

LMM Y GLMM

¿Cuándo necesitamos aplicar un GLMM?

La combinación de dos características requerirán la aplicación de modelos lineares generalizados mixtos:

- La variable respuesta no tiene distribución normal, pero su distribución es conocida.
- Más de una fuente de error aleatorio.

Distribución no normal  Modelos Lineales Generalizados

Múltiples fuentes de error aleatorio  Modelos de efectos mixtos

Linear regression & additive modelling

Generalised linear modelling & generalised additive modelling

Asumen que hay independencia en las observaciones

Allow for:

- Nested data
- Temporal correlation
- Spatial correlation
- Heterogeneity
- Repeated measurements

Mixed modelling & additive mixed modelling

Generalised linear mixed modelling & generalised additive mixed modelling



¿Qué son los efectos aleatorios?

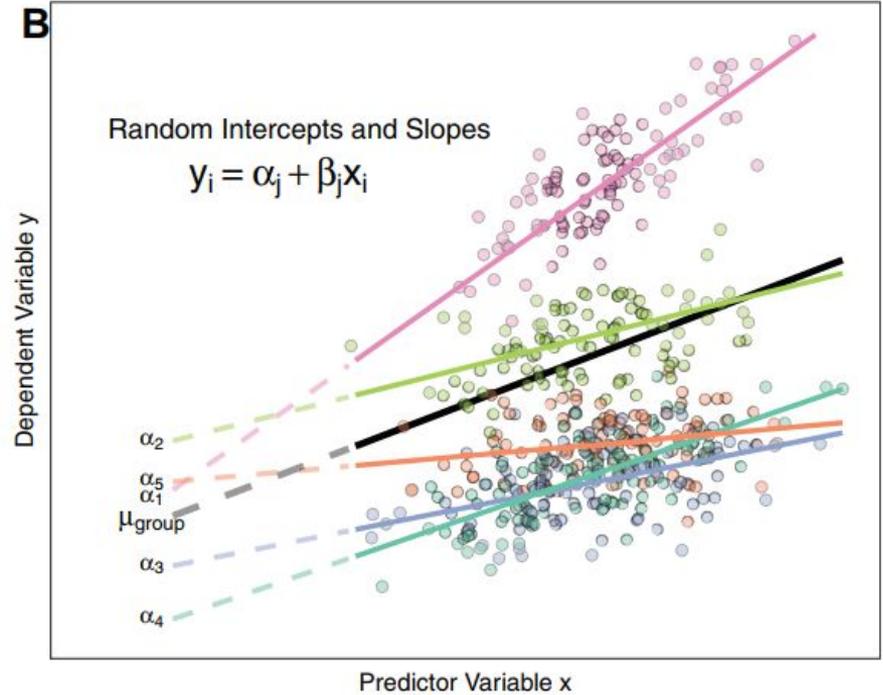
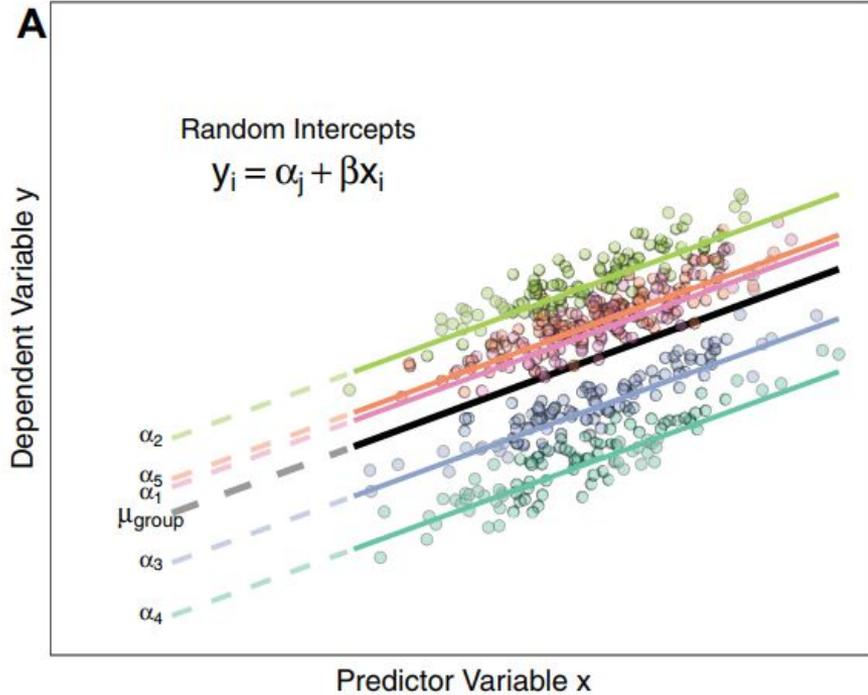


