(\boldsymbol{\theta})$  mediante una normal multivariable con esperanza que es lineal en  $\mathbf{b}$ . Para lograr esto se aproxima el residual  $\mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b})$  cerca a  $\hat{\mathbf{b}}$  como

$$\mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}) \approx \mathbf{Y} - \left[ \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Z}}\mathbf{b} - \hat{\mathbf{Z}}\hat{\mathbf{b}} \right]$$

donde

$$\hat{\mathbf{Z}}_i = \hat{\mathbf{Z}}_i(\boldsymbol{\theta}) = \left. \frac{\partial \eta_i}{\partial \mathbf{b}'_i} \right|_{\hat{\beta}, \hat{\mathbf{b}}} = \left( \left. \frac{\partial \eta_i}{\partial \phi'} \right|_{\hat{\beta}, \hat{\mathbf{b}}} \right) \mathbf{B}_i$$

y

$$\hat{\mathbf{Z}} = \hat{\mathbf{Z}}(\boldsymbol{\theta}) = \text{Diag} \left( \hat{\mathbf{Z}}_1, \hat{\mathbf{Z}}_2, \dots, \hat{\mathbf{Z}}_M \right) = \left. \frac{\partial \eta}{\partial \mathbf{b}'} \right|_{\hat{\beta}, \hat{\mathbf{b}}}$$

Note que  $\hat{\mathbf{Z}}$  es una función de  $\boldsymbol{\theta}$  ya que  $\hat{\boldsymbol{\beta}}$  y  $\hat{\mathbf{b}}$  lo son. Entonces

$$\mathbf{Y} - \left[ \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Z}}\mathbf{b} - \hat{\mathbf{Z}}\hat{\mathbf{b}} \right] \Big| \mathbf{b} \stackrel{aprox}{\sim} N(\mathbf{0}, \sigma^2\mathbf{A})$$

y la distribución condicional aproximada de  $\mathbf{Y}$  es

$$\mathbf{Y}|\mathbf{b} \stackrel{aprox}{\sim} N \left( \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Z}}\mathbf{b} - \hat{\mathbf{Z}}\hat{\mathbf{b}}, \sigma^2\boldsymbol{\Delta} \right)$$

Esta expresión junto con la distribución de  $\mathbf{b}$ , permite aproximar la distribución marginal de  $\mathbf{Y}$  como

$$\mathbf{Y} \stackrel{approx}{\sim} N\left(\eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) - \hat{\mathbf{Z}}\hat{\mathbf{b}}, \sigma^2\hat{\mathbf{V}}\right)$$

donde

$$\hat{\mathbf{V}} = \hat{\mathbf{V}}(\boldsymbol{\theta}) = \boldsymbol{\Delta} + \hat{\mathbf{Z}}\tilde{\mathbf{D}}\hat{\mathbf{Z}}'$$

la log-verosimilitud que corresponde a la distribución marginal aproximada de  $\mathbf{Y}$  es

$$\begin{aligned} \ell_F(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta} | \mathbf{Y}) &= -\frac{1}{2} \log\left(|\sigma^2\hat{\mathbf{V}}|\right) \\ &\quad - \frac{1}{2\sigma^2} \left[\mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Z}}\hat{\mathbf{b}}\right]' \hat{\mathbf{V}}^{-1} \left[\mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Z}}\hat{\mathbf{b}}\right] \end{aligned}$$

donde  $\hat{\mathbf{b}}$  y  $\hat{\mathbf{Z}}$  dependen de  $\boldsymbol{\theta}$ .

Denote los e.m.v. de  $\boldsymbol{\beta}$ ,  $\sigma$  y  $\boldsymbol{\theta}$  con respecto a  $\ell_F$  como  $\boldsymbol{\beta}^{(ML)}$ ,  $\sigma^{(ML)}$  y  $\boldsymbol{\theta}^{(ML)}$ . Note que, como en el caso lineal, los dos estimadores para  $\boldsymbol{\beta}$  son equivalentes, esto es,  $\boldsymbol{\beta}^{(ML)} = \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}^{(ML)})$ . Esto se sigue del hecho que ellos maximizan a  $\ell_F(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta} | \mathbf{Y})$ .

La matriz inversa de las segundas derivadas del  $\ell_F$  proporciona una estimada para una matriz de varianzas y covarianzas aproximada para  $\boldsymbol{\beta}^{(ML)}$ ,  $\sigma^{(ML)}$  y  $\boldsymbol{\theta}^{(ML)}$ .

#### 7.4.2. Máxima verosimilitud restringida

El método para definir los estimadores Máxima verosimilitud restringida (REML) es el mismo que para los e.m.v. excepto que la  $\ell_F$  en la ecuación de arriba es

$$\ell_R(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta} | \mathbf{Y}) = -\frac{1}{2} \log\left(|\sigma^2\hat{\mathbf{X}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{X}}|\right) + \ell_F(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta} | \mathbf{Y})$$

donde

$$\hat{\mathbf{X}}_i = \hat{\mathbf{X}}_i(\boldsymbol{\theta}) = \left. \frac{\partial \eta_i}{\partial \boldsymbol{\beta}'} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}} = \left( \left. \frac{\partial \eta_i}{\partial \boldsymbol{\phi}'} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}} \right) \mathbf{A}_i$$

y

$$\hat{\mathbf{X}} = \hat{\mathbf{X}}(\boldsymbol{\theta}) = \begin{bmatrix} \hat{\mathbf{X}}_1 \\ \hat{\mathbf{X}}_2 \\ \vdots \\ \hat{\mathbf{X}}_M \end{bmatrix} = \left. \frac{\partial \eta}{\partial \boldsymbol{\beta}'} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}}$$

Defina los estimadores  $\boldsymbol{\beta}^{(RML)}$ ,  $\sigma^{(RML)}$  y  $\boldsymbol{\theta}^{(RML)}$  como aquéllos que maximizan  $\ell_R$ . Además  $\boldsymbol{\beta}^{(RML)} = \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}^{(RML)})$ .

La verosimilitud restringida está basada en  $N - p$  contrastes del error linealmente independientes. En el modelo no lineal la matriz derivada  $\hat{\mathbf{X}}$ , que define estos contrastes de los errores y depende de  $\hat{\boldsymbol{\beta}}$  y  $\hat{\mathbf{b}}$ . Sin embargo, ya que el espacio generado por las columnas de esta matriz dependen de la no linealidad intrínseca y no de los parámetros de efectos no lineales, será casi constante cerca de las estimadas.

## 7.5. Algoritmo de dos pasos

Se estudia un algoritmo de dos pasos para hallar  $\boldsymbol{\theta}^{(ML)}$ .

### 1. Paso pseudo-datos (PD)

Dada la actual estimada  $\boldsymbol{\theta}^{(m)}$  para  $\boldsymbol{\theta}$ , calcule

$$\begin{aligned} \boldsymbol{\beta}^{(m)} &= \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}^{(m)}) \\ \mathbf{b}^{(m)} &= \hat{\mathbf{b}}(\boldsymbol{\theta}^{(m)}) \\ \mathbf{Z}^{(m)} &= \hat{\mathbf{Z}}(\boldsymbol{\theta}^{(m)}) \end{aligned}$$

### 2. Paso efectos lineales mixtos (LME)

Dados  $\mathbf{b}^{(m)}$ ,  $\mathbf{Z}^{(m)}$ , sean  $\boldsymbol{\beta}^{(m+1)}$ ,  $\sigma^{(m+1)}$  y  $\boldsymbol{\theta}^{(m+1)}$  son los valores que maximizan

$$\begin{aligned} \ell_F(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta} | \mathbf{Y}) &= -\frac{1}{2} \log \left( \left| \sigma^2 \left( \boldsymbol{\Delta} + \mathbf{Z}^{(m)} \tilde{\mathbf{D}} \mathbf{Z}^{(m)'} \right) \right| \right) \\ &\quad - \frac{1}{2\sigma^2} \left[ \mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}^{(m)}) + \mathbf{Z}^{(m)}\mathbf{b}^{(m)} + \mathbf{Z}^{(m)}\mathbf{b}^{(m)} \right]' \\ &\quad \times \left( \boldsymbol{\Delta} + \mathbf{Z}^{(m)} \tilde{\mathbf{D}} \mathbf{Z}^{(m)'} \right)^{-1} \\ &\quad \times \left[ \mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}^{(m)}) + \mathbf{Z}^{(m)}\mathbf{b}^{(m)} + \mathbf{Z}^{(m)}\mathbf{b}^{(m)} \right] \end{aligned}$$

Se escribe la dependencia de  $\ell_F$  de  $\boldsymbol{\theta}^m$  en el cual son evaluados  $\hat{\mathbf{b}}$  y  $\hat{\mathbf{Z}}$ .

El algoritmo entonces consiste en iterar los dos pasos hasta lograr convergencia.

Lo que se llamó el Paso LME no corresponde exactamente a una estimación de efectos lineales mixtos ya que el residual

$$\mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}^{(m)}) + \mathbf{Z}^{(m)}\mathbf{b}^{(m)}$$

es no lineal en  $\boldsymbol{\beta}$ . Se puede usar el método de Newton-Raphson para optimizar  $\ell_F$  pero esto requeriría la segunda derivada de la función  $\eta$ , con respecto a los efectos fijos  $\boldsymbol{\beta}$ . Para evitar esto, se puede aproximar los residuales cerca a  $\boldsymbol{\beta}^{(m)}$  por

$$\begin{aligned} \mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}^{(m)}) + \mathbf{Z}^{(m)}\mathbf{b}^{(m)} &\approx \mathbf{Y} - \left[ \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}^{(m)}) \right. \\ &\quad \left. + \mathbf{X}^{(m)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}) - \mathbf{Z}^{(m)}\mathbf{b}^{(m)} \right] \\ &= \mathbf{w}^{(m)} - \mathbf{X}^{(m)}\boldsymbol{\beta} \end{aligned}$$

donde

$$\begin{aligned} \mathbf{X}^{(m)} &= \hat{\mathbf{X}}(\boldsymbol{\theta}^{(m)}) \\ \mathbf{w}^{(m)} &= \mathbf{Y} - \eta(\mathbf{A}\boldsymbol{\beta}^{(m)} + \mathbf{B}\mathbf{b}^{(m)}) + \mathbf{X}^{(m)}\boldsymbol{\beta}^{(m)} + \mathbf{Z}^{(m)}\mathbf{b}^{(m)} \end{aligned}$$

Si se define

$$\begin{aligned} \tilde{\ell}_F(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta} | \mathbf{Y}) &= -\frac{1}{2} \log \left( \left| \sigma^2 (\boldsymbol{\Delta} + \mathbf{Z}^{(m)} \tilde{\mathbf{D}} \mathbf{Z}^{(m)'}) \right| \right) \\ &\quad - \frac{1}{2\sigma^2} \left[ \mathbf{w}^{(m)} - \mathbf{X}^{(m)}\boldsymbol{\beta} \right]' \left( \boldsymbol{\Delta} + \mathbf{Z}^{(m)} \tilde{\mathbf{D}} \mathbf{Z}^{(m)'})^{-1} \left[ \mathbf{w}^{(m)} - \mathbf{X}^{(m)}\boldsymbol{\beta} \right] \end{aligned}$$

entonces el paso LME sustituyendo  $\ell_F$  con  $\tilde{\ell}_F$  es un problema de estimación de efectos lineales mixtos del tipo de Laird y Ware. Este nuevo paso LME resultará en las estimaciones deseadas ya que  $\boldsymbol{\beta}^{(ML)}$ ,  $\sigma^{(ML)}$  y  $\boldsymbol{\theta}^{(ML)}$  maximizan  $\tilde{\ell}_F$ .

El método para obtener las estimadas RML es exactamente el mismo que para los e.m.v. excepto que  $\tilde{\ell}_F$  en la ecuación anterior es reemplazado por

$$\begin{aligned} \tilde{\ell}_F(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta} | \mathbf{Y}) &= -\frac{1}{2} \log \left( \left| \sigma^2 \mathbf{X}^{(m)'} \left( \boldsymbol{\Delta} + \mathbf{Z}^{(m)} \tilde{\mathbf{D}} \mathbf{Z}^{(m)'} \right)^{-1} \mathbf{X}^{(m)} \right| \right) \\ &\quad + \tilde{\ell}_F(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta} | \mathbf{Y}) \end{aligned}$$

## 7.6. Ejemplo

Considere el siguiente conjunto de datos acerca de logaritmos del crecimiento de árboles de abeto

Obs	Grupo	Id	$y_{152}$	$y_{174}$	$y_{201}$	$y_{227}$
1	Control	1	2.79	3.10	3.30	3.38
2	Control	2	3.30	3.90	4.34	4.96
3	Control	3	3.98	4.36	4.79	4.99
4	Control	4	4.36	4.77	5.10	5.30
5	Control	5	4.34	4.95	5.42	5.97
6	Control	6	4.59	5.08	5.36	5.76
7	Control	7	4.41	4.56	4.95	5.23
8	Control	8	4.24	4.64	4.95	5.38
9	Control	9	4.82	5.17	5.76	6.12
10	Control	10	3.84	4.17	4.67	4.67
11	Control	11	4.07	4.31	4.90	5.10
12	Control	12	4.28	4.80	5.27	5.55
13	Ozono	13	4.53	5.05	5.18	5.41
14	Ozono	14	4.97	5.32	5.83	6.29
15	Ozono	15	4.37	4.81	5.03	5.19
16	Ozono	16	4.58	4.99	5.37	5.68
17	Ozono	17	4.00	4.50	4.92	5.44
18	Ozono	18	4.73	5.05	5.33	5.92
19	Ozono	19	5.15	5.63	6.11	6.39
20	Ozono	20	4.10	4.46	4.84	5.29
21	Ozono	21	4.56	5.12	5.40	5.69
22	Ozono	22	5.16	5.49	5.74	6.05
23	Ozono	23	5.46	5.79	6.12	6.41
24	Ozono	24	4.52	4.91	5.04	5.71

Observe que en este conjunto de datos, se tiene el mismo número de árboles por grupo (12) y además se registraron 4 medidas de crecimiento a los días

152, 174, 201 y 227, respectivamente. Es decir, cada árbol contribuye con 4 mediciones. El objetivo del estudio es comparar el crecimiento de los árboles en dos ambientes diferentes: Ambiente de control y Ambiente rico en ozono. Ajuste un MLM a estos datos para comparar el crecimiento en estos dos ambientes. Ajuste un modelo no lineal a estos datos.

