

# Introducción a los modelos mixtos (SIA 3011003)

Profesor Juan Carlos Salazar-Uribe  
jcsalaza@unal.edu.co



**Descripción del curso**<sup>1</sup>. El objetivo de este curso es aprender a usar el modelo de regresión mixto para la generación de conocimiento valioso y de calidad, en presencia de datos longitudinales y medidas repetidas. Además de mostrar aplicaciones a datos reales, se describen aspectos relacionados con la teoría matemática y estadística de estos modelos. El curso tiene un nivel matemático intermedio, en el cual se hacen cálculos y demostraciones matemáticas del nivel de pregrado en estadística. Se hará uso de los software estadísticos SAS<sup>®</sup>, a través de SAS Studio<sup>®</sup> y R a través de R Studio<sup>®</sup> para ilustrar y practicar los contenidos aprendidos.

---

<sup>1</sup>This material is largely based on Fitzmaurice, Laird, and Ware (2011). *Applied Longitudinal Analysis*. 2nd Ed. Wiley: New Jersey. and JSM Professional Development Program Continuing Education Course on *Applied Longitudinal Analysis*. Other references were also used.

## Objetivos

- Estudiar los conceptos básicos tanto de modelos lineales mixtos (LMM) como de modelos lineales mixtos generalizados (GLMM).
- Implementar modelos mixtos lineales en la práctica usando datos longitudinales y mediciones repetidas.
- Familiarizar a los estudiantes con el software para implementar tanto LMM como GLMM (SAS<sup>©</sup> y R<sup>©</sup> principalmente).

**Evaluación** Asistencia (10 %), tareas y posiblemente exámenes (90 %). El material de las tareas se publicará al menos 2 o 3 semanas antes de la presentación de las mismas.

Se recomienda encarecidamente la asistencia, cada vez que no asista a la clase, se deducirá 0.1 de la nota final.

**Software.** Software que recomiendo usar:

- **SAS Studio**® (software SAS® en la nube-gratis) al que se puede acceder de forma gratuita (requiere registro) desde:  
[https://www.sas.com/en\\_us/software/on-demand-for-academics.html](https://www.sas.com/en_us/software/on-demand-for-academics.html)
- **R Studio**® (R Studio Desktop, Open Source License Free) que se puede descargar desde  
<https://www.rstudio.com/products/rstudio/download/>.  
Tiene que instalar primero R® desde  
<https://www.r-project.org/>

Puede usar cualquiera de ellos, pero a veces un procedimiento puede implementarse más fácilmente en SAS® y viceversa.

## Contenido

- Introducción: El modelo lineal clásico
- Ejemplos de motivación
- Datos longitudinales. Conceptos básicos
- Modelos Lineales Mixtos (MLM o LMM). Efectos fijos y aleatorios.
- Modelos mixtos lineales generalizados (MLMG o GLIMMIX o GLMM)

Antes de proceder a responder preguntas de investigación (o de interés), es aconsejable prestar atención a la manera en que se deben producir datos.



Figura 1: Generación de conocimiento a partir de datos

Las preguntas de investigación sobre cambios y ocurrencia de eventos son típicas en la práctica. En algunos casos, los investigadores preguntan cómo maduran o crecen las personas y cómo se desarrollan sus habilidades cognitivas en el tiempo; en otros preguntamos si ocurre un evento de interés y cuándo. Por lo general, este tipo de situaciones interesantes se abordan utilizando **datos longitudinales**.



En otras palabras, responder a un tipo particular de pregunta de investigación podría requerir un enfoque estadístico diferente. (Singer y Willett, 2003). Por ejemplo, las preguntas sobre **cambio** podrían requerir métodos como **modelo de crecimiento individual**, **modelo multinivel**, **modelo lineal jerárquico**, **modelo de coeficientes aleatorios** y **modelo mixto**. Por otro lado, las preguntas sobre **ocurrencias de eventos** se podrían abordar por medio de **análisis de supervivencia**, **análisis de historial de eventos**, **análisis de tiempo de falla** y **modelamiento de funciones hazard**. En este curso nos enfocamos principalmente en **Modelos Mixtos**.

