

Modelos mixtos y otras extensiones

Javier Fernández-López, Profesor Ayudante Doctor
Unidad de Matemática Aplicada
Departamento de Biodiversidad, Ecología y Evolución, UCM

Programas de doctorado:

Agrobiología Ambiental

Calidad y Seguridad Alimentaria

Biodiversidad, Funcionamiento y Gestión de Ecosistemas

Bloque 3: Modelos mixtos y otras extensiones

- Factores de efectos aleatorios VS factores de efectos fijos
- Efectos aleatorios: intercepto y pendiente (interacciones)
- Concepto de autocorrelación
- Otros controles de la no-dependencia: autocorrelación temporal y espacial

EFFECTOS FIJOS VS EFFECTOS ALEATORIOS

- Hasta ahora hemos aprendido a utilizar predictores categóricos en los cuales **estábamos interesados**: sexo, tipo de hábitat, etc. Estos eran tratados como **efectos fijos**.
- Por el contrario, los **efectos aleatorios** son parámetros estadísticos que intentan explicar el **ruido** causado por agrupamientos o *clústeres* en los datos que intentamos modelizar.

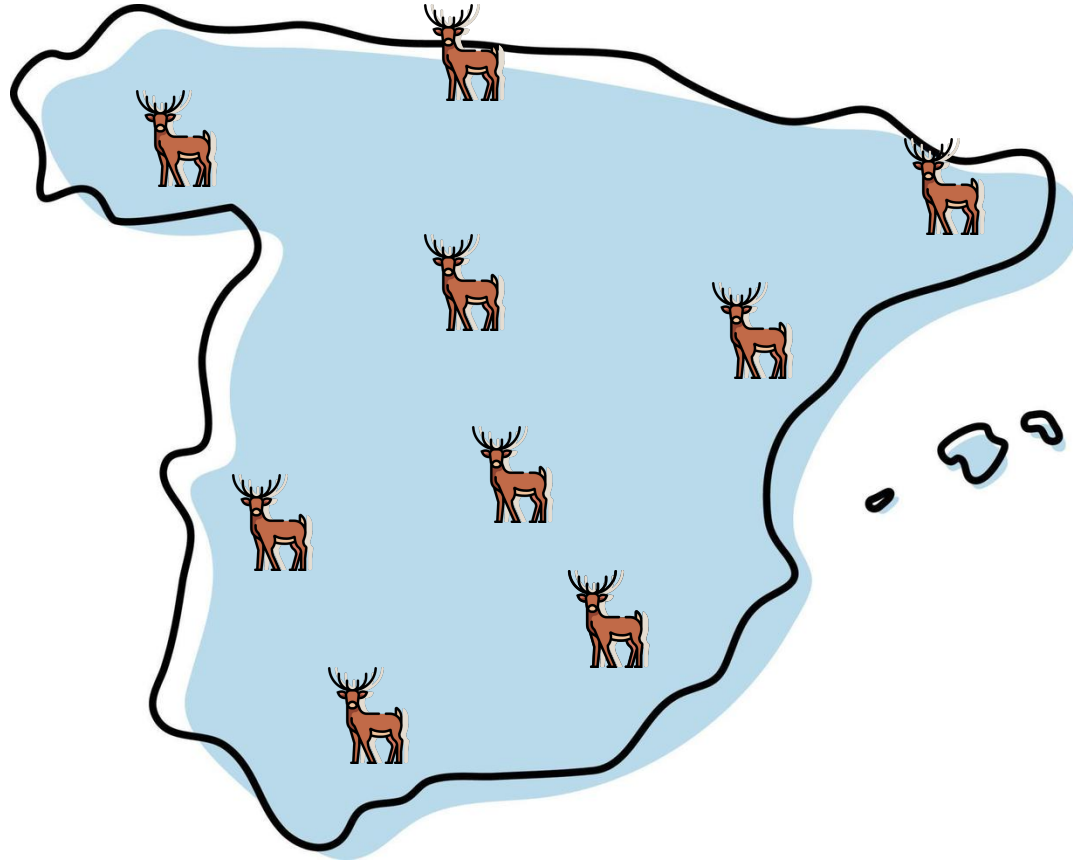
EFFECTOS FIJOS VS EFFECTOS ALEATORIOS

- Hasta ahora hemos aprendido a utilizar predictores categóricos en los cuales **estábamos interesados**: sexo, tipo de hábitat, etc. Estos eran tratados como **efectos fijos**.