### Solución

Para ilustrar, se usará un modelo logístico con efectos aleatorios:

$$Y_{ij} = \frac{\beta_1 + b_{i1}}{1 + (\beta_2 + b_{i2})e^{-x_{ij}/\beta_3}} + \epsilon_{ij}$$

Donde:

- $i = 1, 2, \dots, n$  (número de sujetos) y  $j = 1, 2, \dots, m$  (número de observaciones tomadas al sujeto  $i$ )
- $\beta_1$ : Asíntota superior poblacional
- $\beta_2$ : Parámetro de forma
- $\beta_3$ : Parámetro de escala
- $b_{i1}$  y  $b_{i2}$ : Parámetros de efectos aleatorios. Estos efectos aleatorios permiten que cada sujeto tenga su propia asíntota y forma centradas en  $\beta_1$  y  $\beta_2$ .

El siguiente programa en SAS permite ajustar este modelo,

```
*Modelo no lineal mixto. Arboles de abeto;
data abetos;
input Obs Grupo$ Id y152 y174 y201 y227;
arbol=id;
*clase=0;
*if grupo='Control' then clase=1;
cards;
1 Control 1 2.79 3.10 3.30 3.38
2 Control 2 3.30 3.90 4.34 4.96
3 Control 3 3.98 4.36 4.79 4.99
4 Control 4 4.36 4.77 5.10 5.30
5 Control 5 4.34 4.95 5.42 5.97
6 Control 6 4.59 5.08 5.36 5.76
```

7	Control	7	4.41	4.56	4.95	5.23
8	Control	8	4.24	4.64	4.95	5.38
9	Control	9	4.82	5.17	5.76	6.12
10	Control	10	3.84	4.17	4.67	4.67
11	Control	11	4.07	4.31	4.90	5.10
12	Control	12	4.28	4.80	5.27	5.55
13	Ozono	13	4.53	5.05	5.18	5.41
14	Ozono	14	4.97	5.32	5.83	6.29
15	Ozono	15	4.37	4.81	5.03	5.19
16	Ozono	16	4.58	4.99	5.37	5.68
17	Ozono	17	4.00	4.50	4.92	5.44
18	Ozono	18	4.73	5.05	5.33	5.92
19	Ozono	19	5.15	5.63	6.11	6.39
20	Ozono	20	4.10	4.46	4.84	5.29
21	Ozono	21	4.56	5.12	5.40	5.69
22	Ozono	22	5.16	5.49	5.74	6.05
23	Ozono	23	5.46	5.79	6.12	6.41
24	Ozono	24	4.52	4.91	5.04	5.71

```

;
run;
data abetos1;
set abetos;
if tiempo=1 then tiempo=152;
if tiempo=2 then tiempo=174;
if tiempo=3 then tiempo=201;
if tiempo=4 then tiempo=227;
array xxxx{4} y152 y174 y201 y227;
do i=1 to 4;
y=xxxx{i};
tiempo=i;
output;
end;
run;
proc sort data=abetos1;
by grupo;
run;
proc nlmixed data=abetos1 method=fo;
by grupo;
parms beta1=190 beta2=10 beta3=100 s2b1=500 c12=1.62 s2b2=10 s2e=11.1;
bounds s2b1>=0, s2b2>=0, s2e>=0;

```

```

num=beta1+b1;
ex=exp(-(tiempo)/beta3);
den=1+(beta2+b2)*ex;
model y~normal((num/den),s2e);
random b1 b2 ~normal([0,0],[s2b1,c12,s2b2]) subject=arbol;
title 'Modelo no lineal. Arboles de abeto';
run;

```

Modelo no lineal. Arboles de abeto

----- Grupo=Control -----

The NL MIXED Procedure

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood	-23.4
AIC (smaller is better)	-9.4
AICC (smaller is better)	-6.6
BIC (smaller is better)	-6.0

Parameter Estimates

Parameter	Standard		DF	t Value	Pr >  t	Alpha	Lower	Upper	Gradient
	Estimate	Error							
beta1	6.6267	0.6352	10	10.43	<.0001	0.05	5.2114	8.0420	-0.00002
beta2	0.8233	0.1477	10	5.57	0.0002	0.05	0.4941	1.1524	5.568E-6
beta3	3.6053	0.7539	10	4.78	0.0007	0.05	1.9254	5.2851	0.000017
s2b1	1.0238	0.5786	10	1.77	0.1072	0.05	-0.2653	2.3129	5.976E-6
c12	0.1387	0.1255	10	1.11	0.2950	0.05	-0.1410	0.4184	-2.28E-7
s2b2	0.05676	0.03979	10	1.43	0.1842	0.05	-0.03189	0.1454	-0.00029
s2e	0.005946	0.001716	10	3.46	0.0061	0.05	0.002122	0.009769	0.000804

----- Grupo=Ozono -----

Fit Statistics

-2 Log Likelihood	-34.8
AIC (smaller is better)	-20.8
AICC (smaller is better)	-18.0
BIC (smaller is better)	-17.4

Parameter Estimates

Parameter	Standard		DF	t Value	Pr >  t	Alpha	Lower	Upper	Gradient
	Estimate	Error							
beta1	8.7297	2.9178	10	2.99	0.0135	0.05	2.2284	15.2310	0.000346
beta2	1.0288	0.6479	10	1.59	0.1434	0.05	-0.4149	2.4724	-0.00219
beta3	5.6977	2.8785	10	1.98	0.0759	0.05	-0.7159	12.1114	0.000301
s2b1	0.6973	1.6254	10	0.43	0.6770	0.05	-2.9244	4.3190	-0.00013
c12	0.1754	0.5056	10	0.35	0.7359	0.05	-0.9513	1.3020	-0.00306
s2b2	0.08188	0.1725	10	0.47	0.6452	0.05	-0.3024	0.4662	0.010424
s2e	0.007008	0.002032	10	3.45	0.0062	0.05	0.002481	0.01154	0.003209

Observe que la estimación de la asíntota superior para los árboles que fueron medidos en ambiente rico en ozono (8.7297) es mayor que la estimación de la asíntota para los árboles del ambiente control (6.6267). Es de anotar que esta conclusión también se obtuvo con el modelo lineal mixto que se ajustó a estos mismos datos en el capítulo 4 .

En general, los modelos no lineales son muy sensibles a la elección de los valores iniciales para los parámetros y encontrarlos representa un reto adicional. Sin embargo, tienen la ventaja de que en muchos casos describen de una mejor manera los fenómenos que los modelos lineales y además sus parámetros tienen, en general, interpretaciones muy claras y de gran utilidad en la práctica.

# Capítulo 8

## Ejercicios Propuestos

Los ejercicios propuestos se presentan en un capítulo separado ya que cada uno de ellos necesita de los elementos presentados en los diversos capítulos expuestos. La idea es implementar primero un completo análisis descriptivo de los datos y luego formular (escribiendo el modelo), ajustar y validar un modelo apropiado en cada situación. Se puede usar R o SAS para su solución.

### 8.1. Tratamiento con Enalapril

Chen et al. (1999) presentan datos correspondientes a un experimento realizado para determinar si el Enalapril podría decrecer la tasa de degeneración de los riñones en diabéticos. Los sujetos en el ensayo fueron asignados aleatoriamente al grupo de Enalapril (tratamiento) o a un grupo control. Las mediciones de la tasa de filtración glomerular (GFR) de los sujetos fueron hechas al comienzo del estudio y cada tres meses durante dieciocho meses. La hipótesis de investigación era que la GFR en el grupo del Enalapril no decrecería tan rápidamente como en el grupo de control.

Sujeto	Meses						
	0	3	6	9	12	15	18
Grupo de Tratamiento							
1	65.70	65.20	64.80	74.90	58.80	60.50	56.60
2	26.80	25.30	22.60	24.30	20.40	19.10	20.60
3	50.20	44.20	40.00	44.40	44.50	41.90	37.60
4	93.60	72.60	87.90	91.30	96.50	101.30	104.40
5	44.00	47.00	68.00	44.80	38.90	29.20	40.60
6	62.50	67.30	59.50	38.70	50.00	36.40	34.50
7	70.60	71.00	70.80	81.80	61.60	74.00	71.10
8	100.40	106.00	100.10	101.20	98.20	105.40	88.00
9	104.60	100.00	88.20	82.10	95.50	87.20	113.40
10	89.20	93.10	88.30	77.60	70.90	86.00	74.00
11	32.50	31.30	27.90	27.80	20.50	23.70	21.60
Grupo de Control							
12	54.00	51.10	49.30	38.70	49.50	41.30	55.00
13	71.00	72.20	71.20	52.80	62.00	70.70	69.00
14	103.00	92.70	98.90	83.10	114.30	105.70	82.80
15	107.00	66.60	77.50	53.00	80.70	52.40	61.40
16	55.30	43.10	55.40	44.20	36.70	38.40	35.10
17	54.60	37.30	33.90	36.00	40.20	36.00	32.00
18	88.30	65.00	63.80	77.90	69.60	71.50	75.00
19	51.00	39.10	31.70	41.40	40.60	39.90	39.10
20	46.00	39.20	44.70	45.30	49.80	48.50	55.70
21	60.60	57.00	62.30	68.40	59.30	60.20	68.90
22	29.30	29.30	26.90	17.30	25.30	40.20	30.80

**Tabla 8.1.** Datos tratamiento con enalapril. Mediciones de tasas de filtración glomerular.

## 8.2. Resultados de tratamientos en clínicas

Booth y Hobert (1998) presentan la siguiente tabla que presenta los resultados de dos tratamientos en varias clínicas.

Clínica	1	2	3	4	5	6	7	8
Trt. 1 (Éxitos/Total)	11/36	16/20	14/19	2/16	6/17	1/11	1/5	4/6
Trt. 2 (Éxitos/Total)	10/37	22/32	7/19	1/17	0/12	0/10	1/9	6/7

**Tabla 8.2.** Resultados de tratamientos en varias clínicas.

### 8.3. Duración de pilas alcalinas

Se seleccionaron 6 tipos de pilas alcalinas: RAYOVAC<sup>®</sup>, DURACELL<sup>®</sup>, VARTA<sup>®</sup>, EVEREADY ALKALINE<sup>®</sup>, PANASONIC POWER ALKALINE<sup>®</sup> y ENERGIZER MAX<sup>®</sup>. De cada una se tomaron 2 pilas y se conectaron a un sistema con resistencias paralelas que permitió descargar cada una de las 12 pilas. Cada hora y durante 11 horas se registró la disminución en el voltaje (milivoltios). Los datos están en las siguientes tablas:

	RAYOVAC		DURACELL		VARTA	
	VOLTAJE		VOLTAJE		VOLTAJE	
TIEMPO	PILA 1	PILA 2	PILA 1	PILA 2	PILA 1	PILA 2
0	1217	1288	1419	1450	1430	1308
1	1181	1248	1359	1393	1207	1150
2	1077	1170	1284	1296	1112	1052
3	1040	1138	1222	1235	1056	993
4	1043	1099	1173	1178	1025	953
5	1004	1047	1103	1106	992	917
6	949	968	877	814	952	818
7	873	870	156	138	870	805
8	780	740	107	112	246	172
9	641	618	89	84	96	77
10	62	62	71	78	69	65
11	50	58	61	75	64	60

	EVEREADY		PANASONIC		ENERGIZER	
	VOLTAJE		VOLTAJE		VOLTAJE	
TIEMPO	PILA 1	PILA 2	PILA 1	PILA 2	PILA 1	PILA 2
0	1276	1212	1217	1279	1286	1268
1	1086	1001	1005	1044	1102	1060
2	1034	955	943	980	1036	995
3	992	910	897	958	1000	962
4	959	879	871	930	972	940
5	886	844	804	879	934	905
6	825	788	718	823	874	853
7	693	679	602	736	759	763
8	118.8	142.4	718	612	166	109.2
9	75.5	75.1	602	95.5	75.4	70.4
10	64.7	64.3	117.6	73	62.6	66
11	61.2	61.5	81.7	70.3	60.1	65.1

**Tabla 8.3.** Datos de registros de tiempos de duración de algunas baterías alcalinas.

# Apéndice A

## Algunas bases de datos

En este anexo se presentan algunas de las bases de datos usadas en el texto.