Cuando la **variable de respuesta es continua**, los modelos de regresión lineal clásicos pueden extenderse para manejar los resultados correlacionados que son típicos en los datos longitudinales. Esta correlación entre medidas repetidas podría modelarse **explícitamente** (por ejemplo, mediante modelos de patrón de covarianza) o **implícitamente** (mediante la introducción de efectos aleatorios). El último enfoque produce una clase versátil de modelos de regresión para datos longitudinales conocidos como **modelos lineales de efectos mixtos** (Fitzmaurice y Laird, 2014).

En los últimos años se han producido avances destacados en los métodos para trabajar con datos longitudinales ya que no todos los estudios longitudinales utilizarán los mismos métodos estadísticos. MLM, con excepciones, por supuesto, **aún no se ha utilizado ampliamente**. Según Singer y Willett (2001), una de las causas de esta situación es que muchos libros populares de estadística aplicada no describen estos métodos, creando la impresión errónea de que las técnicas familiares, como el análisis de regresión, serán suficientes en estas aplicaciones longitudinales. **La falta de uso de nuevos métodos es un problema; no usarlos bien es otro.**

MLM es una poderosa herramienta estadística en la investigación aplicada, ya que no solo está ampliamente disponible en el software estándar, sino que también tiene una sólida base teórica. El término **modelo mixto** se refiere a la naturaleza de las partes del modelo que explican la media de un modelo estadístico (más adelante revisaremos este concepto y lo discutiremos a profundidad).

Por lo general, en los datos longitudinales, el resultado se observa repetidamente en diferentes momentos para cada sujeto en estudio. Los sujetos no son necesariamente observados ni al mismo tiempo ni el mismo número de veces (datos no balanceados). Las covariables también podrían ser diferentes para cada sujeto. Este tipo de datos tiene varias características:

- Cada sujeto tiene una tendencia (o trayectoria) particular y no son necesariamente líneas rectas.
- Si obtenemos la media de todas las observaciones, obtenemos una estimación del perfil promedio de la población.

Para este tipo de datos longitudinales se formulan los llamados **modelos específicos de sujeto** o **modelos poblacionales promedios** y dependiendo de los objetivos del estudio se puede estar interesado en uno u otro.

## ¿Por qué usar LMM para modelar datos longitudinales?}

- Porque es una herramienta flexible para modelar datos correlacionados.
- Porque la estimación se basa en la función de verosimilitud.
- Porque produce una prueba válida en diseños complejos con datos desbalanceados.
- Porque permite la estimación REML de las componentes de la varianza (Restricted Maximum Likelihood REML, que se analizará en detalle más adelante).
- Porque permite la inferencia de efectos aleatorios.
- Porque está disponible tanto en R<sup>©</sup> como en SAS<sup>©</sup>.

Se presentan algunos ejemplos motivacionales que ilustran situaciones prácticas en las que MLM podría usarse potencialmente para extraer información útil y de valor.

**Spruce tree growth data: crecimiento de árboles de abeto.** Este experimento se llevó a cabo para comparar el crecimiento de los abetos en dos entornos: control y entorno enriquecido con ozono. Hay 12 árboles por grupo. Para cada árbol se registraron 4 medidas de crecimiento a los días 152, 174, 201 y 227, respectivamente. En otras palabras, cada árbol contribuye con 4 medidas. El conjunto de datos:



# INTRODUCCIÓN

Obs	Group	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

Obs	Group	Id	$y_{152}$	$y_{174}$	$y_{201}$	$y_{227}$
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

