Módulo 4

INTRODUCCIÓN A GLM y GLMM

Curso de Posgrado: “Modelado y estimación de ocupación para poblaciones y comunidades de especies bajo enfoque Bayesiano”

CCT CONICET Mendoza

24 - 28 Abril 2023

CONICET MENDOZA

Instituto Nacional de Tecnología Agropecuaria Argentina

INTA

GTBA

Grupo Transdisciplinario de Biodiversidad y Agroecosistemas

GIEPCO

Grupo de Investigación en Ecología Poblacional y Comportamental

CONICET

INTRODUCCIÓN A GLM

- Modelo lineal generalizado
- Extiende el concepto de modelo lineal con respuesta normal, como el análisis de varianza (ANOVA) y una regresión, a otras distribuciones de respuesta (ej. Poisson y Binomial)
- La comprensión sólida de los GLM es base crucial para cualquier ecólogo

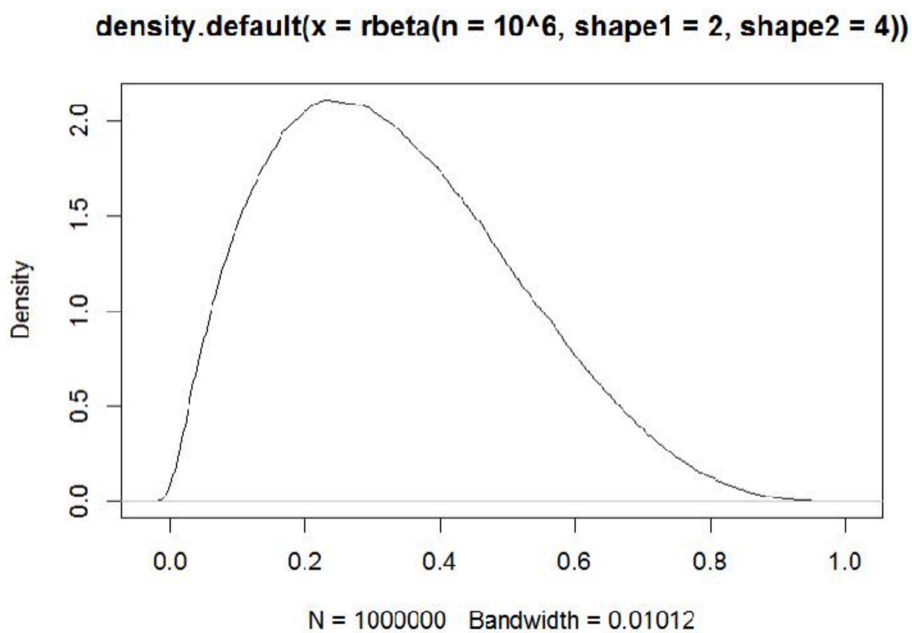
INTRODUCCIÓN A GLM

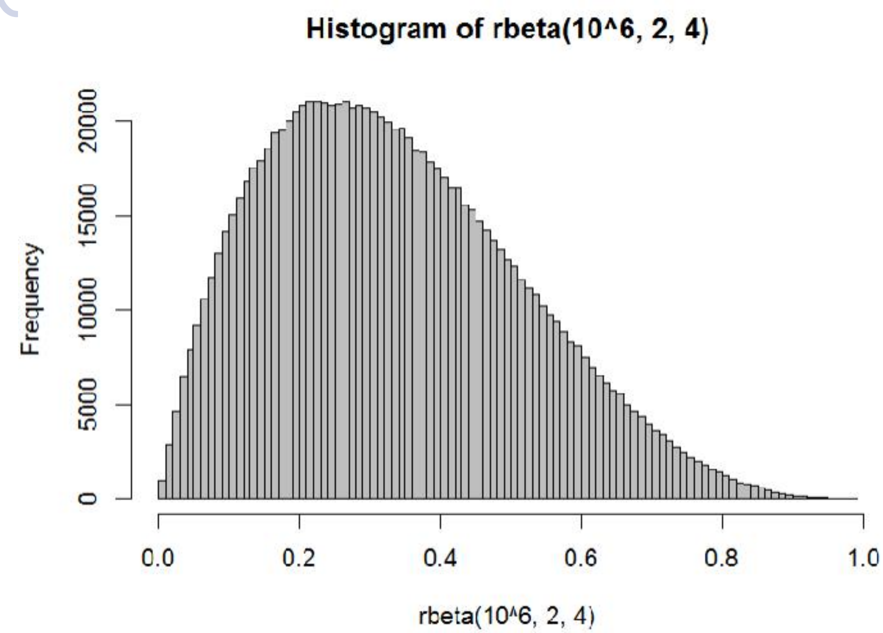
- La respuesta de un modelo con dos componentes: **SEÑAL + RUIDO**
- "RUIDO" es la variabilidad aleatoria en una respuesta observada
- Se podría entender como: la parte fija y aleatoria de un modelo, la media y la varianza, la dispersion, etc.
- Un GLM estrictamente tiene un sólo componente de **ruido**, pero a veces necesitamos varios en un modelo estadístico (ej. Modelos mixtos o efectos aleatorios, Jerárquicos, etc)
- Para describir el "ruido" usamos distribuciones estadísticas
- Los descriptores de la parte aleatoria de nuestro modelo son sus propios parámetros de dichas distribuciones.
- Más utilizados: Poison, binomial, normal, multinomial, exponencial