### A.1. Electroplateado

Placa	Medición			Tratamiento
	1	2	3	
1	0.52	0.49	0.49	I
2	0.53	0.51	0.50	I
3	0.52	0.49	0.51	I
4	0.51	0.53	0.52	I
5	0.54	0.52	0.50	I
6	0.52	0.51	0.52	I
7	0.54	0.51	0.53	II
8	0.54	0.56	0.54	II
9	0.53	0.55	0.52	II
10	0.52	0.53	0.53	II
11	0.54	0.54	0.53	II
12	0.52	0.51	0.54	II
13	0.56	0.57	0.57	III
14	0.58	0.55	0.57	III
15	0.57	0.58	0.57	III
16	0.56	0.58	0.56	III
17	0.55	0.57	0.58	III
18	0.57	0.54	0.56	III

**A.2. Hueso ramus**

Individuo	Edad			
	8.0	8.5	9.0	9.5
1	47.80	48.80	49.00	49.70
2	46.40	47.30	47.70	48.40
3	46.30	46.80	47.80	48.50
4	45.10	45.30	46.10	47.20
5	47.60	48.50	48.90	49.30
6	52.50	53.20	53.30	53.70
7	51.20	53.00	54.30	54.50
8	49.80	50.00	50.30	52.70
9	48.10	50.80	52.30	54.40
10	45.00	47.00	47.30	48.30
11	51.20	51.40	51.60	51.90
12	48.50	49.20	53.00	55.50
13	52.10	52.80	53.70	55.00
14	48.20	48.90	49.30	49.80
15	49.60	50.40	51.20	51.80
16	50.70	51.70	52.70	53.30
17	47.20	47.70	48.40	49.50
18	53.30	54.60	55.10	55.30
19	46.20	47.50	48.10	48.40
20	46.30	47.60	51.30	51.80

**A.3. Toros**

Toro	Porcentaje/100								
1	0.46	0.31	0.37	0.62	0.30				
2	0.70	0.59							
3	0.52	0.44	0.57	0.40	0.67	0.64	0.70		
4	0.47	0.21	0.70	0.46	0.14				
5	0.42	0.64	0.50	0.69	0.77	0.81	0.87		
6	0.35	0.68	0.59	0.38	0.57	0.76	0.57	0.29	0.60

### A.4. Secado de madera

Listón	Grosor	Inicio	7	14	21	28	35	42	49	56	62	70	92
1	1	15.30	13.05	12.10	11.50	11.00	10.70	10.50	10.40	10.40	10.40	10.40	10.40
2	1	13.30	12.00	11.60	11.20	11.20	11.00	10.95	10.90	10.90	10.90	10.90	10.90
3	1	15.20	13.25	12.20	11.50	11.00	10.60	10.60	10.40	10.40	10.40	10.40	10.40
4	1	19.25	16.30	14.90	13.70	13.00	12.20	11.70	11.10	10.90	10.50	10.45	10.45
5	1	15.50	13.90	13.20	12.50	12.20	11.80	11.80	11.65	11.55	11.55	11.55	11.55
6	1	12.00	11.05	10.60	10.30	10.10	9.90	9.90	9.90	9.90	9.90	9.90	9.90
7	1	13.05	11.75	10.95	10.50	10.00	9.80	9.70	9.60	9.58	9.50	9.50	9.50
8	1	15.20	13.08	12.40	11.82	11.40	11.20	11.20	11.10	11.10	11.00	11.00	11.00
9	1	13.55	11.20	10.50	10.02	9.70	9.60	9.50	9.40	9.60	9.50	9.50	9.50
10	1	14.60	11.20	10.20	9.90	9.80	9.60	9.60	9.60	9.55	9.55	9.55	9.55
11	1	14.70	12.90	12.00	11.10	10.65	10.40	10.30	10.00	10.00	10.00	10.00	10.00
12	1	14.70	12.70	11.80	11.05	10.80	10.40	10.30	10.20	10.15	10.10	10.10	10.10
13	1	15.30	13.30	12.40	12.00	11.60	11.45	11.40	11.40	11.40	11.40	11.35	11.35
14	1	16.80	13.90	13.10	12.40	12.00	11.65	11.50	11.30	11.30	11.30	11.30	11.30
15	1	12.00	11.15	10.80	10.40	10.20	10.15	10.15	10.10	10.20	10.10	10.10	10.10
16	1	15.65	13.65	12.90	12.20	11.80	11.40	11.20	10.90	10.90	10.80	10.80	10.80
17	2	15.30	13.10	12.40	11.80	11.50	11.20	11.10	10.95	10.95	10.95	10.95	10.95
18	2	8.60	15.90	14.70	13.50	12.80	12.10	11.60	11.10	11.10	10.70	10.50	10.15
19	2	12.90	11.80	11.20	10.80	10.60	10.40	10.40	10.20	10.20	10.25	10.25	10.25
20	2	13.00	10.80	10.10	9.80	9.60	9.50	9.40	9.40	9.40	9.40	9.40	9.40
21	2	13.00	11.25	10.05	9.62	9.15	8.83	8.66	8.46	8.25	8.00	8.02	8.02
22	2	14.35	12.60	11.60	10.75	10.30	9.90	9.89	9.20	9.10	9.00	9.03	9.00
23	2	10.10	9.28	8.72	8.50	8.35	8.30	8.30	8.10	7.90	7.85	7.88	7.88
24	2	12.70	10.65	9.80	9.15	8.73	8.44	8.28	7.96	7.65	7.65	7.67	7.67
25	2	11.50	10.20	9.82	9.55	9.37	9.32	9.31	8.95	8.85	8.90	8.94	8.94
26	2	10.40	10.05	9.89	9.72	9.61	9.59	9.58	9.39	9.20	9.20	9.23	9.23
27	2	11.60	10.25	9.60	9.18	8.93	8.79	8.74	8.30	8.30	8.30	8.32	8.30
28	2	10.60	9.72	9.12	8.91	8.78	8.74	8.73	8.56	8.40	8.40	8.41	8.40
29	2	11.55	10.00	9.44	9.26	9.13	9.11	9.11	8.65	8.45	8.65	8.68	8.68
30	2	12.40	10.10	9.69	9.52	9.40	9.40	9.54	8.95	9.10	9.00	9.00	9.00
31	2	10.25	9.09	8.40	8.20	8.09	8.06	8.05	7.70	7.65	7.65	7.67	7.67
32	2	10.60	9.25	8.49	8.15	7.95	7.90	7.89	7.45	7.60	7.40	7.40	7.40
33	2	10.80	10.00	9.66	9.44	9.29	9.24	9.29	8.75	8.95	8.80	8.83	8.83
34	2	9.90	8.90	8.32	8.16	8.02	8.00	8.00	7.80	7.80	7.60	7.63	7.60
35	2	12.55	9.75	9.35	9.21	9.10	9.10	9.10	8.97	8.85	8.75	8.79	8.80
36	2	10.50	9.38	8.70	8.35	8.14	8.06	8.06	7.60	7.60	7.55	7.55	7.50
37	2	14.00	11.80	10.63	9.80	9.68	9.30	9.13	8.89	8.65	8.60	8.62	8.60
38	2	10.50	8.28	8.68	8.44	8.32	8.22	8.22	7.95	7.80	7.80	7.80	7.80
39	2	11.70	10.40	9.80	9.77	9.58	9.49	9.43	9.16	9.10	9.00	9.00	9.00
40	2	11.08	9.85	9.15	8.82	8.60	8.51	8.48	8.00	7.95	7.85	7.89	7.90
41	3	8.39	7.09	6.54	6.50	6.40	6.40	6.40	6.40	6.40	6.39	6.42	6.38
42	3	9.32	7.80	7.04	6.82	6.68	6.67	6.67	6.65	6.62	6.62	6.65	6.62
43	3	7.63	6.55	6.22	6.20	6.13	6.12	6.16	6.16	6.14	6.13	6.17	6.12
44	3	10.39	8.79	7.82	7.34	7.06	6.90	6.88	6.85	6.84	6.80	6.83	6.80
45	3	7.81	6.52	6.06	5.97	5.89	5.89	5.95	5.90	5.88	5.88	5.91	5.89
46	3	7.74	6.45	6.04	5.98	5.94	5.91	5.94	5.94	5.92	5.92	5.94	5.92
47	3	8.41	7.31	6.92	6.88	6.84	6.84	6.87	6.85	6.83	6.82	6.87	6.83
48	3	8.90	7.62	6.98	6.82	6.72	6.70	6.73	6.70	6.68	6.68	6.67	6.66
49	3	8.92	7.51	6.84	6.68	6.57	6.54	6.52	6.52	6.50	6.50	6.54	6.51
50	3	7.26	6.49	6.12	6.09	6.01	6.00	6.00	5.98	5.97	5.97	5.95	5.95

Listón	Grosor	Inicio	7	14	21	28	35	42	49	56	62	70	92
51	3	9.23	6.93	6.58	6.56	6.50	6.50	6.50	6.50	6.50	6.50	6.53	6.50
52	3	7.59	6.44	6.00	5.90	5.88	5.88	5.88	5.88	5.86	5.86	5.86	5.86
53	3	7.80	6.46	6.80	6.06	6.00	6.00	6.05	6.00	6.00	5.95	6.01	5.99
54	3	7.86	6.66	6.26	6.20	6.14	6.14	6.19	6.16	6.13	6.13	6.16	6.13
55	3	8.76	7.69	7.20	7.10	7.08	7.03	7.03	7.02	7.00	7.00	7.03	7.00
56	3	9.42	8.32	7.72	7.54	7.45	7.44	7.44	7.39	7.39	7.39	7.42	7.40
57	3	8.00	6.71	6.36	6.32	6.28	6.28	6.28	6.26	6.25	6.25	6.27	6.25
58	3	7.60	6.22	5.80	5.74	5.66	5.66	5.70	5.70	5.67	5.67	5.67	5.66
59	3	7.68	6.10	5.67	5.70	5.66	5.66	5.66	5.66	5.66	5.66	5.68	5.66
60	3	7.89	6.80	6.48	6.46	6.40	6.40	6.39	6.37	6.37	6.37	6.39	6.37
61	4	6.82	5.65	5.31	5.28	5.23	5.23	5.25	5.24	5.23	5.23	5.26	5.22
62	4	6.96	5.69	5.28	5.23	5.18	5.18	5.19	5.17	5.16	5.16	5.19	5.16
63	4	8.54	7.05	6.47	6.30	6.20	6.20	6.20	6.17	6.15	6.15	6.19	6.17
64	4	8.71	7.26	6.57	6.33	6.19	6.18	6.18	6.16	6.13	6.13	6.18	6.14
65	4	8.84	7.22	6.46	6.11	5.92	5.88	5.88	5.86	5.83	5.80	5.83	5.89
66	4	6.72	5.60	5.26	5.24	5.19	5.19	5.20	5.20	5.19	5.19	5.21	5.19
67	4	8.88	7.39	6.53	6.07	5.80	5.65	5.65	5.62	5.58	5.56	5.59	5.58
68	4	7.58	5.91	5.41	5.46	5.40	5.40	5.41	5.41	5.40	5.39	5.43	5.39
69	4	6.64	5.94	5.68	5.66	5.62	5.62	5.63	5.62	5.62	5.62	5.64	5.62
70	4	8.58	6.88	6.00	5.69	5.57	5.56	5.53	5.53	5.52	5.52	5.53	5.50
71	4	7.38	5.79	5.47	5.43	5.39	5.39	5.41	5.40	5.39	5.39	5.42	5.38
72	4	8.56	6.16	5.77	5.74	5.69	5.69	5.75	5.72	5.70	5.70	5.72	5.69
73	4	8.48	6.99	6.26	6.01	5.85	5.79	5.78	5.76	5.74	5.74	5.76	5.74
74	4	8.68	7.22	6.56	6.33	6.19	6.13	6.15	6.12	6.10	6.09	6.12	6.08
75	4	7.61	5.71	5.44	5.40	5.32	5.32	5.36	5.33	5.30	5.29	5.30	5.29
76	4	6.80	5.84	5.67	5.55	5.48	5.47	5.50	5.50	5.49	5.46	5.49	5.47
77	4	7.18	5.75	5.46	5.43	5.36	5.36	5.40	5.39	5.39	5.35	5.39	5.37
78	4	8.74	6.95	6.16	5.83	5.68	5.62	5.62	5.59	5.57	5.57	5.60	5.58
79	4	8.19	7.21	6.61	6.34	6.17	6.12	6.06	6.04	6.02	6.02	6.04	6.02
80	4	9.09	7.59	6.82	6.50	6.30	6.27	6.27	6.26	6.25	6.24	6.25	6.23
81	5	4.92	4.27	4.06	4.04	3.99	3.99	4.02	4.01	4.01	4.01	4.03	4.00
82	5	5.10	4.52	4.24	4.23	4.18	4.18	4.23	4.19	4.19	4.18	4.21	4.18
83	5	5.89	4.46	4.12	4.06	4.02	4.01	4.03	4.02	4.02	4.01	4.03	4.01
84	5	5.82	4.37	4.03	3.96	3.89	3.89	3.89	3.89	3.89	3.89	3.91	3.88
85	5	6.51	4.41	4.20	4.20	4.16	4.16	4.16	4.16	4.16	4.16	4.17	4.16
86	5	7.07	4.52	4.28	4.28	4.24	4.22	4.27	4.24	4.24	4.24	4.27	4.23
87	5	5.79	4.20	3.88	3.86	3.81	3.80	3.82	3.80	3.80	3.78	3.82	3.80
88	5	5.59	4.26	4.08	4.08	4.04	4.04	4.06	4.04	4.04	4.04	4.06	4.03
89	5	7.94	5.69	4.61	4.25	4.14	4.13	4.18	4.15	4.13	4.12	4.15	4.12
90	5	6.02	4.30	3.90	3.85	3.81	3.81	3.84	3.81	3.81	3.81	3.83	3.80
91	5	5.82	4.32	4.02	3.96	3.93	3.93	3.95	3.94	3.92	3.92	3.93	3.91
92	5	5.86	4.30	4.08	4.08	4.04	4.04	4.06	4.06	4.06	4.04	4.05	4.04
93	5	5.79	4.29	4.06	4.06	4.04	4.03	4.03	4.02	4.02	4.01	4.05	4.20
94	5	5.95	4.39	4.00	3.95	3.90	3.90	3.90	3.90	3.90	3.90	3.92	3.90
95	5	5.88	4.74	4.32	4.22	4.16	4.16	4.19	4.18	4.18	4.17	4.17	4.15
96	5	5.20	4.33	4.12	4.10	4.07	4.06	4.09	4.08	4.08	4.07	4.09	4.06
97	5	5.70	4.33	4.08	4.06	4.04	4.04	4.04	4.02	4.02	4.02	4.06	4.04
98	5	5.80	4.38	4.10	4.07	4.03	4.02	4.05	4.02	4.02	4.02	4.04	4.02
99	5	5.74	4.23	3.88	3.83	3.78	3.78	3.82	3.79	3.79	3.79	3.81	3.80
100	5	5.98	5.12	4.68	4.54	4.47	4.44	4.45	4.44	4.44	4.44	4.45	4.45