## Grupo de Control.

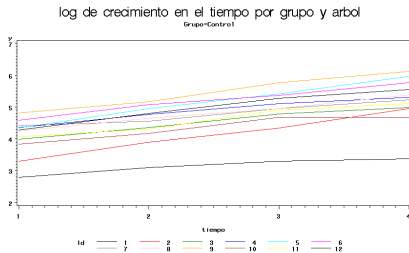


Figura 2: Perfiles grupo de Control

## Grupo de Ozono.

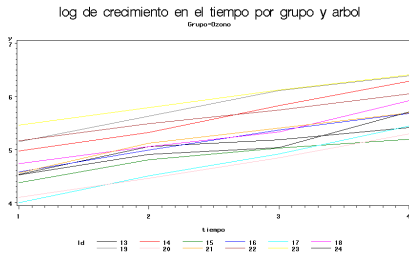


Figura 3: Perfiles grupo de Ozono

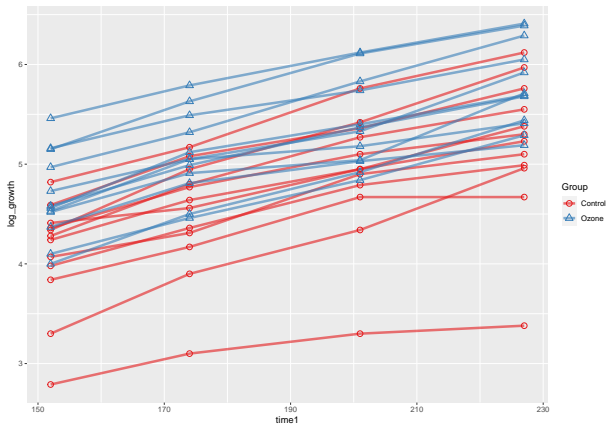


Figura 4: Perfiles ambos grupos. Abetos

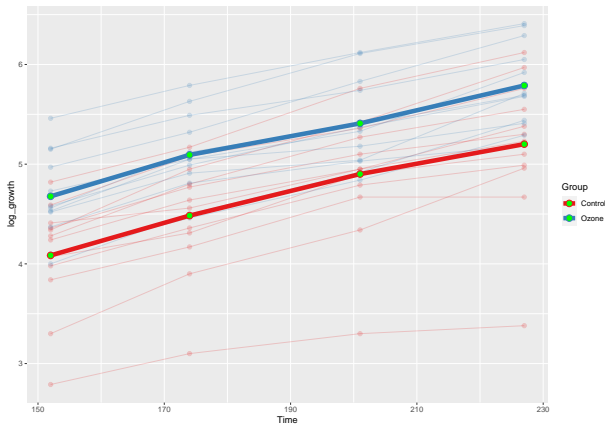


Figura 5: Perfiles Y promedios observados ambos grupos. Abetos

Estos gráficos muestran una tendencia lineal a lo largo del tiempo y cada árbol parece tener su propio intersepto aleatorio (las pendientes parecen ser las mismas). El conjunto de datos anterior, entre otros datos, se utilizará en clases posteriores para ilustrar diferentes técnicas de análisis típicas en MLM.

**Duración de baterías alcalinas.** Se seleccionaron seis marcas diferentes de pilas alcalinas para estudiar su duración en condiciones normales: RAYOVAC<sup>®</sup>, DURACELL<sup>®</sup>, VARTA<sup>®</sup>, ENERGIZER MAX<sup>®</sup>, EVEREADY ALKALINE<sup>®</sup>, PANASONIC POWER ALKALINE<sup>®</sup>. Para cada marca, se seleccionaron 2 baterías y se conectaron a un dispositivo paralelo especialmente diseñado para sacar la carga de cada batería. Cada hora, durante 11 horas, se midió el voltaje. Los datos son los siguientes:



# INTRODUCCIÓN

	RAYOVAC		DURACELL	
	VOLTAJE		VOLTAGE	
TIME	Battery 1	Battery 2	Battery 1	Battery 2
0	1217	1288	1419	1450
1	1181	1248	1359	1393
2	1077	1170	1284	1296
3	1040	1138	1222	1235
4	1043	1099	1173	1178
5	1004	1047	1103	1106
6	949	968	877	814
7	873	870	156	138
8	780	740	107	112
9	641	618	89	84
10	62	62	71	78
11	50	58	61	75

# INTRODUCCIÓN

\begin{center}

	VARTA		ENERGIZER	
	VOLTAGE		VOLTAGE	
TIME	Battery 1	Battery 2	Battery 1	Battery 2
0	1430	1308	1286	1268
1	1207	1150	1102	1060
2	1112	1052	1036	995
3	1056	993	1000	962
4	1025	953	972	940
5	992	917	934	905
6	952	818	874	853
7	870	805	759	763
8	246	172	166	109.2
9	96	77	75.4	70.4
10	69	65	62.6	66
11	64	60	60.1	65.1