Medir el crecimiento de los polluelos a lo largo del tiempo

Los efectos aleatorios suelen representar alguna variable de agrupamiento (Breslow Clayton 1993) y permiten estimar la varianza de la variable de respuesta dentro de estos grupos y entre ellos.

Esto reduce la probabilidad de falsos positivos (Error de tipo I) y falsos negativos (Error de tipo II, e.g. Crawley 2013).

Random intercept y Random slope



Modelos con Random intercept y Random slope

```
M3 <- glmer(successful.breed ~ body.mass + (1|sample.site),  
family=binomial)
```



Random intercept

```
M4 <- glmer(successful.breed ~ body.mass + (body.mass|sample.site),  
family=binomial)
```



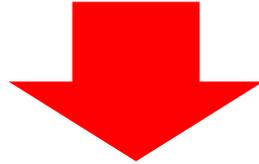
Random slope

Nos permiten entonces...

- ★ Controlar la no independencia entre los datos.
- ★ Mejorar la precisión en la estimación de los parámetros.
- ★ Estimar los componentes de la varianza.
- ★ Realizar predicciones para grupos no medidos.

Consideraciones importantes:

1. Requieren al menos cinco "niveles" (grupos) de un término de intercepción aleatoria para lograr estimaciones sólidas de la varianza. Con **<5 niveles**, es posible que el modelo mixto no pueda estimar con precisión la varianza entre poblaciones.



En este caso, la estimación de la varianza colapsará a cero, haciendo que el modelo sea equivalente a un MLG ordinario, o bien será distinta de cero pero incorrecta si el pequeño número de grupos muestreados no es representativo de la distribución real.

2. En segundo lugar, los modelos pueden ser inestables si los tamaños de las muestras entre los grupos están muy desequilibrados, es decir, si algunos grupos contienen muy pocos datos.

3. En tercer lugar, una cuestión importante es la dificultad de decidir la “significancia” o “importancia” de la varianza entre grupos.



Los **GLMM** son herramientas potentes, pero una parametrización incorrecta de los efectos aleatorios en el modelo podría producir estimadores en el modelo que sean tan poco fiables como ignorar por completo la necesidad de efectos aleatorios.



WARNING

Seleccionando los efectos aleatorios: ¿cruzados o anidados?

Va a estar determinado por el diseño experimental

Ejemplo: Un investigadora está interesada en entender los factores que afectan la masa de la nidada de un ave passerina.



Diseño experimental:



Una población de estudio en 5 fragmentos de bosque diferentes.

En cada uno de los fragmentos de bosques colocaron 30 cajas nido.

-Tasa de alimentación de las hembras

-Subsecuente nidada

Cada semana durante la época reproductiva

Supongamos que es un estudio de largo término...