## A.5. Conteo de pasos en humanos

	Sexo	Estatura	Peso	P1	P2	P3
1	M	1.65	57.00	58.00	54.00	63.00
2	F	1.63	56.00	75.00	74.50	74.00
3	F	1.60	59.00	64.00	70.00	67.00
4	M	1.68	69.00	70.00	71.50	70.00
5	F	1.58	52.00	73.50	72.00	71.50
6	F	1.58	52.00	70.00	71.00	70.00
7	M	1.82	89.00	65.00	69.50	71.00
8	M	1.73	73.00	63.00	61.50	61.00
9	M	1.66	66.00	80.00	82.00	82.00
10	F	1.54	60.00	65.50	68.50	67.50
11	M	1.70	56.00	67.00	66.00	66.50
12	F	1.66	54.00	65.00	64.00	65.00
13	M	1.72	62.00	74.50	75.00	72.50
14	F	1.71	65.00	66.00	64.00	64.00
15	F	1.70	55.00	69.50	67.50	68.00
16	M	1.74	56.00	68.50	68.50	67.00
17	F	1.70	65.00	73.00	72.00	74.00
18	M	1.63	77.00	70.00	71.50	71.50
19	M	1.73	94.00	71.00	64.00	64.00
20	M	1.72	67.00	66.50	65.00	66.00
21	F	1.66	71.00	70.00	69.00	68.00
22	M	1.84	70.00	66.00	59.00	59.00
23	F	1.55	48.00	74.50	75.00	75.00
24	F	1.69	54.00	62.00	61.00	59.50
25	M	1.71	68.00	65.00	64.00	63.00
26	M	1.80	74.00	63.00	60.00	60.50
27	M	1.73	56.00	76.00	76.00	76.00
28	F	1.60	54.00	72.00	70.00	71.00
29	F	1.60	52.00	87.00	87.00	87.00
30	M	1.70	81.00	76.50	76.00	77.00
31	M	1.75	55.00	72.50	71.50	72.50
32	M	1.68	79.00	66.00	64.50	66.00
33	M	1.78	69.00	59.00	59.00	58.00
34	M	1.71	60.00	72.00	71.00	72.00
35	F	1.56	53.00	72.00	68.50	68.50
36	F	1.60	55.00	72.00	71.50	70.50
37	M	1.71	62.00	74.00	71.00	74.00
38	M	1.74	65.00	72.00	68.00	69.00
39	F	1.76	66.00	76.00	75.00	74.00
40	M	1.72	92.00	66.00	66.50	66.00
41	F	1.61	72.00	81.50	83.00	84.00
42	F	1.52	54.00	75.00	74.00	74.00
43	F	1.57	46.00	67.00	69.00	70.00
44	F	1.54	45.00	72.00	71.50	72.00
45	M	1.67	70.00	78.50	78.00	79.00

	Sexo	Estatura	Peso	P1	P2	P3
46	M	1.74	75.00	64.00	63.00	63.00
47	M	1.66	63.00	62.00	63.00	62.50
48	M	1.70	59.00	68.00	67.00	67.00
49	M	1.75	60.00	67.00	68.00	70.50
50	M	1.72	64.00	67.00	70.00	70.00
51	M	1.77	74.00	61.50	61.00	60.00
52	M	1.73	73.00	72.00	72.00	71.00
53	M	1.73	64.00	65.00	66.00	64.00
54	M	1.75	60.00	69.00	69.00	68.50
55	M	1.82	68.00	65.00	66.00	65.00
56	M	1.76	64.00	71.00	72.00	75.00
57	F	1.77	70.00	57.00	56.00	55.00
58	M	1.79	70.00	63.00	64.00	63.00
59	M	1.74	68.00	66.00	68.00	66.50
60	F	1.57	50.00	76.00	75.50	76.00
61	F	1.59	58.00	68.00	68.50	66.50
62	M	1.74	69.00	61.00	60.00	60.00
63	F	1.61	49.00	73.00	73.50	71.50
64	F	1.60	62.00	72.00	72.00	71.00
65	M	1.83	91.00	52.00	53.00	53.00
66	M	1.87	80.00	67.00	69.00	59.00
67	M	1.74	64.00	69.00	66.00	66.00
68	F	1.70	50.00	67.00	69.00	69.50
69	F	1.63	48.00	65.00	63.50	65.00
70	M	1.84	86.00	68.00	67.00	67.00
71	M	1.69	73.00	64.00	65.00	66.00
72	F	1.64	55.00	67.00	68.50	68.00
73	F	1.66	63.00	65.00	64.00	65.50
74	M	1.67	74.00	72.00	77.00	74.00
75	M	1.77	74.00	57.00	56.00	56.00
76	M	1.81	79.00	64.00	60.00	63.00
77	M	1.77	78.00	64.00	62.50	59.50
78	M	1.72	60.00	71.00	71.00	70.00
79	M	1.76	88.00	63.00	63.00	62.00
80	F	1.60	63.00	71.00	71.00	70.00
81	M	1.79	59.00	80.00	77.00	76.50
82	M	1.75	65.00	61.00	62.00	61.00
83	M	1.70	82.00	68.00	67.00	66.00
84	M	1.78	73.00	64.00	66.00	67.00
85	F	1.55	60.00	66.00	70.50	68.50
86	M	1.77	71.00	64.50	63.00	63.00
87	M	1.78	66.00	63.00	64.00	64.00
88	M	1.72	77.00	67.00	68.00	66.00
89	M	1.67	86.00	62.00	63.00	71.00
90	M	1.70	61.00	58.50	58.50	60.00
91	M	1.58	59.00	69.50	71.00	70.00
92	F	1.66	54.00	72.00	73.00	72.00

	Sexo	Estatura	Peso	P1	P2	P3
93	F	1.61	53.00	76.00	75.50	75.00
94	M	1.84	72.00	60.50	59.00	59.00
95	M	1.74	80.00	69.00	69.00	67.00
96	F	1.50	57.00	81.50	82.00	82.00
97	M	1.79	59.00	72.50	70.00	68.00
98	M	1.80	78.00	63.00	63.00	63.00
99	M	1.85	76.00	68.00	69.00	71.00
100	M	1.68	66.00	71.00	73.00	73.00
101	M	1.76	90.00	70.00	69.00	68.00
102	F	1.53	49.00	79.00	78.00	78.00
103	M	1.69	59.00	67.00	70.00	70.00
104	F	1.73	65.00	67.00	69.00	68.00
105	F	1.63	58.00	78.50	77.50	78.50
106	F	1.59	55.00	76.00	75.00	75.00
107	F	1.68	55.00	72.50	73.00	74.00
108	F	1.61	56.00	66.00	68.00	68.00
109	F	1.70	65.00	65.00	65.00	64.00
110	F	1.58	62.00	67.00	64.00	65.00
111	F	1.68	52.00	76.00	78.00	78.00
112	F	1.56	43.00	69.00	72.00	68.50
113	M	1.80	67.00	69.00	69.00	68.00
114	F	1.59	58.00	74.00	77.00	74.00
115	F	1.58	56.00	75.00	75.00	75.00
116	F	1.67	59.00	81.50	79.50	77.00
117	F	1.64	50.00	71.00	72.50	73.00
118	F	1.66	70.00	76.00	75.00	74.00

# Apéndice **B**

## Resultados de álgebra Lineal

### Elementos matriciales

#### Formas cuadráticas

Una forma cuadrática en las variables  $x_1, x_2, \dots, x_d$  es una expresión del tipo

$$f(x_1, x_2, \dots, x_d) = \sum_{i=1}^d \sum_{j=1}^d a_{ij} x_i x_j$$

donde  $a_{ij} = a_{ji}$ . Algunos  $a_{ij}$  pueden ser ceros. En forma matricial la forma cuadrática se escribe como

$$\mathbf{x}' \mathbf{A} \mathbf{x}$$

donde

1.  $\mathbf{x} = (x_1, x_2, \dots, x_d)'$  y
2.  $\mathbf{A}$  es una matriz simétrica de dimensión  $n \times n$  que contiene los coeficientes.

La forma cuadrática más simple es  $f(x) = a_{11}x^2$ . La ecuación de una parábola en la única variable  $x$ .

**Matriz definida positiva** Una matriz simétrica  $\mathbf{A}$  y su forma cuadrática asociada es llamada *definida positiva* si  $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$  para todo vector no nulo  $\mathbf{x}$ .

**Matriz semidefinida positiva** Una matriz simétrica  $\mathbf{A}$  y su forma cuadrática asociada es llamada *semidefinida positiva* si  $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$  para todo vector no nulo  $\mathbf{x}$ .

**Matriz definida negativa** Una matriz simétrica  $\mathbf{A}$  y su forma cuadrática asociada es llamada *definida negativa* si  $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$  para todo vector no nulo  $\mathbf{x}$ .

**Matriz semidefinida negativa** Una matriz simétrica  $\mathbf{A}$  y su forma cuadrática asociada es llamada *semidefinida negativa* si  $\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0$  para todo vector no nulo  $\mathbf{x}$ .

También es posible hablar de formas indefinidas, esto pasa cuando la forma cuadrática es positiva y negativa para conjuntos de vectores

### Resultados útiles

1. Formas cuadráticas positivas tienen matrices de rango completo.
2. Sea  $\mathbf{C}_{m \times n}$  de rango  $r$ . El rango de  $\mathbf{C}\mathbf{C}'$  es  $r$  igual que el de  $\mathbf{C}'\mathbf{C}$
3. Sea  $\mathbf{C}_{m \times n}$  de rango  $r$ .  $\mathbf{C}\mathbf{C}'$  y  $\mathbf{C}'\mathbf{C}$  son definidas positivas o semidefinidas positivas. Si el rango de  $\mathbf{C}\mathbf{C}'$  o  $\mathbf{C}'\mathbf{C}$  es igual a su tamaño entonces la matriz es definida positiva, en otro caso es semidefinida positiva.

## Las raíces y vectores característicos de una matriz

Las *raíces características* (valores propios o eigenvalores) de una matriz  $\mathbf{A}_{p \times p}$  son las soluciones al determinante

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

El determinante es un polinomio de grado  $p$  en  $\lambda$ , y así  $\mathbf{A}$  tiene  $p$  raíces características

$$|\mathbf{A} - \lambda\mathbf{I}| = (-\lambda)^p + S_1(-\lambda)^{p-1} + S_2(-\lambda)^{p-2} + \cdots - S_{p-1}(-\lambda) + |\mathbf{A}|$$

donde  $S_i$  es la suma de todos los determinantes de las menores principales  $i \times i$ .  $S_1$  es la suma de todos los elementos de la diagonal principal de  $\mathbf{A}$ , o sea la *traza*( $\mathbf{A}$ ).