- Por el contrario, los **efectos aleatorios** son parámetros estadísticos que intentan explicar el **ruido** causado por agrupamientos o *clústeres* en los datos que intentamos modelizar.
 - Estudiantes en diferentes colegios. Niveles del factor colegio?
 - Individuos en diferentes poblaciones. Niveles del factor población?
 - Muestras en diferentes establecimientos. . Niveles del factor establecimiento?
 - Hojas en diferentes árboles. Niveles del factor árbol?

EFFECTOS FIJOS VS EFFECTOS ALEATORIOS

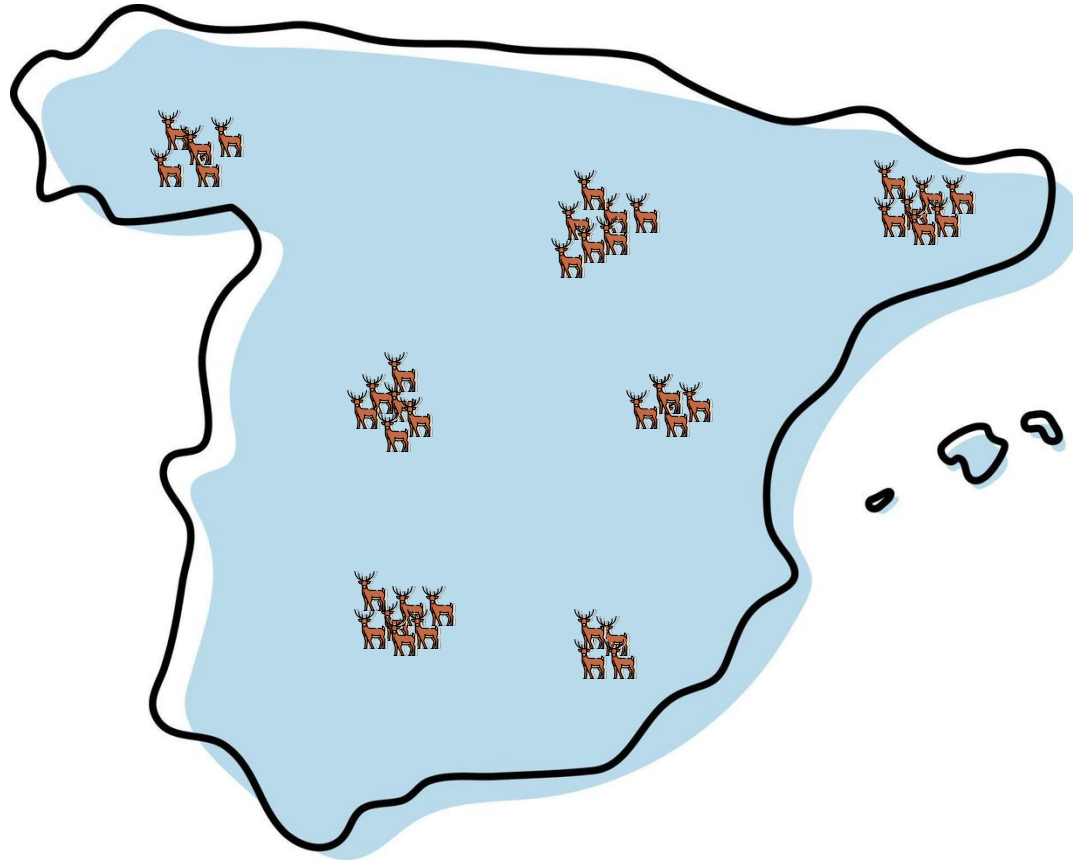
- Hasta ahora hemos aprendido a utilizar predictores categóricos en los cuales **estábamos interesados**: sexo, tipo de hábitat, etc. Estos eran tratados como **efectos fijos**.
- Por el contrario, los **efectos aleatorios** son parámetros estadísticos que intentan explicar el **ruido** causado por agrupamientos o *clústeres* en los datos que intentamos modelizar.
 - Estudiantes en diferentes colegios. Niveles del factor colegio?
 - Individuos en diferentes poblaciones. Niveles del factor población?
 - Muestras en diferentes establecimientos. . Niveles del factor establecimiento?
 - Hojas en diferentes árboles. Niveles del factor árbol?
- Sexo o clases de edad no son aleatorios: normalmente estamos interesados en su efecto, y si volviésemos a muestrear, encontraríamos los mismos niveles para estos factores.