# INTRODUCCIÓN

	EVEREADY		PANASONIC	
	VOLTAGE		VOLTAGE	
TIME	Battery 1	Battery 2	Battery 1	Battery 2
0	1276	1212	1217	1279
1	1086	1001	1005	1044
2	1034	955	943	980
3	992	910	897	958
4	959	879	871	930
5	886	844	804	879
6	825	788	718	823
7	693	679	602	736
8	118.8	142.4	718	612
9	75.5	75.1	602	95.5
10	64.7	64.3	117.6	73
11	61.2	61.5	81.7	70.3

Un gráfico de perfiles (o de espagueti) de los datos. Alrededor de las 5 horas se observa un decaimiento de los perfiles asociados a las diferentes marcas.

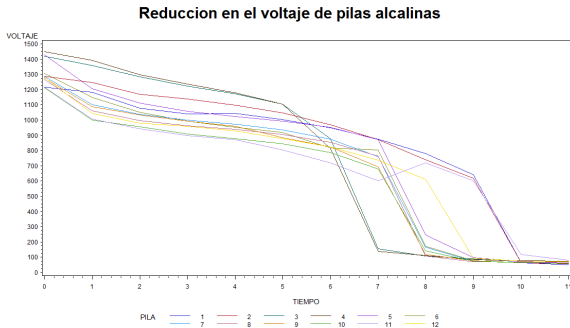


Figura 6: Perfiles duración pilas alcalinas

## Ensayo de tratamiento de niños expuestos al plomo

- La exposición al plomo durante la infancia se asocia con déficits sustanciales en las pruebas de capacidad cognitiva.
- Quelación<sup>2</sup>. El tratamiento de niños con altos niveles de plomo por lo general requiere inyecciones y hospitalizaciones.
- Un nuevo agente, **Succimer**, se puede administrar por vía oral.

---

<sup>2</sup>Eliminar (un metal pesado, como el plomo o el mercurio) del torrente sanguíneo por medio de un quelato, como el EDTA (ácido etilendiaminotetraacético; un ácido cristalino que actúa como un fuerte agente quelante y que forma una sal de sodio utilizada como antídoto para el envenenamiento por metales y como anticoagulante.

## Ensayo de tratamiento de niños expuestos al plomo

- Ensayo aleatorizado controlado con placebo que examina los cambios en el nivel de plomo en la sangre durante el curso del tratamiento.
- 100 niños asignados al azar a placebo o **Succimer**.
- Medidas del nivel de plomo en sangre al inicio, 1, 4 y 6 semanas.

**Ensayo de tratamiento de niños expuestos al plomo** Datos de ensayos de tratamiento de niños expuestos al plomo. Niveles medios de plomo en sangre (y SD) al inicio, semana 1, 4 y 6.

Grupo	Línea base	Semana 1	Semana 4	Semana 6
Succimer	26.5 (5.0)	13.5 (7.7)	15.5 (7.8)	20.8 (9.2)
Placebo	26.3 (5.0)	24.7 (5.5)	24.1 (5.7)	23.2 (6.2)

Grupo placebo: No reciben ingrediente activo. Grupo Succimer: reciben el nuevo fármaco (diferente a la quelación basada en inyecciones).

## Ensayo de tratamiento de niños expuestos al plomo

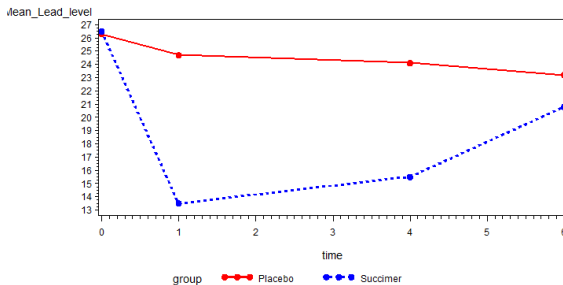


Figura 7: Gráfico de los niveles medios de plomo en sangre al inicio (Tiempo = 0), semana 1, 4 y 6 en los grupos Succimer y Placebo



**Ensayo de tratamiento de niños expuestos al plomo** Los niños del grupo placebo tienden a reducir ligeramente sus niveles de plomo en sangre durante el período de observación; por otro lado, los niños del grupo Succimer muestran una fuerte disminución en sus niveles de plomo en sangre, especialmente durante las semanas 1 y 4 y luego esos niveles aumentan, pero en general observamos, en el grupo Succimer, una disminución en los niveles de plomo en sangre que es mucho más sustancial que la observada en el grupo placebo. ¿Cómo medir esta aparente diferencia?