**Resultados:**

- El producto de las raíces características de  $\mathbf{A}$  es igual a  $|\mathbf{A}|$ .
- La suma de las raíces características de  $\mathbf{A}$  es igual a la *traza*( $\mathbf{A}$ ).

Cuando  $\mathbf{A} = \mathbf{S}$ , entonces  $|\mathbf{S}|$  recibe el nombre de la varianza generalizada, y *traza*( $\mathbf{S}$ ) recibe el nombre de variación total.

- Las raíces características de una matriz simétrica real son todas reales.
- Las raíces características de una matriz positiva son todas positivas.
- Si una matriz simétrica de  $n \times n$  es semidefinida positiva de rango  $r$ , contiene exactamente  $r$  raíces positivas y  $n - r$  raíces iguales a cero.
- Las raíces características no ceros del producto  $\mathbf{AB}$  son iguales a las raíces no ceros de  $\mathbf{BA}$ . Luego las trazas de  $\mathbf{AB}$  y  $\mathbf{BA}$  son iguales.
- Las raíces características de una matriz diagonal son los mismos elementos de la diagonal.
- Las raíces características de una matriz triangular son los elementos de la diagonal.

Asociada con cada raíz característica  $\lambda_i$  de la matriz cuadrada  $\mathbf{A}$  hay un vector característico  $\mathbf{x}_i$  cuyos elementos satisfacen el sistema de ecuaciones homogéneo

$$[\mathbf{A} - \lambda_i \mathbf{I}] \mathbf{x}_i = \mathbf{0}$$

## La norma de una matriz

Una función  $\|\mathbf{A}\|$  definida sobre todas las matrices  $m \times m$   $\mathbf{A}$ , reales o complejas, es una norma de matriz si las siguientes condiciones cumplen para todas las matrices  $\mathbf{A}$  y  $\mathbf{B}$ :

1.  $\|\mathbf{A}\| \geq 0$
2.  $\|\mathbf{A}\| = 0$  si, y solo sí,  $\mathbf{A} = \mathbf{0}$
3.  $\|c\mathbf{A}\| = |c| \|\mathbf{A}\|$  para cualquier escalar.
4.  $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$
5.  $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$

La norma matricial euclídeana se define como

$$\|\mathbf{A}\|_E = \left( \sum_{i=1}^m \sum_{j=1}^m |a_{ij}|^2 \right)^{1/2}$$

La norma matricial máximo de columna es

$$\|\mathbf{A}\|_1 = \max_{1 \leq j \leq m} \sum_{i=1}^m |a_{ij}|$$

La norma matricial máximo de fila es

$$\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^m |a_{ij}|$$

## Descomposición espectral de una matriz

Si  $\mathbf{A}_{d \times d}$  es una matriz simétrica entonces  $\mathbf{A} = \mathbf{T}\mathbf{\Delta}\mathbf{T}'$ , donde

1.  $\mathbf{T}$  es ortogonal
2.  $\mathbf{\Delta} = \text{Diag}[\lambda_1, \lambda_2, \dots, \lambda_d]$  es una matriz diagonal real donde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$  y corresponden a las raíces propias de  $\mathbf{A}$ .
3. Las columnas de  $\mathbf{T}$  son los vectores propios de  $\mathbf{A}$ .

**Resultados:**

1.  $tr(\mathbf{A}) = \sum_{i=1}^d \lambda_i$
2.  $Det(\mathbf{A}) = |\mathbf{A}| = \prod_{i=1}^d \lambda_i$
3. Si  $\mathbf{A}$  es una matriz de proyección ( $\mathbf{A}\mathbf{A} = \mathbf{A}$ ), entonces
  - a)  $\lambda_i \in \{0, 1\}$ , para  $i = 1, 2, \dots, d$
  - b)  $tr(\mathbf{A}) = rango(\mathbf{A})$
  - c)  $tr(\mathbf{A})$  es la dimensión del subespacio sobre el cual  $\mathbf{A}$  proyecta.
4. Si  $\mathbf{A} \geq 0$  entonces  $\lambda_d \geq 0$
5. Si  $\mathbf{A} > 0$  entonces  $\lambda_d > 0$

Estas propiedades de matrices simétricas tienen una implicación importante para las formas cuadráticas. Si se aplica la transformación ortogonal

$$\mathbf{x} = \mathbf{P}\mathbf{y}$$

a las  $p$  variables de la forma cuadrática  $\mathbf{x}'\mathbf{A}\mathbf{x}$  tenemos

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{y} = \mathbf{y}'\mathbf{D}\mathbf{y} = \lambda_1 y_1^2 + \dots + \lambda_r y_r^2$$

donde las  $\lambda_i$  son las raíces características de la matriz de coeficientes y  $r$  es el rango de la forma.

### Traza de una matriz

- $Tr(\mathbf{A} + \mathbf{B}) = Tr(\mathbf{A}) + Tr(\mathbf{B})$
- $Tr(\mathbf{AC}) = Tr(\mathbf{CA})$
- Si  $\mathbf{A}$  es una matriz simétrica de dimensión  $n \times n$  con valores propios  $\lambda_i, i = 1, \dots, n$  entonces:

1.

$$tr(\mathbf{A}) = \sum_{i=1}^n \lambda_i$$

2.

$$tr(\mathbf{A}^s) = \sum_{i=1}^n \lambda_i^s$$

3.

$$\operatorname{tr}(\mathbf{A}^{-1}) = \sum_{i=1}^n \lambda_i^{-1}$$

si  $\mathbf{A}$  es no singular.

- Suponga que  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k$  representan matrices de dimensión  $n \times n$  y  $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_k$ . Si  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k$  son idempotentes y si  $\mathbf{A}_i \mathbf{A}_j = \mathbf{0}$  para todo  $i \neq j$ , entonces  $\mathbf{A}$  es idempotente con  $\operatorname{rango}(\mathbf{A}) = \operatorname{rango}(\mathbf{A}_1) + \dots + \operatorname{rango}(\mathbf{A}_k)$
- En particular  $\operatorname{rango}(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}') = n - 1$

## El rango

- Si  $\mathbf{A}$  y  $\mathbf{B}$  son matrices conformables, entonces

$$\operatorname{rango}(\mathbf{AB}) \leq \min\{\operatorname{rango}(\mathbf{A}), \operatorname{rango}(\mathbf{B})\}$$

- Si  $\mathbf{A}$  es cualquier matriz y  $\mathbf{P}$  y  $\mathbf{Q}$  son matrices no singulares conformables, entonces

$$\operatorname{rango}(\mathbf{PAQ}) = \operatorname{rango}(\mathbf{A})$$

## Matrices particionadas

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \cdots & \mathbf{A}_{mn} \end{bmatrix}$$

donde  $\mathbf{A}_{ij}$  es una submatriz que contiene  $r_i$  filas y  $c_j$  columnas.

### Operaciones con matrices particionadas

#### Suma

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \cdots & \mathbf{A}_{1n} + \mathbf{B}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} + \mathbf{B}_{m1} & \cdots & \mathbf{A}_{mn} + \mathbf{B}_{mn} \end{bmatrix}$$

donde  $\mathbf{A}_{ij}$  y  $\mathbf{B}_{ij}$  tienen la misma dimensión para todo  $i$  y  $j$ .

#### Producto

$$\mathbf{AB} = \begin{bmatrix} \sum_{j=1}^n \mathbf{A}_{1j}\mathbf{B}_{j1} & \cdots & \sum_{j=1}^n \mathbf{A}_{1j}\mathbf{B}_{jn} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^n \mathbf{A}_{mj}\mathbf{B}_{j1} & \cdots & \sum_{j=1}^n \mathbf{A}_{mj}\mathbf{B}_{jn} \end{bmatrix}$$

#### Inversa

$$\mathbf{A}^{-1} = \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & -(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix}$$

donde  $\mathbf{A}_{11}$  y  $\mathbf{A}_{22}$  son ambas cuadradas y no singulares

#### Resultado

Si  $\mathbf{A}_{11}$  es no singular

$$|\mathbf{A}| = |\mathbf{A}_{11}| \cdot |\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|$$

Si  $\mathbf{A}_{22}$  es no singular

$$|\mathbf{A}| = |\mathbf{A}_{22}| \cdot |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|$$

## Inversa generalizada

Sea  $\mathbf{A}$  una matriz. Existe una única matriz  $\mathbf{K}$  que satisface las siguientes condiciones:

1.  $\mathbf{AKA} = \mathbf{A}$
2.  $\mathbf{KAK} = \mathbf{K}$
3.  $\mathbf{KAK} = \mathbf{K}$

Condiciones Satisfechas	Nombre de la matriz	Símbolo que se puede usar
1	Inversa generalizada	$A^{-}$ ó $A^{\{g\}}$
1 y 2	Inversa generalizada reflexiva	$A^{\{r\}}$
1, 2 y 3	Inversa generalizada normalizada	$A^{\{n\}}$
1, 2, 3 y 4	Inversa generalizada Penrose	$A^{\{p\}}$

## Diferenciación de formas lineales y cuadráticas

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}'} = \mathbf{a}'$$

$$\frac{\partial \mathbf{B}\mathbf{x}}{\partial \mathbf{x}'} = \mathbf{B}$$

$$\frac{\partial (\mathbf{B}\mathbf{x})'}{\partial \mathbf{x}} = \mathbf{B}'$$

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$$

Si  $\mathbf{A}$  es simétrica

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

## Operadores esperanza y covarianza

Sea  $\mathbf{X} = [(x_{ij})]$  una matriz de variables aleatorias. Se define el operador esperanza  $\mathcal{E}$  por la siguiente matriz

$$\mathcal{E}[\mathbf{X}] = [(E[X_{ij}])]$$

o sea una matriz con los valores esperados de las variables aleatorias. De la propiedad de linealidad de  $E$ , se tiene

$$\mathcal{E}[\mathbf{AXB} + \mathbf{C}] = \mathbf{A}\mathcal{E}[\mathbf{X}]\mathbf{B} + \mathbf{C}$$

**Prueba:**

Sea

$$\mathbf{W} = \mathbf{AXB} + \mathbf{C}$$

Entonces

$$W_{ij} = \sum_{r=1}^m \sum_{s=1}^n a_{ir} X_{rs} b_{sj} + c_{ij}$$

Entonces

$$\begin{aligned} \mathcal{E}[\mathbf{AXB} + \mathbf{C}] &= [E(W_{ij})] \\ &= \left[ \sum_{r=1}^m \sum_{s=1}^n a_{ir} E[X_{rs}] b_{sj} + c_{ij} \right] \\ &= \left[ \left( (\mathbf{A}\mathcal{E}[\mathbf{X}]\mathbf{B})_{ij} \right) \right] + [(c_{ij})] \\ &= \mathbf{A}\mathcal{E}[\mathbf{X}]\mathbf{B} + \mathbf{C} \end{aligned}$$

**Resultado**

Si  $\mathbf{x}$  es un vector de  $m \times 1$  de variables aleatorias (vector aleatorio), entonces

$$\mathcal{E}[\mathbf{Ax}] = \mathbf{A}\mathcal{E}[\mathbf{x}]$$

para cualquier matriz conformable  $\mathbf{A}$ .

**Resultado**

Si  $\mathbf{x}$  y  $\mathbf{y}$  son vectores aleatorios de  $n \times 1$  de  $\mathbf{A}$  y  $\mathbf{B}$  son matrices de constantes de dimensiones  $m \times n$ , entonces

$$\mathcal{E}[\mathbf{Ax} + \mathbf{By}] = \mathbf{A}\mathcal{E}[\mathbf{x}] + \mathbf{B}\mathcal{E}[\mathbf{y}]$$

**Resultado**

Si  $\mathbf{x}$  y  $\mathbf{y}$  son vectores aleatorios de  $n \times 1$  de  $a$  y  $b$  son escalares, entonces

$$\mathcal{E}[a\mathbf{x} + b\mathbf{y}] = a\mathcal{E}[\mathbf{x}] + b\mathcal{E}[\mathbf{y}]$$

**Covarianza generalizada**

El concepto de covarianza también puede generalizarse para dos vectores de variables aleatorias  $\mathbf{x} = [(x_{ij})]$  y  $\mathbf{y} = [(y_{ij})]$  (no necesariamente de las mismas dimensiones), este operador generalizado se denota por  $\mathcal{C}$  y se define como:

$$\begin{aligned} \mathcal{C}[\mathbf{x}, \mathbf{y}] &= [(cov[x_i, y_j])] \\ &= \mathcal{E}[(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{y} - \mu_{\mathbf{y}})'] \\ &= \mathcal{E}[\mathbf{xy}'] - \mu_{\mathbf{x}}\mu_{\mathbf{y}}' \end{aligned}$$

donde  $\mu_{\mathbf{x}} = \mathcal{E}[\mathbf{x}]$ . Si  $\mathbf{x}$  y  $\mathbf{y}$  son independientes estadísticamente, entonces  $\mathcal{C}[\mathbf{x}, \mathbf{y}] = \mathbf{0}$ .