EFFECTOS FIJOS VS EFFECTOS ALEATORIOS



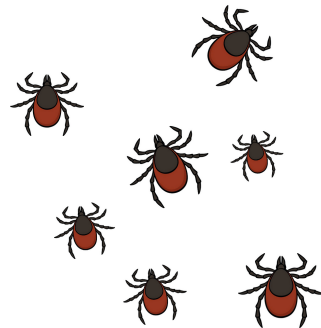
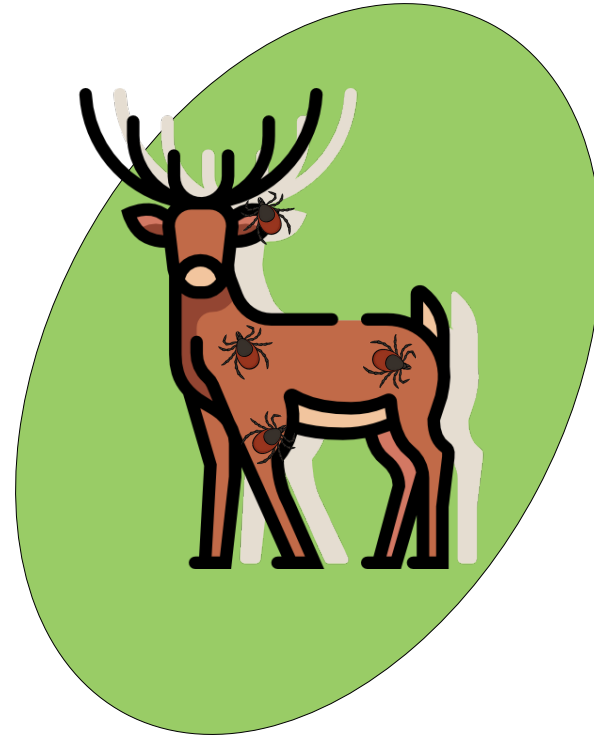
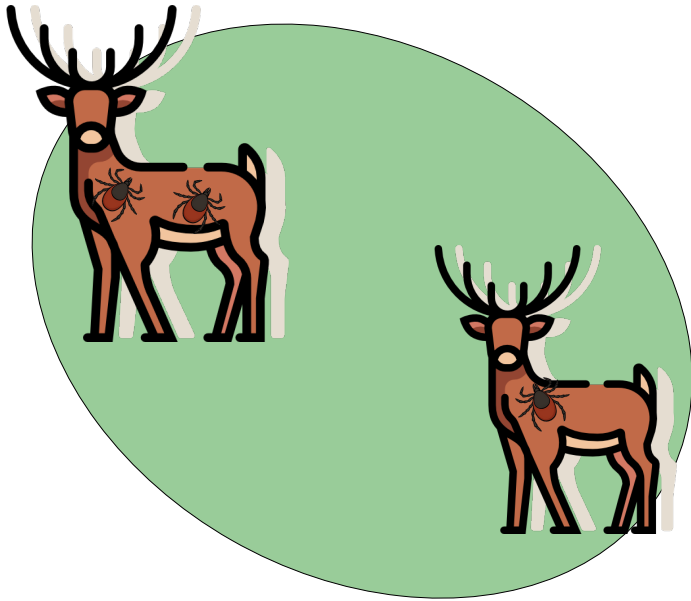
¿Cómo afecta la temperatura a la abundancia de garrapatas en ciervo?

EFFECTOS FIJOS VS EFFECTOS ALEATORIOS



¿Cómo afecta la temperatura a la abundancia de garrapatas en ciervo?

¿Cómo afecta la temperatura a la abundancia de garrapatas en ciervo?