**Crecimiento craneofacial en sujetos colombianos.** En la práctica, es común recopilar datos sobre un fenómeno a lo largo del tiempo. Por ejemplo, en estudios de ortodoncia, puede ser de interés estudiar las diferencias y cambios en el tiempo entre mujeres y hombres con respecto a las variables craneofaciales cuando no se ha recibido tratamiento de ortodoncia.

(Jimenez et al. 2019)<sup>3</sup>.



Figura 8: Datos de crecimiento facial

<sup>3</sup>Jimenez, ID., Villegas, L., Alvarez, G., Salazar-Uribe, JC. (2019). Modeling facial growth data in 49 untreated Colombian mestizo subjects during 18 years follow up using linear mixed models. *American Journal of Orthodontics and Dentofacial Orthopedics*

**Crecimiento craneofacial en sujetos colombianos.** En los estudios de ortodoncia, es posible observar patrones (perfiles individuales) como este, con respecto a una medida específica (altura facial anterior).  
Datos longitudinales

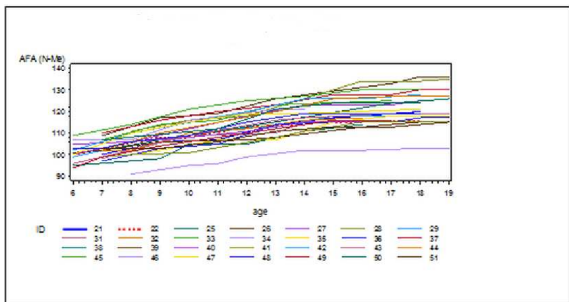


Figura 9: Datos de crecimiento facial

Crecimiento craneofacial en sujetos colombianos. Estos modelos permiten obtener **Perfiles Individuales**. Por ejemplo, considere estos 4 pacientes y un polinomio cuadrático mixto (datos reales):

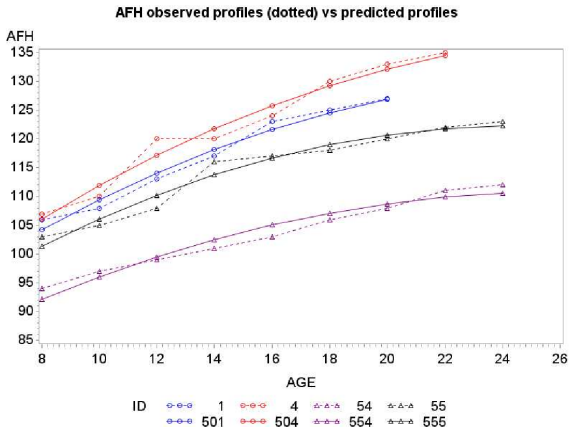


Figura 10: Datos de crecimiento facial

## Algunas ventajas

- Una característica de los estudios longitudinales es que las mediciones en los mismos individuos se toman repetidamente a lo largo del tiempo.
- Los estudios longitudinales permiten el estudio directo del cambio a lo largo del tiempo.
- El objetivo principal es caracterizar los cambios en la respuesta a lo largo del tiempo y los factores que influyen en el cambio.
- Con medidas repetidas, se puede capturar el cambio intra-individuo.

**Algunas dificultades.** Sin embargo, existen varias complicaciones cuando se trata de datos longitudinales:

- Las medidas repetidas en individuos están correlacionadas.
- La variabilidad suele ser heterogénea a lo largo del tiempo.
- Especialmente con respuestas categóricas o de conteos, se pueden presentar inestabilidades numéricas al ajustar este tipo de modelos, que pueden comprometer la calidad de la inferencia o la predicción.