INTRODUCCIÓN A GLM

Hay muchas distribuciones en R, y podemos ver sus densidades y generar números aleatorios bajo una distribución

```
plot(density(rbeta(n=10^6, shape1=2, shape2 = 4)))
hist(rbeta(10^6, 2, 4), nclass=100, col="gray")
```



```
> rpois(10^6, 4)
  [1]  3  0  7  3  8  3  6  2  4  2  1  1  5  6  2  2  6  2  4  3  7  3  6  6  6  6  8  3
 [29]  3  2  3  3  7  4  3  5  5  4  6  3  8  5  3  3  3  0  2  6  7  7  3  2  8  2  2  5
 [57]  0  2  1  6  3  3  4  5  3  4 10  2  2  7  7  2  6  2  6  6  2  4  4  2  3  5  5  3
 [85]  5  4  4  8  4  5  3  3  3  7  4  3  2  4  3  3  7  4  4  5  2  2  1  6  4  2  5  4
[113]  5  3  6  1  2  6  6  6  4  0  3  4  5  3  4  5  5  7  5  4  3  6  4  4  1  5  6  5
[141]  6  2  2  4  3  1  3  3  2  2  3  6  5  3  4  6  3  5  7  4  1  5  4  2  5  1  6  6
[169]  1  4  4  3  5  7  3  7  4  7  4  6  2  3  4  6  4  4  4  4  2  4  1  7  1  4  2  7
[197]  3 10  2  3  8  4  3  4  5  3 10  5  4  1  3  5  6  2  3  4  3  4  4  6  2  2  5  3
[225]  2  3  4  4  7  6  3  2  7  5  4  5  5  3  5  3  3  4  5  2  1  2  8  1  4  8  2  4
[253]  5  9  2  3  4  6  3  4  3  7  6  1  4  6  8  5  7  3  2  4  4  5  1  3  4  2  8  4
[281]  5  3  2  0  7  3  6  6  5  4  7  4 10  5  3  9  7  4  4  3  1  4  2  7  4  4  4  3
[309]  3  4  1  4  5  5  1  4  7  3  0  6  9  5  6  8  4  4  4  4  3  6  4  5  4  5  6  4
[337]  4  3  2  4  6  6  2  2  3  5  7  6  3  4  3  2  7  2  5  2  4  0  4  4  3  7  2  7
[365]  5  7  3  1  4  1  3  5  6  4  6  2  6  8  1  5  4  4  3  2  2  2  2  8  5  6  4  4
[393]  4  3  7  4  3  4  2  6  3  4  4  1  4  3  4  5  3  5  0  6  7  6  2  5  6  2  5  6
[421]  2  3  6  4  1  2  3  5  3  5  4  2  4  4  8  4  1  3  3  2  7  5  3  1  6  2  5  1
[449]  3  1  2  8  3  6  5  1  3  3  3  7  4  3  5  6  4  4  3  4  2  4  3  1  1  7  5  5
[477]  7  1  5  8  3  2  1  1  4  1  6  4  4  2  2  4  2  4  6  4  3  2  7  4  4  4  4  8
[505]  3  3  6  3  3  7  3  3  3  7  3  3  4  4  4  3  7  2  8  5  3  5  1  7  2  3  3  7
[533]  3  3  2  3  4  0  2  5  6  7  5  7  4  4  3  3  2  4  6  3  5  6  4  4  5  6  4  3
[561]  6  6  3  8  7  4  5  6  4  3  2  4  5  4  2  5  8  4  3  3  3  7  6  5  2  4  3  5
[589]  6  4  5  2  9  2  4  7  6  2  3  3  5  5  5  5  3  3  7  4  1  4  5  6  4  3  4  4
```