Diferencias entre efectos fijos y efectos aleatorios

Factores de efectos fijos: variables de **interés** específico. Suelen tener un “numero limitado” de niveles. Si volviésemos a obtener una muestra, obtendríamos los mismos niveles.

Factores de efectos aleatorios: capturan la variabilidad de factores que queremos **controlar**, pero que en realidad no estamos interesados en ellos (efectos atribuidos a ser una muestra de una población mayor). Suelen tener “muchos niveles”. Si volviésemos a obtener una muestra, podríamos obtener otros niveles.

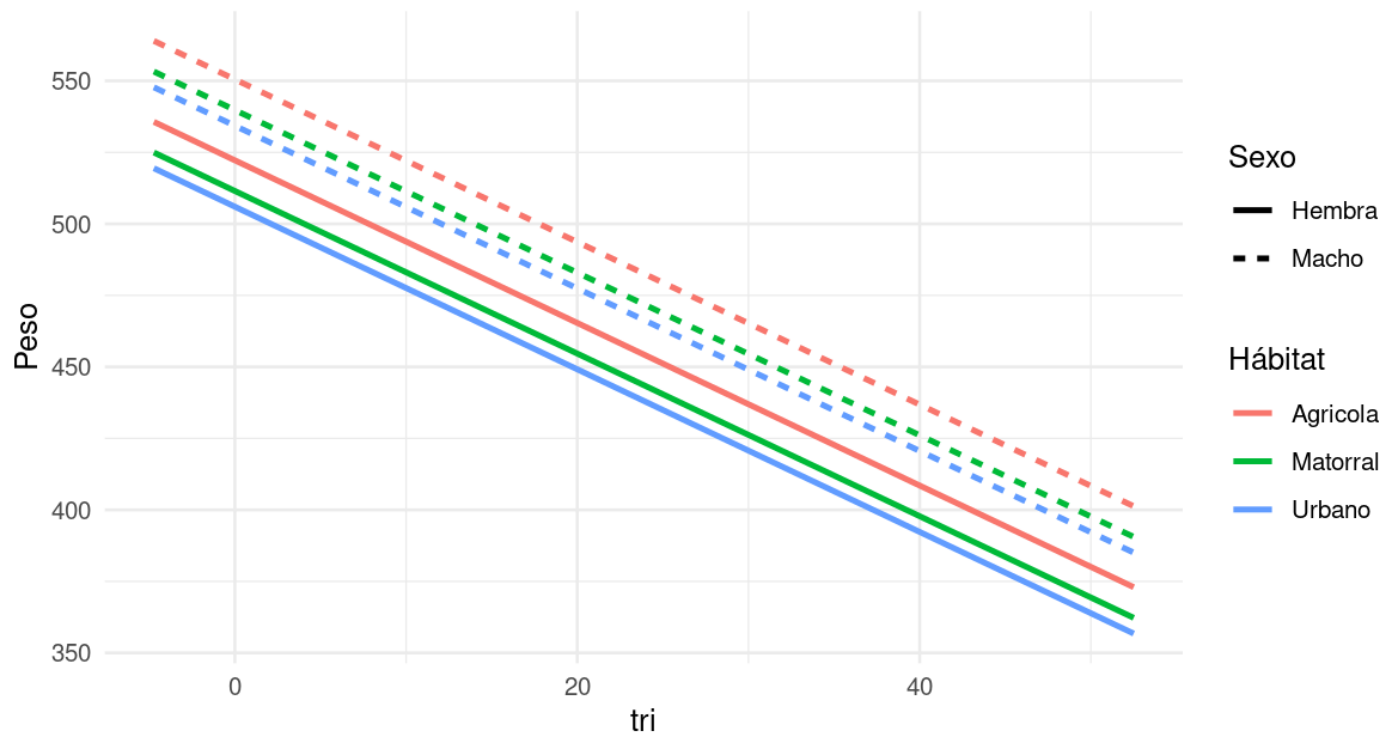
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\text{peso}_i \sim \text{Normal}(\mu_i, \sigma)$$