Cuando  $\mathbf{x} = \mathbf{y}$ , entonces se puede denotar  $\mathcal{C}[\mathbf{x}, \mathbf{x}] = \mathcal{D}(\mathbf{x}) = \Sigma_x$ .

$$\Sigma_x = \mathcal{E}[(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{x} - \mu_{\mathbf{x}})']$$

**Resultados**

1.  $\mathcal{C}[\mathbf{Ax}, \mathbf{By}] = \mathbf{A}\mathcal{C}[\mathbf{x}, \mathbf{y}]\mathbf{B}'$
2.  $\mathcal{D}[\mathbf{Ax}] = \mathbf{A}\mathcal{D}[\mathbf{x}]\mathbf{A}'$
3.  $\mathcal{D}[\mathbf{x} - \mathbf{a}] = \mathcal{D}[\mathbf{x}]$

4. Si  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , es una muestra aleatoria de una distribución  $d$ -dimensional con vector de medias  $\boldsymbol{\mu}$  y matriz de dispersión (o de varianzas y covarianzas)  $\boldsymbol{\Sigma}$ , entonces, si definimos la media muestral y la dispersión muestral como

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = \frac{1}{n-1} \mathbf{Q}$$

Entonces

$$\mathcal{E}[\bar{\mathbf{x}}] = \frac{1}{n} \sum_{i=1}^n \mathcal{E}[\mathbf{x}_i] = \boldsymbol{\mu}$$

$$\mathcal{D}[\bar{\mathbf{x}}] = \frac{1}{n^2} \sum_{i=1}^n \mathcal{D}[\mathbf{x}_i] = \frac{1}{n} \boldsymbol{\Sigma}$$

## Valor esperado de una forma cuadrática

### Resultado

Sea  $\mathbf{y}_{n \times 1}$  un vector aleatorio y sea  $\mathcal{E}(\mathbf{y}) = \boldsymbol{\mu}$  y  $Cov(\mathbf{y}) = \boldsymbol{\Sigma}$ . Entonces

$$\mathcal{E}(\mathbf{y}' \mathbf{A} \mathbf{y}) = \text{traza}(\mathbf{A} \boldsymbol{\Sigma}) + \boldsymbol{\mu}' \mathbf{A} \boldsymbol{\mu}$$

**prueba:**

$$E(y_i y_j) = \sigma_{ij} + \mu_i \mu_j$$

$$\begin{aligned}
\mathcal{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) &= \mathcal{E}\left[\sum_{i=1}^n \sum_{j=1}^n y_i y_j a_{ij}\right] \\
&= \sum_{i=1}^n \sum_{j=1}^n a_{ij} (\sigma_{ij} + \mu_i \mu_j) \\
&= \text{traza}(\mathbf{A}\Sigma) + \mu' \mathbf{A} \mu
\end{aligned}$$

**Resultado**

Sea  $\mathbf{y}_{n \times 1}$  un vector aleatorio y sea  $\mathcal{E}(\mathbf{y}) = \mathbf{0}$  y  $Cov(\mathbf{y}) = \Sigma = \sigma^2 \mathbf{I}$  y  $\mathbf{A}$  idempotente de rango  $k$ . Entonces

$$\mathcal{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) = k\sigma^2$$

**Resultado**

$$(\mathbf{A} + \mathbf{BGB}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B} + \mathbf{G}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1}$$

# Apéndice C

## Repaso de modelos lineales generalizados

### C.1. Familia exponencial

Uno de los objetivos de este tipo de modelamiento consiste en construir un modelo lineal el cual es válido para un miembro de la familia exponencial. Recuerde que si  $y$  es un miembro de la familia exponencial, entonces su densidad se puede expresar como:

$$f(y; \theta, \phi) = e^{\left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}}$$

donde:

$\theta$  = Un parámetro de localización (no necesariamente la media de  $y$ )

y

$\phi$  = Un parámetro de dispersión

( $\phi$  solo aparece en distribuciones que tienen 2 parámetros, tal como la normal). Entonces, si la variable tiene una distribución que depende solo de un parámetro, entonces:

$$f(y; \theta) = e^{\left\{ \frac{y\theta - b(\theta)}{a} + c(y) \right\}}$$

La siguiente tabla muestra algunas distribuciones comunes de la familia exponencial junto con sus respectivas componentes:

Distribución	$a$	$b(\theta)$	$c(y)$
Bernoulli	1	$\log(1 + \exp(\theta))$	$\log(1) = 0$
Binomial	$\frac{1}{n}$	$\log(1 + \exp(\theta))$	$\log \binom{n}{ny}$
Poisson	1	$\exp(\theta)$	$-\log(y!)$
Poisson de media $\mu t$	$\frac{1}{t}$	$\exp(\theta)$	$yt \log(yt) - \log(yt!)$

**Tabla c.1.** Algunas distribuciones comunes de la familia exponencial

Puede mostrarse que la media y varianza de  $y$  están especificadas como:

$$E(y) = \mu = b'(\theta)$$

y

$$V(y) = ab''(\theta) = ab''(b'^{-1}(\mu))$$

Distribución	Media	Varianza
Bernoulli	$(1 + \exp(-\theta))^{-1}$	$\frac{\exp(\theta)}{[1 + \exp(\theta)]^2} = \mu(1 - \mu)$
Binomial	$(1 + \exp(-\theta))^{-1}$	$\exp(\theta) / [1 + \exp(\theta)]^2 / n = \mu(1 - \mu) / n$
Poisson	$\exp(\theta)$	$\exp(\theta) = \mu$
Poisson de media $\mu t$	$\exp(\theta)$	$\exp(\theta) / t = \mu / t$

**Tabla c.2.** Media y varianza para algunas distribuciones comunes de la familia exponencial

Recuerde que para respuestas que se distribuyen normalmente, el *modelo lineal general* es:

$$Y = \mu + \epsilon$$

donde

$$\mu = \mathbf{X}\boldsymbol{\beta}$$

Por ejemplo. Si  $Y \sim \text{Bernoulli}$

$$\begin{aligned} P(Y = j) &= p^j (1 - p)^{1-j}, \quad j = 0, 1 \\ &= \left( \frac{p}{1 - p} \right)^j (1 - p) \\ &= e^{\{j\theta + b(\theta)\}} \end{aligned}$$

$c(y) = 0$ . Sean

$$\theta = \log \left( \frac{p}{1 - p} \right), \quad b(\theta) = \log(1 - p)$$

Entonces

$$p = \frac{e^\theta}{1 + e^\theta}$$

y

$$\begin{aligned} b(\theta) &= \log(1 - p) \\ &= \log\left(1 - e^{-\{j\theta + b(\theta)\}}\right) \\ &= \log\left(\frac{1}{1 + e^\theta}\right) \end{aligned}$$

Entonces

$$\begin{aligned} b'(\theta) &= e^{-\{j\theta + b(\theta)\}} \\ &= p \end{aligned}$$

( $E(y) = \mu = p$ ). En términos de inferencia, es mejor pensar en  $\theta$  que en  $p$ .

## C.2. Modelos lineal general y lineal generalizado

En un *modelo lineal generalizado*, se reemplaza a  $\mu$  por una función  $g(\cdot)$  (llamada *Función de vínculo*):

$$g(\mu) = \mathbf{X}\boldsymbol{\beta}$$

El papel de la función de vínculo es mapear el rango de  $\mu$  en la recta real. Un tipo de función de vínculo es la *Función de vínculo canónica* la cual está dada por  $g = b^{-1}$ . A continuación se presentan algunas distribuciones con su respectiva función de vínculo:

Distribución	$g(\mu)$	Nombre
Bernoulli	$\log\left(\frac{\mu}{1-\mu}\right)$	Logit
Binomial	$\log\left(\frac{\mu}{1-\mu}\right)$	Logit
Poisson	$\log(\mu)$	Log
Poisson de media $\mu t$	$\log(\mu)$	Log

**Tabla c.3.** Función de vínculo (o link) para algunos miembros de la familia exponencial

Recuerde que  $\phi$  es un escalar. En el modelo lineal general

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Donde  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  y  $V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$ . Mientras que en el modelo lineal generalizado

$$y = \mu + \epsilon$$

donde  $E(\epsilon) = 0$ . Note que

$$V(y) = ag'^{-1}(\mu)$$

También note

$$V(\mathbf{Y}) = V(\boldsymbol{\epsilon}) = \mathbf{V} \quad (\text{Matriz diagonal})$$

A diferencia del modelo lineal general para respuestas normalmente distribuidas, en el modelo lineal generalizado, la matriz de covarianzas para  $\mathbf{Y}$  tiene diferentes varianzas por observación.

$$V(\boldsymbol{\epsilon}) = \phi\mathbf{A}\mathbf{B}, \quad g(\mu) = \mathbf{X}\boldsymbol{\beta}$$

(Una matriz diagonal). La independencia se preserva pero la distribución no se preserva, por lo que no se preserva la varianza constante como en el

modelo lineal general. En situaciones prácticas, se acostumbra tomar:

$$\begin{aligned}\phi &= 1 \\ \mathbf{A} &= \mathbf{I} \\ \mathbf{B} &= \text{Una función de } \mu\end{aligned}$$

En este caso

$$\mathbf{V} = \mathbf{AB}$$

Donde

$$\mathbf{A} = \{a_i\}$$

y

$$\mathbf{B} = \text{diag}\{b''(\theta_i)\} = \text{diag}\{g'^{-1}(\mu_i)\}$$

Por ejemplo, para el caso Bernoulli y Poisson  $a_i = 1$  lo cual implica que  $\mathbf{A} = \mathbf{I}$ .

### 3.3. Funciones de vínculo Logit

En los casos Bernoulli<sup>1</sup> y Binomial, la función de vínculo se acostumbra llamarla *Regresión logística*. Por ejemplo, suponga que se está llevando a cabo un ensayo clínico con una respuesta dicótoma (éxito-fracaso) y se desean

---

<sup>1</sup>Note que en el caso Bernoulli  $\mu = \frac{e^\theta}{1+e^\theta}$  lo cual implica que

$$\theta = \log\left(\frac{\mu}{1-\mu}\right)$$

y por lo tanto  $\mu = P(y_i = 1)$ .

comparar las tasas de éxito para los dos tratamientos. Entonces

$$\log \left[ \frac{\mu}{1 - \mu} \right] = a + bx$$

donde  $x$  es 0 o 1, dependiendo de si el paciente está en uno u otro grupo de tratamiento. Observe que

$$e^b = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}$$

donde  $p_1$  y  $p_2$  son las probabilidades de éxito asociadas a los dos tratamientos. Esta cantidad  $e^b$  se conoce como *Razón de Odds* (Odds Ratio, OR en inglés)

### 3.4. Funciones de vínculo Log

**Caso Poisson.** Los modelos que usan la función de vínculo 'Log' se conocen como *modelos log-lineales*. Los resultados producidos por estos modelos se interpretan frecuentemente como *tasas relativas* o *riesgos relativos* (Relative Risk, RR en inglés). Por ejemplo, suponga que en un ensayo clínico, el resultado de interés es el número de episodios de mareo durante un período de tiempo fijo. En esta instancia,

$$RR = \frac{\text{Tasa de mareo para tratamiento 1}}{\text{Tasa de mareo para tratamiento 2}}$$

### 3.5. Ajuste del modelo lineal generalizado

Si

$$f(y; \theta) = e^{\left[ \frac{y\theta - b(\theta)}{a} + c(y) \right]}$$

entonces el logaritmo de la función de verosimilitud está dado por:

$$\begin{aligned}\log(L) &= \sum_i \left( \frac{y_i \theta_i - b(\theta_i)}{a_i} + K \right) \\ &= \mathbf{Y}' \mathbf{A}^{-1} \boldsymbol{\theta} - b(\boldsymbol{\theta})^{\frac{1}{2}'} \mathbf{A}^{-1} b(\boldsymbol{\theta})^{\frac{1}{2}} + K\end{aligned}$$

donde  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)'$  y  $b(\boldsymbol{\theta}) = (b(\theta_1), b(\theta_2), \dots, b(\theta_n))'$  y  $K = \text{constante}$ . Esto puede re-escribirse en términos de las covariables y los coeficientes  $\boldsymbol{\beta}$  como:

$$\log(L) = \mathbf{Y}' \mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta} - b(\mathbf{X} \boldsymbol{\beta})^{\frac{1}{2}'} \mathbf{A}^{-1} b(\mathbf{X} \boldsymbol{\beta})^{\frac{1}{2}} + K$$

Diferenciando con respecto a  $\boldsymbol{\beta}$ , se encuentra que el MLE de  $\boldsymbol{\beta}$  debe satisfacer

$$\frac{d \log(L)}{d \boldsymbol{\beta}} = \mathbf{X}' \mathbf{A}^{-1} \left( \mathbf{Y} - b'(\mathbf{X} \boldsymbol{\beta}) \right)$$

Esta ecuación es no lineal en  $\boldsymbol{\beta}$  y por lo tanto se necesita un método numérico iterativo para obtener a  $\hat{\boldsymbol{\beta}}$ . También,

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{B} \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} \mathbf{B})^{-1}$$

donde

$$\mathbf{B} = \text{diag}\{b''(\boldsymbol{\theta})\}$$

la cual es una matriz diagonal que contiene las varianzas de las observaciones individuales.

Las pruebas estadísticas para los efectos fijos se realizan usando pruebas tipo Wald o chi-cuadrado los cuales usualmente requieren de muestras grandes.