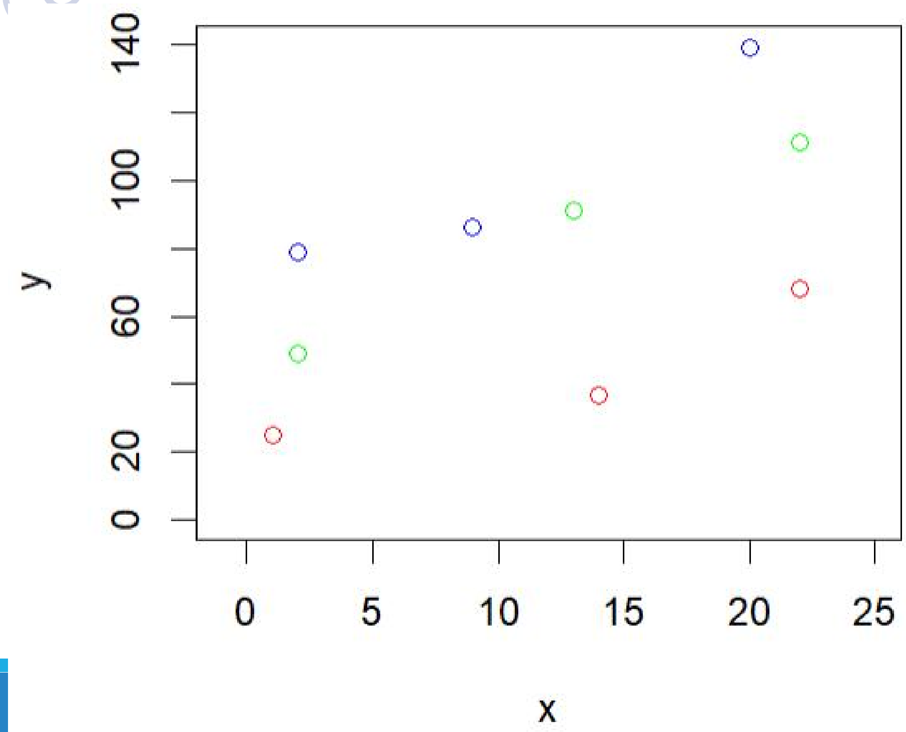
INTRODUCCIÓN A GLM

- El componente de la **"señal"** contiene las parte predecibles de la respuesta, o la estructura de la media de un modelo (el más conocido: modelo lineal)
- Los parámetros afectan a la respuesta media de una manera aditiva

$$y = \alpha * x_1 + \beta * x_2$$

- Los más conocidos test-t,

ANOVA, ANCOVA



INTRODUCCIÓN A GLM

$$y_i = \alpha_j + \beta * X_i + \varepsilon_i \qquad \varepsilon_i \sim Normal\,(0, \sigma^2)$$

Unidad de respuesta

Factor A_j

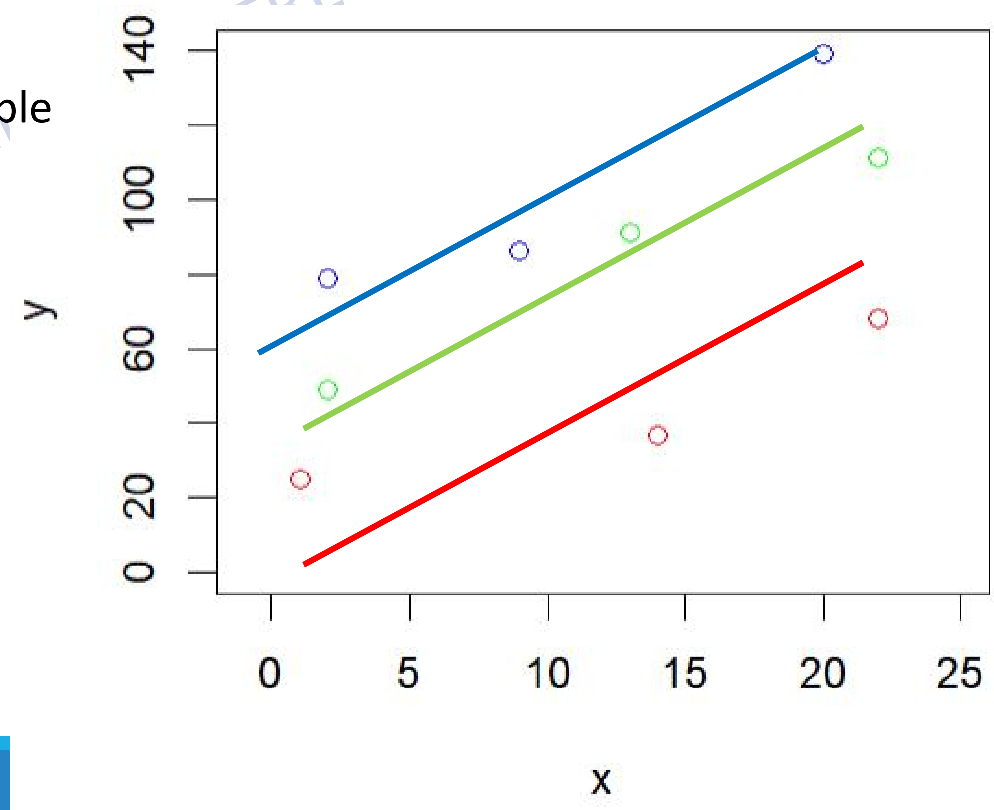
Valor de una variable continua

Residuos (no explicado por la combinación lineal)