Masa de la nidada ~ tasa de alimentación + (1|Fragmento de bosque /Hembra ID)

EFFECTOS ALEATORIOS ANIDADOS

Masa de la nidada ~ tasa de alimentación + (1|Fragmento de bosque /Hembra ID) + (1| Año)

EFFECTO ALEATORIO CRUZADO (AÑO)



En nuestro ejemplo anterior de las aves, podríamos ajustar una pendiente aleatoria para el efecto de la tasa de búsqueda de alimento en la masa de la puesta dada la identificación individual de cada ave. Es decir, la magnitud del efecto de la tasa de búsqueda de alimento sobre la masa de puesta resultante difiere entre las aves.

Masa de la nidada ~ Tasa de alimentación + (1|Fragmento de bosque /Hembra ID) + (1| Año)+ (Tasa de alimentación|Hembra ID)

Por ejemplo, si hubiéramos aplicado una alimentación de dos niveles de tratamiento a las aves de cada bosque, también podríamos esperar que la magnitud del efecto de recibir suplementos vitamínicos difiriese dependiendo del bosque en el que se aplicase. En este caso, especificaremos pendientes aleatorias aleatorias para la variable de tratamiento en función del ID del bosque.

Masa de la nidada ~ tasa de alimentación + (1|Fragmento de bosque /Hembra ID) + (1| Año)+ (Tasa de alimentación|Hembra ID)+(Tratamiento|Fragmento de bosque)

Cómo especificar la parte random:

Fórmula	Significado
$(1 group)$	Random group intercept
$(x group) = (1+x group)$	Random slope of x within group with correlated intercept
$(0+x group) = (-1+x group)$	Random slope of x within group: no variation in intercept
$(1 group) + (0+x group)$	Uncorrelated random intercept and random slope within group
$(1 site/block) = (1 site)+(1 site:block)$	Intercept varying among sites and among blocks within sites (nested random effects)
$site+(1 site:block)$	Fixed effect of sites plus random variation in intercept among blocks within sites

Fórmula	Significado
$(x site/block) = (x site)+(x site:block) = (1 + x site)+(1+x site:block)$	slope and intercept varying among sites and among blocks within sites
$(x1 site)+(x2 block)$	two different effects, varying at different levels
$x*site+(x site:block)$	fixed effect variation of slope and intercept varying among sites and random variation of slope and intercept among blocks within sites
$(1 group1)+(1 group2)$	intercept varying among crossed random effects (e.g. site, year)

Válido para los siguientes paquetes: lme4-nlme (nested effects only, although crossed effects can be specified with more work)-glmmADMB and glmmTMB. MCMCglmm uses a different specification, inherited from AS-REML.

Construimos el modelo:

1. Establecemos la distribución de la variable respuesta (e.g. Poisson, Binomial, Gamma).
2. Link function.
3. Número y tipo de efectos aleatorios (estrechamente relacionado al Diseño).

Importante: La sobredispersión debe probarse en el modelo más complejo.

FIN



Upper respiratory tract disease, force of infection, and effects on survival of gopher tortoises

ARPAT OZGUL,^{1,4} MADAN K. OLI,¹ BENJAMIN M. BOLKER,² AND CAROLINA PEREZ-HEYDRICH³

¹*Department of Wildlife Ecology and Conservation, 110 Newins-Ziegler Hall, University of Florida, Gainesville, Florida 32611 USA*

²*Department of Zoology, 223 Bartram Hall, University of Florida, Gainesville, Florida 32611 USA*

³*Department of Infectious Diseases and Pathology, University of Florida, Gainesville, Florida 32610 USA*

Pregunta/Hipótesis: La enfermedad del tracto respiratorio superior (URTD) causada por *Mycoplasma agassizii* ha contribuido al declive de algunas poblaciones salvajes de tortugas de tierra (*Gopherus polyphemus*).

Diseño: Año y Sitio como variables aleatorias cruzadas.

Variabes: Respuesta: Recuento de caparazones frescos Independiente: Seroprevalencia de micoplasma (es decir, la fracción de tortugas portadoras de anticuerpos contra la enfermedad).