**Solución alternativa:** Use el método de mínimos cuadrados ponderados (la idea es 'voltear' el problema de uno no lineal a uno lineal). Para esto, escriba

$$\mathbf{Z} = g(\boldsymbol{\mu}) + (\mathbf{Y} - \boldsymbol{\mu})g'(\boldsymbol{\mu}) = g(\boldsymbol{\mu}) + (\mathbf{Y} - \boldsymbol{\mu})\mathbf{B}^{-1}$$

Esta es una expansión de Taylor de primer orden para  $g(\boldsymbol{\mu})$  alrededor de  $\boldsymbol{\mu}$ . Puesto que  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$  se sigue que:

$$\begin{aligned} V(\mathbf{Z}) &= V(\mathbf{X}\boldsymbol{\beta}) + \mathbf{B}^{-1}\mathbf{V}(\mathbf{Y} - \boldsymbol{\mu})\mathbf{B}^{-1} \\ &= \mathbf{A}\mathbf{B}^{-1} \end{aligned}$$

Entonces, tome

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \epsilon$$

(los términos  $\epsilon$  son independientes pero no idénticamente distribuidos). En este caso  $\mathbf{Z}$  se analiza iterativamente usando mínimos cuadrados ponderados cuyos pesos (o ponderaciones) están dadas por la inversa de la matriz  $V(\mathbf{Z})$ . La iteración se requiere ya que  $\mathbf{Z}$  y  $V(\mathbf{Z})$  dependen de  $\boldsymbol{\beta}$ . Se obtiene

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'V(\mathbf{Z})^{-1}\mathbf{X})^{-1}\mathbf{X}'V(\mathbf{Z})^{-1}\mathbf{Z}$$

De manera continua se actualizan  $V(\mathbf{Z})$  y  $\tilde{\boldsymbol{\beta}}$  hasta que se obtiene convergencia. Al final se calcula:

$$V(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'V(\mathbf{Z})^{-1}\mathbf{X})^{-1}$$

Este método requiere tamaños muestrales grandes.

# Capítulo 4

## Software para modelos mixtos

### 4.1. Software R

El software R<sup>1</sup> es tal vez uno de los mejores para realizar análisis e investigación y tiene la ventaja de ser de dominio público. Entre sus muchas posibilidades, cuenta con librerías especializadas para implementar modelos mixtos lineales y no lineales. En *R* existen varias librerías que permiten ajustar modelos mixtos de una manera eficiente. Se presentan algunas de las más usadas.

### 4.2. Función `lme()`

Esta función se encuentra en la librería `lme`. Permite ajustar un modelo lineal mixto.

#### Utilización:

---

<sup>1</sup>[www.r-project.org](http://www.r-project.org)

```
lme(fixed, data, random, correlation, weights, subset, method,  
    na.action, control, contrasts = NULL, keep.data = TRUE)  
## S3 method for class 'lme':  
update(object, fixed., ..., evaluate = TRUE)
```

### 4.3. Función `glmmML`

Esta función se encuentra en la librería `glmmML`. Permite ajustar un modelo lineal generalizado con intercepto aleatorio vía máxima verosimilitud e integración numérica vía cuadratura de Gauss-Hermite.

#### Utilización:

```
glmmML(formula, family = binomial, data, cluster, weights,  
cluster.weights, subset, na.action,  
offset, prior = c("gaussian", "logistic", "cauchy"),  
start.coef = NULL, start.sigma = NULL, fix.sigma = FALSE,  
control = list(epsilon = 1e-08, maxit = 200, trace = FALSE),  
method = c("Laplace", "ghq"), n.points = 8, boot = 0)
```

### 4.4. Función `glmm()`

Esta función se encuentra en la librería `glmm`. Permite ajustar un Modelo Lineal Mixto Generalizado GLMM utilizando el muestreador de Gibbs.

#### Utilización:

```
glmm(formula, family, data, weights, offset, icm = 50,
```

```
burnin = 1000, keep = 1000, model.show = FALSE,
progress.info = 0, store.results = FALSE, thin = 1,
bugsfile, seed1 = 6762, seed2 = 92928, seed3 = 19729)
```

## 4.5. Función lmer

### Utilización:

Esta función se encuentra en la librería `lmer`.

```
lmer(formula, data, family, method, control, start,
      subset, weights, na.action, offset, contrasts,
      model, ...)
```

`\section{lmer2}`

Esta función se encuentra en la librería `{\tt lmer}`.

`{\bf Utilización:}`

```
lmer2(formula, data, family, method, control, start,
       subset, weights, na.action, offset, contrasts,
       model, ...)
```

**formula** Una fórmula lineal con dos partes que describen la parte de los efectos fijos del modelo, con la respuesta en el lado izquierdo del operador “~” y los términos, separados por operadores “+”, a la derecha. El símbolo “|” separa una expresión para la matriz del modelo y el factor de agrupamiento.

**data** Un marco de datos opcional que contiene las variables nombradas en la fórmula. Por defecto las variables son tomadas del ambiente en el cual fue llamada `lmer`.

**family** Una familia GLM. Si este argumento no se proporciona entonces se ajusta un modelo lineal mixto; en otro caso se ajusta un modelo lineal generalizado mixto.

**method** Un nombre. Para un modelo lineal mixto por defecto es “REML” que indica que el modelo debe ajustarse maximizando la log-verosimilitud restringida. La alternativa es “ML” lo que indica que la log-verosimilitud debe ser maximizada. (este método es llamado algunas veces de máxima verosimilitud “completa”). Para un GLMM el criterio es siempre la log-verosimilitud pero este criterio no tiene una forma cerrada y debe ser aproximado. La aproximación por defecto es “PQL” o quasiverosimilitud penalizada. Alternativas son “Laplace” o “AGQ” que indican aproximaciones laplacianas o cuadraturas gaussianas adaptables, respectivamente. El método “PQL” es el más rápido pero el menos exacto. El método “Laplace” es intermedio en velocidad y exactitud. El método “AGQ” es el más exacto pero es considerablemente más lento que los otros.

**control** Una lista de parámetros de control.

**start** Una lista de matrices de precisión relativas para los efectos aleatorios.

**subset, weights, na.action, offset, contrasts** Argumentos de especificación adicionales.

**model** Argumento lógico que indica si la componente modelo de ser retornado.

... Otros posibles argumentos. En la actualidad no está en uso.

## 4.6. Otros paquetes para modelos lineales mixtos

En algunos capítulos anteriores se discutió en detalle el uso de algunos procedimientos del SAS tales como el PROC MIXED y el PROC GLIMMIX. En relación a otro software relacionado con el uso de e implementación de modelos lineales mixtos, Bolker *et al.* (2008), dan recomendaciones acerca de cuál software usar y en que situaciones. Estos autores se enfocan especialmente en problemas relacionados con modelos lineales mixtos generalizados. Ellos discuten en detalle las distintas técnicas de estimación de parámetros dentro de este modelo disponibles en SAS, R, GenStat, AS-REML, AD Model Builder, HLM, GLLAAMM del STATA y WinBUGS y presentan ventajas y desventajas asociadas a cada uno de los distintos métodos y cada uno de estos paquetes. Concluyen que este tipo de modelos se deberían usar con mayor frecuencia en biología ya que ellos constituyen poderosas herramientas de análisis; sin embargo, invitan a los investigadores del área de biología a usarlos con precaución teniendo en cuenta las recomendaciones anotadas en su artículo.

# Capítulo 5

## Referencias

- Andrews, D. F. and Pregibon, D. (1978). Finding outliers that matter. *Journal of the Royal Statistical Society, Series B*, 40, 85–93.
- Agresti, A. (2002). *Categorical data analysis*. (2nd edn.). New York: Wiley Wiley. U.S.A.
- Aitkin, M.A. Alfó, M. (1998). Regression models for binary longitudinal responses. *Statistics and Computing*, 8, 289–307.
- Aitkin, M.A. Alfó, M. (2003). Longitudinal analysis of repeated binary data using autoregressive and random effect modeling. *Statistical Modeling* 3, 291–303.
- Bates, D.M., Watts, D.G. (1988). *Nonlinear regression analysis and its applications*. John Wiley & Sons, Inc. New York.
- Bolker BM, Brooks, ME, Clark, CJ, Geange, SW, Poulsen, JR, Stevens, HH, and White, JS. (2008). Generalized linear mixed models: a

practical guide for ecology and evolution. *Trends in Ecology and Evolution*, Vol 24, No. 3. pp. 127–135.

- Booth, J. G. y Hobert, J. P. (1998). Standard errors of prediction in generalized linear mixed models *Journal of the American Statistical Association*, Vol. 93, No. 441, pp. 262–272.
- Botero, J. G. (1993). Secado del ciprés (*Cupressus lusitanica* Mill) para usos industriales: estibas, Molduras y Muebles. Tesis Universidad Nacional de Colombia, Medellín.
- Bulirsch, R. and Stoer, J. (1991). *Introduction to numerical analysis*. New York: Springer-Verlag.
- Cantoni, E. and Hastie, T. (2002). Degrees of freedom tests for smoothing splines. *Biometrika*, 89 (2): 251–263.
- Chen, Y. J. et al. (1999). A rank test for equality of two multivariate populations vs. a particular ordered alternative. *Computational Statistics & Data Analysis*, Vol. 29, No. 2, pp. 129–144.
- Cook, R.D.S., and Weisberg, S. (1982). *Residuals and influence in regression*. Chapman and Hall, London.
- Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994). *Analysis of longitudinal data*. Oxford: Oxford University Press.
- Dempster, AP, Laird, NM, and Rubin, DB. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the*

*Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 1–38.

- Dempster, AP, Rubin, DB., and Tsutakawa, RK. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, Vol. 76, No. 374, pp. 341–353.
- Giesbrecht, F.G., Burns, J.C. (1985). Two-stage analysis based on a mixed model: large sample asymptotic theory and small sample simulation results. *Biometrics*, **41**, 477–486.
- Grizzle, J.E., and Allen, D.M. (1969). Analysis of growth and dose response curves. *Biometrics*, vol. 25, no. 2: 357–381.
- Hemmerle, W. J. and Hartley, H.A. (.1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. *Technometrics*, 15, 819-832.
- Harville, D.A. (1974). Optimal procedures for some constrained selection problems. *J. Amer. Statist. Assoc.* **69**, No. 346, pp. 449–452.
- Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of American Statistics Association*, 72:320–340
- Harville, D.A. Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model, *J. Amer. Statist. Assoc.* **87**, pp. 724–731.

- Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Chapman and Hall, London.
- Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression models for multilevel analysis. *Biometrics* 50, 933–944.
- Hedeker, D. Siddiqui, O. and Hu, F.B. (2000). Random effects regression analysis of correlated grouped-time survival data. *Statistical Methods in Medical Research* 9,161-179.
- C. R. Henderson, C. R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, Vol. 31, No. 2 , pp. 423–447
- Jacquim-Gadda et al. (2007). Robustness of linear mixed model to misspecified error distribution. *Computational Statistics & Data Analysis*, Vol. 51, No. 10, pp. 5142–5154.
- Jennrich, RI, Schluchter, MD. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, Vol. 42, No. 4, pp. 805–820.
- Kalbfleisch, J. D. Lawless, J. F. and Vollmer, W. M. (1983). Estimation in Markov models from aggregate data. *Biometrics* 39, 907–919.
- Kendall E. Atkinson. (1989). *An introduction to numerical analysis*. John Wiley & Son.
- Kenward, M. G., Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood *Biometrics* 53, pp.

983–997.

- Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics*, Vol. 38, 963–974.
- Lange, N. y Ryan, L. (1989). Assessing normality in random effects models. *The Annals of Statistics*, Vol. 17, No. 2, pp. 624–642
- Lindstrom, M.J. and Bates, D.M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83, 1014–1022.
- Lindley, V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B*, 34 1–41.
- Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed effects models for repeated-measures data. *Biometrics*, Vol. 46, No. 3, pp. 673–687
- Littell, R. et al. (2006). *SAS for mixed models, Second Edition*. SAS Institute Inc., Cary, NC.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models* (2nd edn). New York: Chapman & Hall/CRC. U.S.A.
- McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of American Statistical Associations* 92, 162–170.
- McGilchrist, C.E. (1994). Estimation in Generalized Mixed Models. *Journal of the Royal Statistical Society B*, 56, 61–69.