$$\boxed{y_i = \alpha_j + \beta * X_i} + \boxed{\varepsilon_i}$$

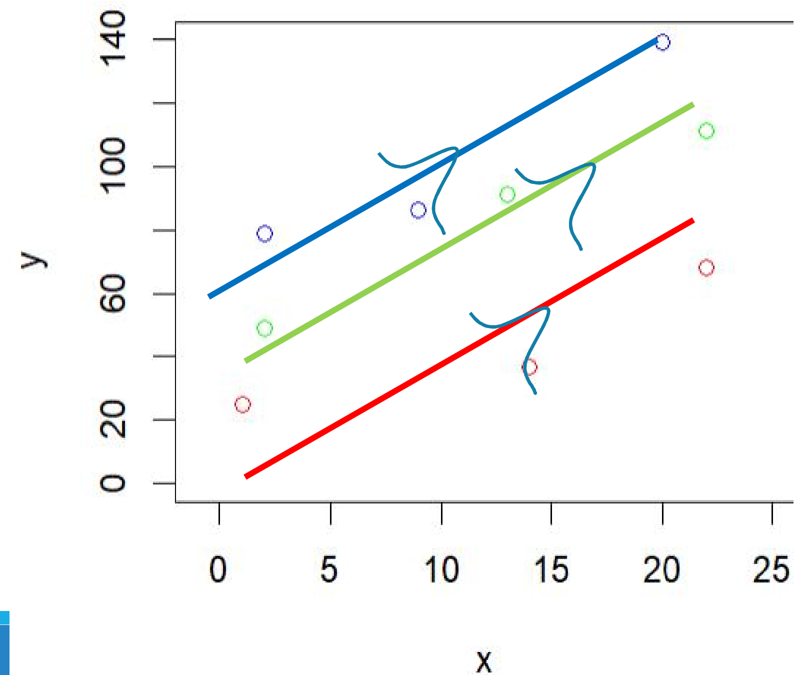
Señal

Ruido



FUNCIÓN "LINK" O ENLACE

- Hasta ahora asumimos una distribución normal para la parte aleatoria de la respuesta
- Cuando las respuestas no siguen una distribución normal, ya no podemos usar directamente el modelo aditivo (ej. Agregar ruido a una poison o binomial)
- Podemos aplicar indirectamente un modelo lineal, a través de una función de enlace que transforma el valor esperado
- La función de enlace, nos permite "enlazar" los componentes de señal y de ruido en un modelo



FUNCIÓN "LINK" O ENLACE

$$y_i \sim f(\mu_i)$$ ← Alguna distribución

Alguna función g → $$g(\mu_i) = \underbrace{\alpha + \beta * x_i}_{\text{Modelo lineal}}$$

GLM Poisson

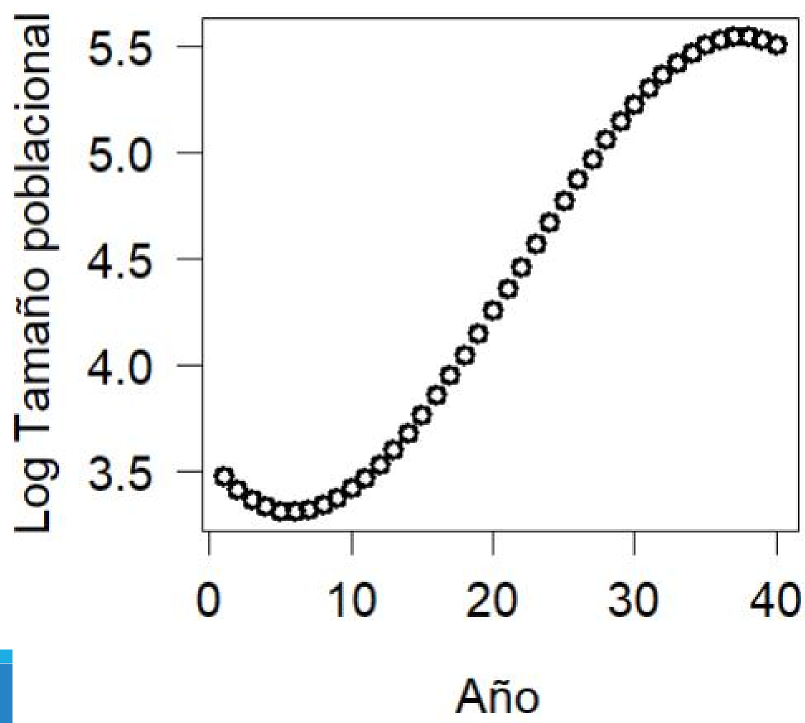
Poisson (ej. conteos)

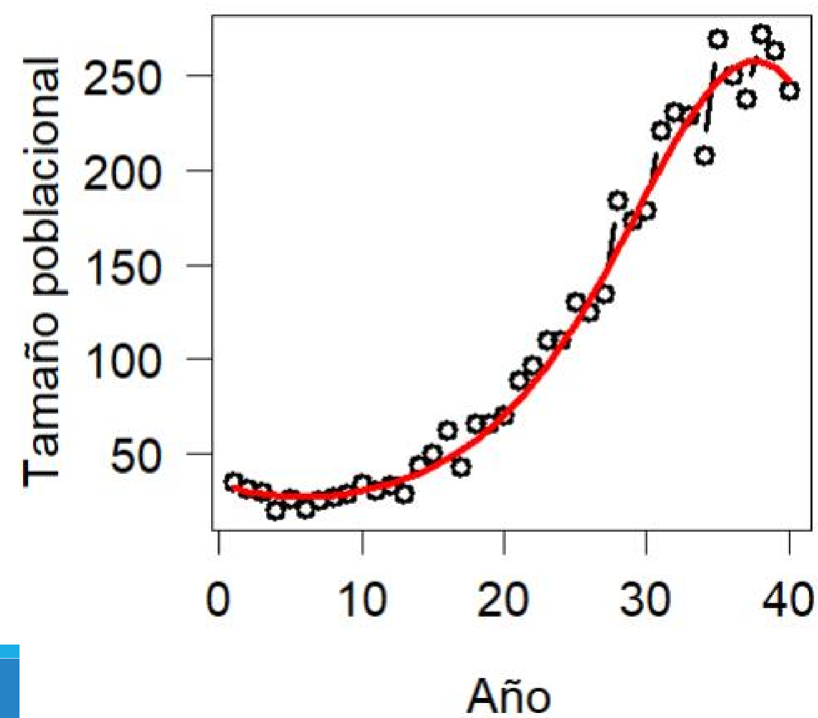
Función log-link

$$C_i \sim Poisson(\lambda_i)$$

$$\boldsymbol{log(\lambda_i) = \alpha + \beta * x_i}$$

λ_i es el conteo esperado (respuesta media)





GLM Binomial

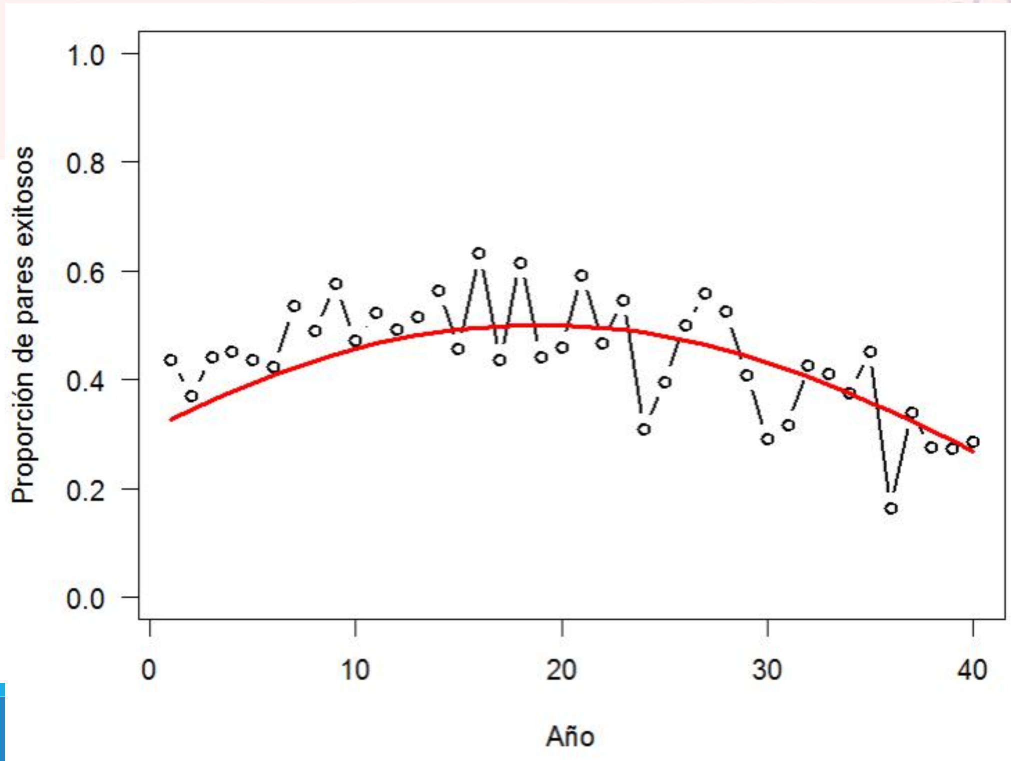
Binomial (ej. detecciones)