- Morrell, C.H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics*, **54**, 1560–1568.
- Nelder, John A.; R. Mead (1965). A simplex method for function minimization. *Computer Journal* 7: 308–313
- Nelson, DL., Cox, MM. (2000). *Lehninger Principles of Biochemistry*. 3rd Ed., Worth Publishers, USA
- Patterson, H.D.; Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58 (3): 545–554.
- Pinheiro, J.C., and Bates, D.M. (2000). *Mixed-effects models in S and S-PLUS*. Springer. New York.
- Prasad, N.G.N. and J. N. K. Rao (1990). The Estimation of mean squared errors of small area estimators, *Journal of the American Statistical Association*, **85**, pp. 163–171.
- R Development Core Team (2008). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, ISBN 3-900051-07-0. <http://www.R-project.org>. Vienna, Austria.
- Robinson, A. (2008). *icebreakR*, available at: [cran.r-project.org/doc/contrib/Robinson-icebreaker.pdf](http://cran.r-project.org/doc/contrib/Robinson-icebreaker.pdf)
- Schall, R. (1991). Estimation in generalized Linear Models with random Effects. *Biometrika*, Vol. 78, No. 4, pp. 719–727

- Scheipl, F., Greeven, S. y Küchenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis*, Vol. 52, No. 7, pp. 3283–3299
- Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed model. *Biometrics*, **50**, pp. 1171–1177.
- Salazar, J. C. Schmitt, F. A. Yu, L. Mendiondo, M. M. and Kryscio, J. (2005). Shared random effects analysis of multi-state Markov models: applications to a longitudinal study of transitions to dementia. *Statistics in Medicine* Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.21002/ sim2437.
- Skrondal, A. and Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika* 68, 267–287.
- Smith, A. F. M. (1973). A general bayesian linear model. *Journal of the Royal Statistical Society, Series B*, Vol. 35, 67–75.
- Ten Have, T. R. Miller, M. Reboussin, B. and James, M. K. (2000). Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the Longitudinal Study of Aging. *Biometrics* 56, 279–287.
- Verbeke G., and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer-Verlag. New York.
- Watts, D. G. (1981). A task analysis approach to designing a regression analysis course. *The American Statistician*, Vol. 35, No. 2, pp. 77–84.

- Wolfinger, RD. (1996). Heterogeneous variance covariance structure for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, **1** (2): 205–230.
- Youden, W.J. (1964). *Statistical methods for chemist*. Sixth Ed. John Wiley & Sons. New York.
- Zewotir, T. and Galpin, JS. (2005). Influence diagnostics for linear mixed models. *Journal of Data Science*, **32**, 153–177.

# Índice alfabético

- Ajuste del modelo lineal generalizado, 234
- Algoritmo de dos pasos, 200
- Algoritmo de Newton-Raphson, 65
- Algoritmo EM, 63, 174, 175
- Algoritmo EM Monte Carlo (MCEM), 174, 176
- Algoritmo Metropolis, 175
- Algoritmo Monte Carlo Newton Raphson (MCNR), 177, 178
- Algoritmo scoring, 177
- Análisis de incrementos, 31
- Análisis de residuales, 120
- Aproximación de mezclas de  $\chi^2$ 's, 117
- Aproximación de Taylor de primer orden, 172
- Area under curve, 31
- Chequeo de normalidad, 86
- Chi-cuadrado invertida, 90
- Chi-cuadrado invertida escalada, 94
- Cluster, 10
- Clustered data, 11
- Coefficiente de correlación intraclase, 42
- Componentes de varianza, 35, 128
- Consistencia, 62
- Conteo de pasos en humanos, 214
- Contraste de error, 55
- Covarianza generalizada, 226
- Covarianza marginal, 169
- Covarianza sin estructura, 126
- Criterios de información, 69
- Datos 100 metros planos, 27
- Datos árboles de abeto, 29
- datos agrupados, 11
- Datos conteo de pasos, 26

- Datos desbalanceados, 11
- Datos duración de pilas, 26, 209
- Datos electroplateado, 19
- Datos hueso ramus, 22
- Datos longitudinales, 11
- Datos secado de madera, 25
- Datos semen de toros, 26
- Densidad posterior, 197
- Descomposición de Cholesky, 120, 121
- Descomposición espectral, 60
- Descomposición espectral de una matriz, 220
- Diagnósticos en el modelo mixto, 162
- Diagnósticos vía eliminación de observaciones, 162
- Diferenciación de formas cuadráticas, 224
- Diseños ANOVA mixtos balanceados, 62
- Distancia de Cook, 164
- Distribución a priori, 85
- Distribución conjunta y condicional, 93
- Distribución hiperapriori, 88
- Distribución marginal de las componentes de varianza, 100
- Distribución posterior, 85
- EBLUP, 130
- Ecuación estimadora, 56
- Ecuaciones del modelo mixto de Henderson, 130
- Efectos aleatorios, 12
- Efectos fijos, 12, 13, 32
- El criterio de información bayesiano de Schwarz (BIC), 70
- El criterio de información corregido de Akaike (AICC), 70
- El criterio de información de Akaike (AIC), 69
- Electroplateado, 132, 210
- Error residual, 129
- Esperanza condicional, 129
- Esperanza Incondicional, 129
- Esperanza incondicional, 129
- Estadístico de Andrews-Pregibon, 166
- Estadístico de Cook-Weisberg, 165
- Estadísticos de resumen, 31
- Estimación en el MLMG, 170

- Estimación por máxima verosimilitud, 57
- Estudios longitudinales, 45
- Expectation, 63
- Estimador Bayes empírico, 169
- Extensión al caso de efectos aleatorios, 194
- Estimador de mínimos cuadrados generalizados, 57
- Factor de Cholesky, 197
- Estimador empírico de Bayes, 86
- Familia exponencial, 229
- Estimador naive, 105
- Fisher's Scoring, 66
- Estimador robusto, 105
- Forma matricial del MLM, 35
- Estimador sandwich, 105
- Forma matricial del modelo lineal clásico, 36
- Estimadores Bayes empíricos, 87
- Forma matricial del modelo no lineal, 195
- Estimadores de contracción o encogimiento, 86
- Formas Cuadráticas, 217
- Estimadores Shrinkage, 86
- Estructura ante-dependiente, 122
- Formas cuadráticas, 217
- Estructura AR(1), 123
- Formulaciones especiales del MLM, 39
- Estructura ARMA(1, 1), 123
- Función de linkeo, 172
- Estructura auto-regresiva heterogénea de primer orden, 123
- Función de vínculo, 187, 231
- Estructura de factor analítico, 125
- Función de vínculo canónica, 231
- Estructura de factor analítico con diagonal igual, 126
- Función de verosimilitud marginal, 57
- Estructura de factor analítico no diagonal, 125
- Función objetivo perfilada, 57
- Estructura Huynh-Feldt, 127
- Funciones de vínculo Log, 234
- Estudios de corte transversal, 45
- Funciones de vínculo Logit, 233
- Funciones estimables, 129, 130
- Funciones predecibles, 130

- Genética porcina, 179
- Gráficos Q-Q, 121
- Grados de libertad residuales, 107
- Hessiana, 65, 111
- Hiperparámetro, 90
- Hueso ramus, 146, 211
- Inferencia, 87
- Inferencia para la media, 104
- Inferencia para las componentes de  
varianza, 111
- Inferencias en el MLM, 51
- Influencia en los efectos aleatorios, 166
- Influencia en los estimadores de los  
efectos fijos, 164
- Influencia sobre la verosimilitud, 167
- Innovaciones o shocks, 48
- Intercepto aleatorio, 30
- Intervalos de confianza, 138
- Inversa, 223
- Inversa generalizada, 61, 108, 224
- La aproximación de Crainiceanu y Ruppert, 117
- Last Observation Carried Forward, 31
- Log-verosimilitud marginal, 56
- Máxima verosimilitud residual, 59
- Máxima Verosimilitud Restringida, 57,  
59, 199
- Máxima Verosimilitud Restringida (REML),  
58
- Máxima verosimilitud simulada, 179
- Método Bayes empírico, 84
- Método de contención, 108
- Método de Kenward-Rogers, 111
- Método de los momentos, 68
- Método de Satterthwaite, 109–111
- Método no iterativo, 67
- Método residual, 107
- Métodos Monte Carlo, 174
- Mínimos cuadrados, 197
- Mínimos cuadrados generalizados, 39,  
66, 196
- Matrices particionadas, 222
- Matriz de proyección, 221
- Matriz definida negativa, 218
- Matriz definida positiva, 218
- Matriz semidefinida positiva, 218
- Maximization, 63
- Media a priori, 96
- Media posterior, 96, 196

- Medida de influencia, 165
- Medidas de bondad de ajuste, 120
- Mejor estimador lineal insesgado, 130
- Mejor predictor lineal insesgado, 55, 68, 89, 128, 130
- Mejor predictor lineal insesgado estimado o empírico, 130
- MLM para el modelo de regresión simple, 39
- Moda posterior, 197
- Modelo autorregresivo de primer orden, 114
- Modelo bayesiano, 91
- Modelo de dos etapas, 32
- Modelo de efectos aleatorios simple, 122
- Modelo de efectos mixtos no lineal, 197
- Modelo de interceptos aleatorios, 38, 72
- Modelo de interceptos y tendencias aleatorias, 47
- Modelo de Laird-Ware, 37
- Modelo de medidas repetidas con estructura de covarianza especificada, 77
- Modelo de regresión con errores autocorrelacionados, 48
- Modelo jerárquico bayesiano, 88
- Modelo lineal, 10
- Modelo lineal general, 230
- Modelo lineal generalizado, 232
- Modelo lineal generalizado mixto, 170
- Modelo lineal mixto, 34
- Modelo lineal mixto frecuentista, 91
- Modelo lineal mixto tradicional, 171
- Modelo mixto de Henderson, 53
- Modelo mixto no lineal, 190
- Modelo no lineal, 191
- Modelo no lineal tradicional, 190
- Modelos de población promedios, 12
- Modelos específicos de sujetos, 12
- Modelos jerárquicos, 51
- Modelos lineales generalizados, 229
- modelos lineales mixtos, 11
- modelos log-lineales, 234
- Modelos marginales, 51, 52
- Modelos mixtos, 13, 29
- Modelos no lineales, 190

- Modelos no lineales que son linealiza-  
bles, 191
- Número de errores en una tarea, 188
- Newton-Raphson, 177
- Norma de una matriz, 219
- Norma matricial euclideana, 220
- Norma matricial máximo de colum-  
na, 220
- Norma matricial máximo de fila, 220
- Normalidad asintótica, 62
- Normalidad de los efectos aleatorios,  
168
- Observaciones futuras, 130
- Operadores esperanza y covarianza,  
225
- Outliers, 84
- Parámetro condicionalmente lineal, 193
- Parámetros condicionalmente lineali-  
zables, 193
- Paso efectos lineales mixtos, 200
- Paso pseudo-datos, 200
- Perfil promedio poblacional, 12
- Perfilamiento, 56
- Predicción, 128
- Predictor lineal, 187
- Procesos aleatorios jerárquicos, 11
- Producto, 223
- Producto de Kronecker, 36
- Profiling, 56
- Propiedades de los estimadores REML,  
61
- Prueba de la razón de la verosimili-  
tud restringida, 117
- Prueba de Wald para componentes  
de varianza  $F_W$ , 118
- Pruebas  $F$  aproximadas implementa-  
das en SAS, 119
- Pruebas de hipótesis, 100
- Pruebas de hipótesis sobre las varian-  
zas de los efectos aleatorios,  
116
- Pruebas para los efectos aleatorios,  
113
- Pruebas tipo Wald, 86, 235
- Pseudo-datos, 197, 198
- Puntajes, 159
- Rango, 222
- Razón de determinantes de matrices  
de varianzas, 164

- Razón de Odds, 234
- Recomendación para la construcción del modelo lineal mixto, 119
- Regresión lineal simple, 31
- Regresión logística, 233
- Residuales Cholesky, 121, 136
- Residuales condicionales, 121
- Residuales crudos, 121, 136
- Residuales de Pearson, 136
- Residuales estandarizados, 136
- Residuales normalizados, 136
- Residuales studentizados, 164
- Resultados de álgebra Lineal, 217
- Riesgos relativos, 234
- RLRT vía bootstrap, 118
- Scoring, 177
- Secado de madera, 212
- Sesgo del estimador de ML, 58
- Simetría compuesta, 124
- Simetría Compuesta (Compound Symmetry), 42
- Simetría Compuesta (Compound Symmetry), 47
- Simetría compuesta heterogénea, 124, 127
- Simplex de Nelder-Mead, 66
- Submuestreo, 132
- Subreplicación, 132
- Técnica Bayes empírico, 168
- Tamaño del efecto, 44
- Tamaño muestral, 43
- Tamaño muestral para comparar dos medias, 43
- Test aproximado de Wald, 106
- Test aproximados a la distribución F, 107
- Test de razón de verosimilitud, 69
- Tests de razón de verosimilitud (LRT), 112
- Tests LRT bajo REML, 112
- Toros, 140, 211
- Traza, 165, 221
- Valor esperado de una forma cuadrática, 227
- Valores y vectores característicos o propios, 218
- Variabilidad entre-sujetos, 32
- Variabilidad intra-sujeto, 32
- Variación entre-individuos, 88

Variación intra-individual, 88

Variación total, 219

Varianza generalizada, 219

Vector de efectos fijos, 57

Vector de media posterior, 96

Verosimilitud perfilada, 56

Los modelos mixtos se han convertido en una herramienta fundamental para el investigador en el área de modelamiento estadístico y ha tenido un amplio desarrollo en los últimos veinte años. Con este texto se pretende llegar al público de habla española que tenga la necesidad de modelar datos mediante esta técnica. Entre los tipos de datos más comunes que se benefician de los modelos mixtos, se encuentran los datos longitudinales, datos de panel y medidas repetidas. En este texto se hace un desarrollo a nivel teórico y además se presentan aplicaciones con datos reales, que ilustran la implementación de modelos mixtos en el software estadístico R y SAS.