$$C_i \sim Binomial(N_i, p_i)$$

$$\boldsymbol{logit(p_i) = \alpha + \beta * x_i}$$

Función logit

$$logit(p_i) = log\left(\frac{p_i}{1 - p_i}\right)$$

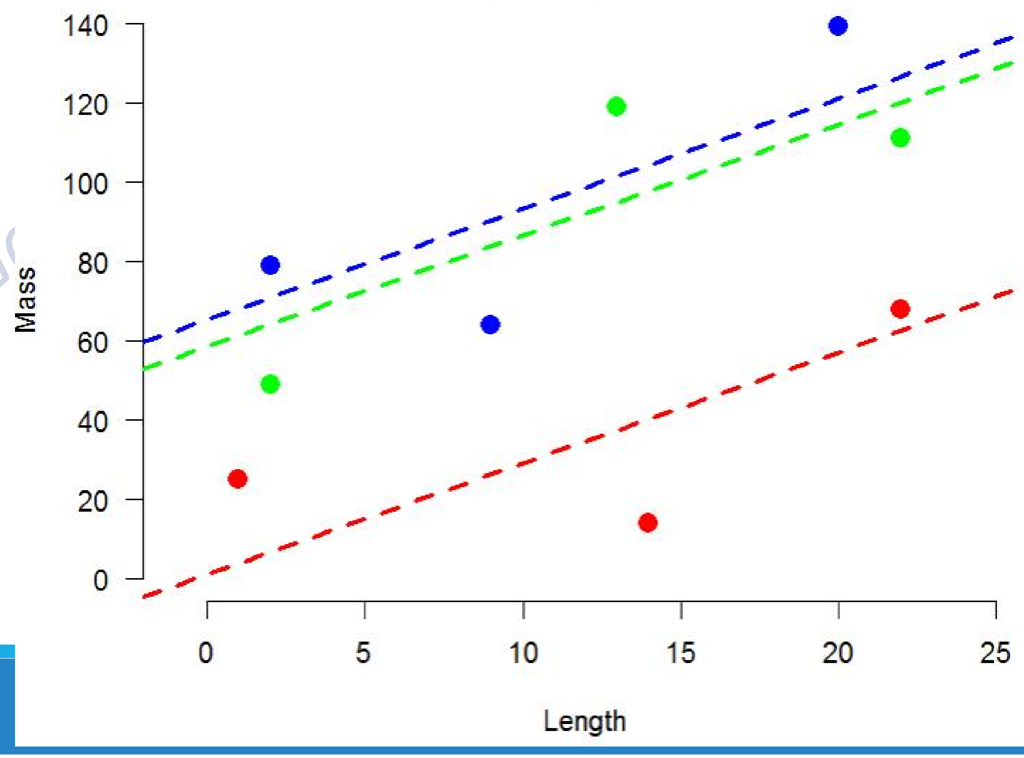


N_i es el número de intentos
pi probabilidad de éxito
(sólo dos eventos)

INTRODUCCIÓN A GLMM – Efectos Aleatorios

- Es un GLM con efectos fijos y aleatorios
- Ejemplo: se ve una relación lineal pero la base varía para poblaciones diferentes (poblaciones *j*, individuos *i*)

$$masa_i = \alpha_{j(i)} + \beta * longitud_i + \varepsilon_i$$



¿Qué son los efectos aleatorios?

1. Los efectos poblacionales $\boldsymbol{\alpha_j}$ son completamente independientes (y hay solo 3 poblaciones que nos interesan)
2. Los 3 $\boldsymbol{\alpha_j}$ no son independientes, y son una muestra de una población mayor
    1. Nos lleva a efectos fijos
    2. Nos lleva a efectos aleatorios

$$\boldsymbol{masa_i = \alpha_{j(i)} + \beta * longitud_i + \varepsilon_i}$$

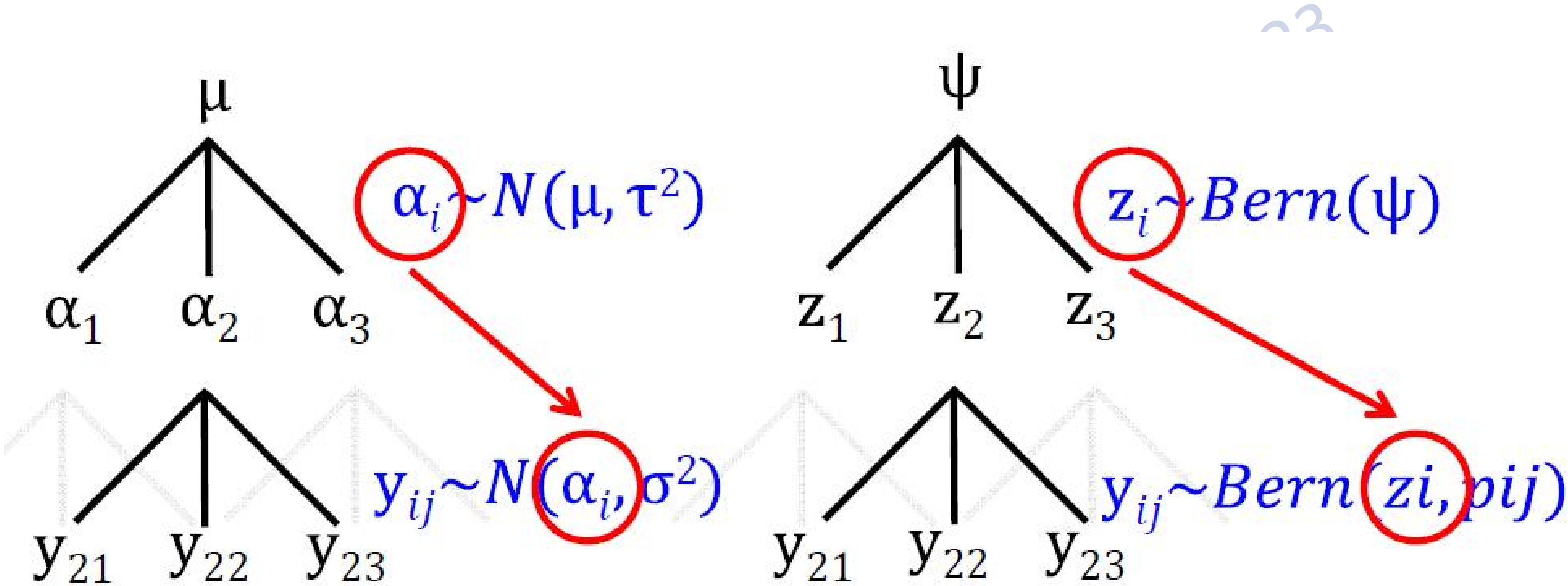
$$\boldsymbol{\alpha_j} \sim Normal\ (\mu_\alpha, \sigma_\alpha^2) \longrightarrow \alpha_j \text{ es aleatorio}$$

INTRODUCCIÓN A GLMM – Efectos Aleatorios

- Son dos o mas efectos que pertenecen juntos de alguna manera – son estimados bajo alguna limitación o una distribución común
- Generalmente somos libres de elegir si un efecto es fijo o aleatorio, pero hay casos en que ese supuesto no tiene sentido biológico.
- En el "lenguaje" Bayesiano una distribución previa que gobierna un set de efectos también tiene parámetros que deben ser estimados. Esos parámetros (de los efectos aleatorios) se denominan ***hiperparámetros,*** y sus previas se denominan ***hiperprevias***
- Cuando introducimos efectos aleatorios, introducimos una jerarquía de efectos.

MODELOS JERÁRQUICOS

- Secuencia dependiente de variables aleatorias (observadas y no observadas) – Ej. ANOVA en bloques aleatorios



  - Otros nombres: estado-espacio, aleatorios, modelos mixtos, variables latentes, etc…

Efectos Aleatorios y fijos

- A veces tenemos que tratar un efecto como fijo porque los efectos no vienen de una distribución común y pueden ser sistemáticamente diferentes
- Los efectos aleatorios tardan más en llegar a la convergencia
- Depende del alcance de la inferencia:
  - ¿Todos los niveles de un factor están incluidos en el modelo?   SI --- probablemente "fijo"
  - Pueden los niveles de un factor ser vistos como una muestra aleatoria de una población mayor?

    SI --- probablemente "aleatorio"

Efectos Aleatorios y fijos

- Las definiciones en la literatura son confusas!
- "depende" – muchas definiciones no siempre proveen la guía adecuada para saber si usar efectos fijos o aleatorios
- Un factor aleatorio con muy pocos efectos, puede tener estimaciones muy imprecisas
- Pensar bien que es lo que se quiere hacer…

Efectos Aleatorios y fijos

“Finalmente, puedes intentar convencerte a ti mismo (y a tu revisores, lectores o supervisor) que entre grupos la variación no es importante ajustando el modelo ignorando bloques y luego examinando la variación de los residuos entre los bloques tanto gráficamente como estadísticamente.”

“Para justificar ignorar la variación entre grupos en el modelo, debe demostrar la variación de los residuos entre grupos es tanto estadística como biológicamente irrelevante”.

“La variación biológicamente relevante es una importante señal de advertencia, incluso si no es estadísticamente significativa”.

Bolker 2008, p. 326

https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

GLMM FAQ

Ben Bolker and others

05 Oct 2022

?

GLMM – Efectos Aleatorios

1. Sólo ordenadas son aleatorias, pero las pendientes son idénticas para los grupos
2. Sólo pendientes son aleatorias, pero las ordenadas son idénticas para los grupos
3. Las ordenadas y pendientes son aleatorias

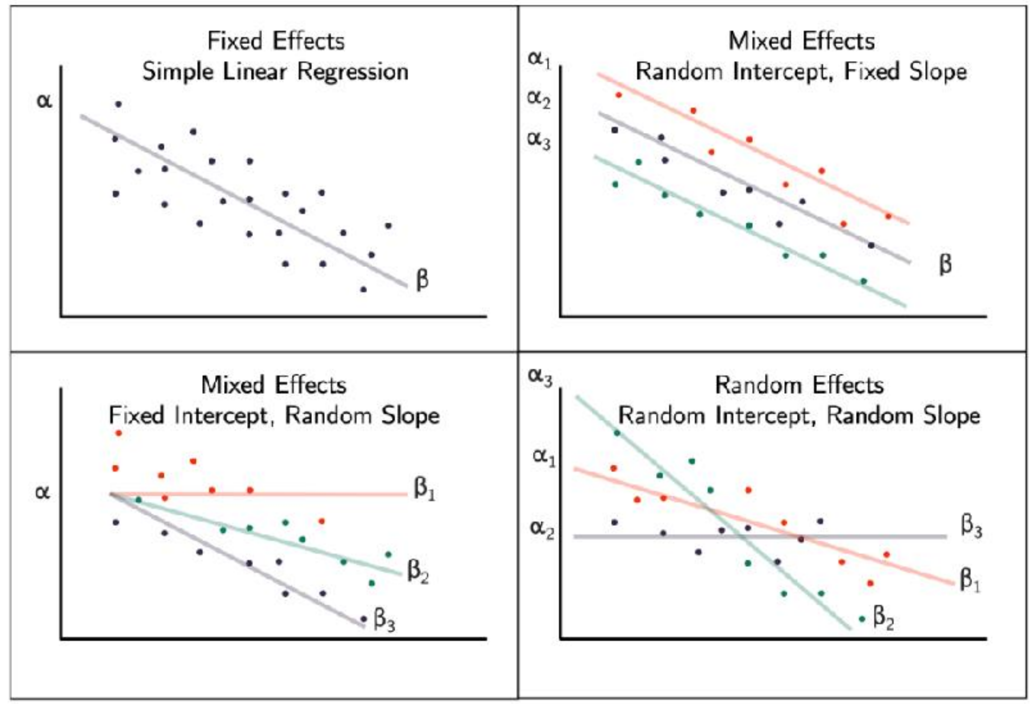


Figure 9.1: Fixed effect, mixed effects, and random effects linear regression models.

GLMM – Efectos Aleatorios

Ejemplo de ordenadas y pendientes aleatorias

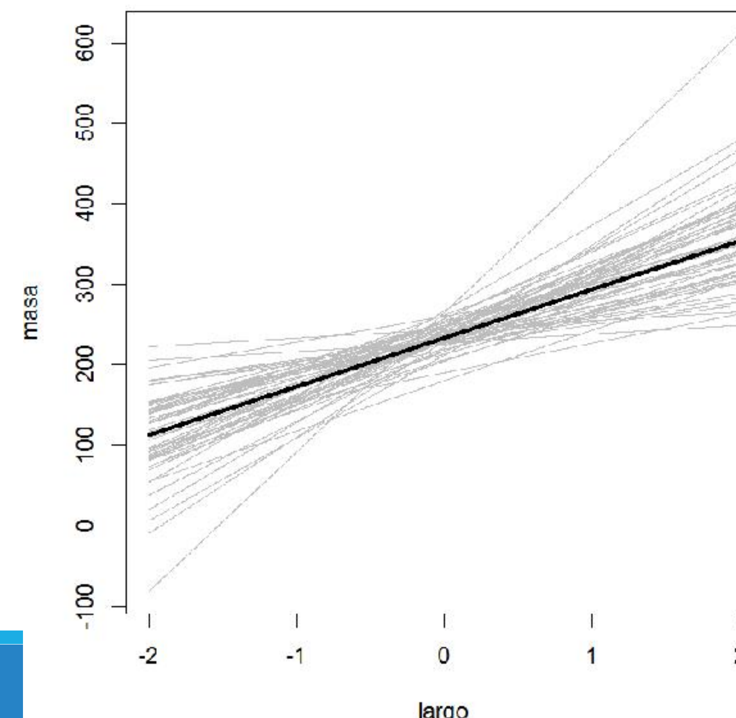
i son individuos
j son poblaciones
X es el tamaño

$$y_i = \alpha_{j(i)} + \beta_{j(i)} * x_i + \varepsilon_i$$

$\alpha_j \sim Normal\ (\mu_\alpha, \sigma_\alpha^2)$ # Efectos aleatorios de las ordenadas

$\beta_j \sim Normal\ (\mu_\beta, \sigma_\beta^2)$ # Efectos aleatorios de las pendientes

$\varepsilon_i \sim Normal\ (0, \sigma^2)$ # Efectos aleatorios residuales



Podemos hacer lo mismo con un GLMM poisson o binomial (no olvidar la función de enlace!)

GLMM – Efectos Aleatorios

Ejemplo de ordenadas aleatoria y pendiente fija

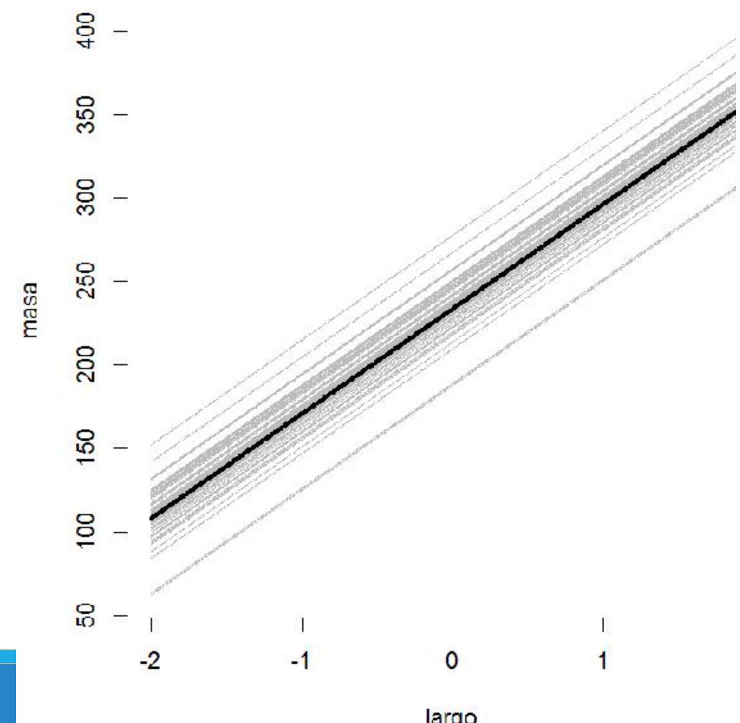
$$y_i = \alpha_{j(i)} + \beta * x_i + \varepsilon_i$$

i son individuos
j son poblaciones
X es el tamaño

$\alpha_j \sim Normal\ (\mu_\alpha, \sigma_\alpha^2)$ # Efectos aleatorios de las ordenadas

$\beta \sim Normal\ (0, \sigma^2)$ # Efectos fijo de la pendiente

$\varepsilon_i \sim Normal\ (0, \sigma^2)$ # Efectos aleatorios residuales



REFERENCIAS

- Bolker, B. 2008. Ecological Models and Data in R. Princeton University Press, Princeton, NJ.
- Marc Kery. 2010. Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses. Academic Press.
- Marc Kery and Michael Schaub. 2012. Bayesian population analysis using WinBUGS. A hierarchical perspective. Academic Press.
- Marc Kery & J. Andy Royle. 2016. Applied hierarchical modeling in ecology. Modeling distribution, abundance and species richness using R and BUGS. Volume 1: Prelude and Static models. Academic Press