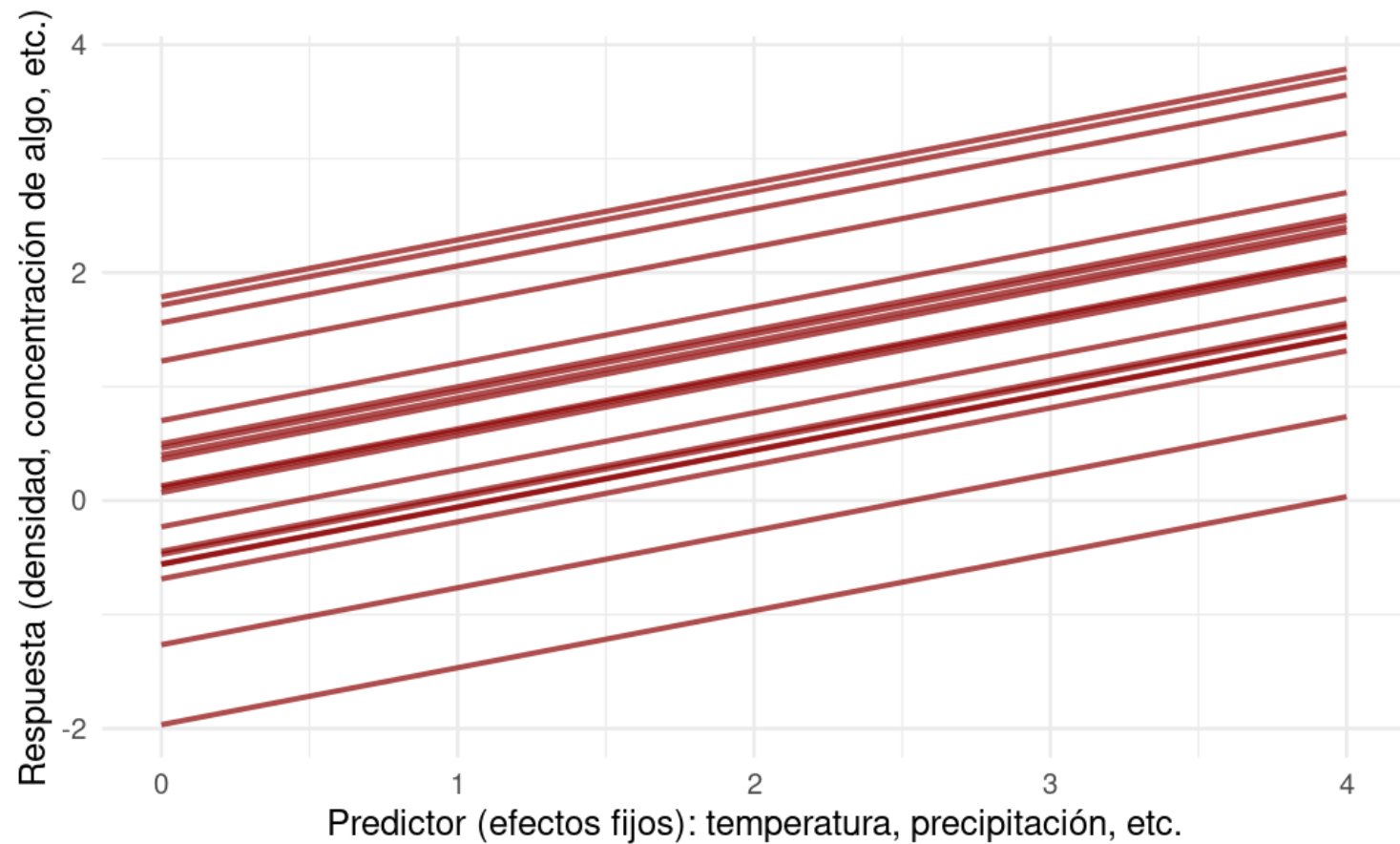
$$\mu_i = \beta_0 + \beta_1 I(\text{sex}_i = \text{Macho}) + \beta_2 I(\text{hab}_i = \text{Matorral}) + \beta_3 I(\text{hab}_i = \text{Urbano}) + \beta_4 \text{tri}_i$$

Paso 3: Misma pendiente para las 6 rectas (modelo: $\text{peso} \sim \text{sex} + \text{hab} + \text{tri}_c$)

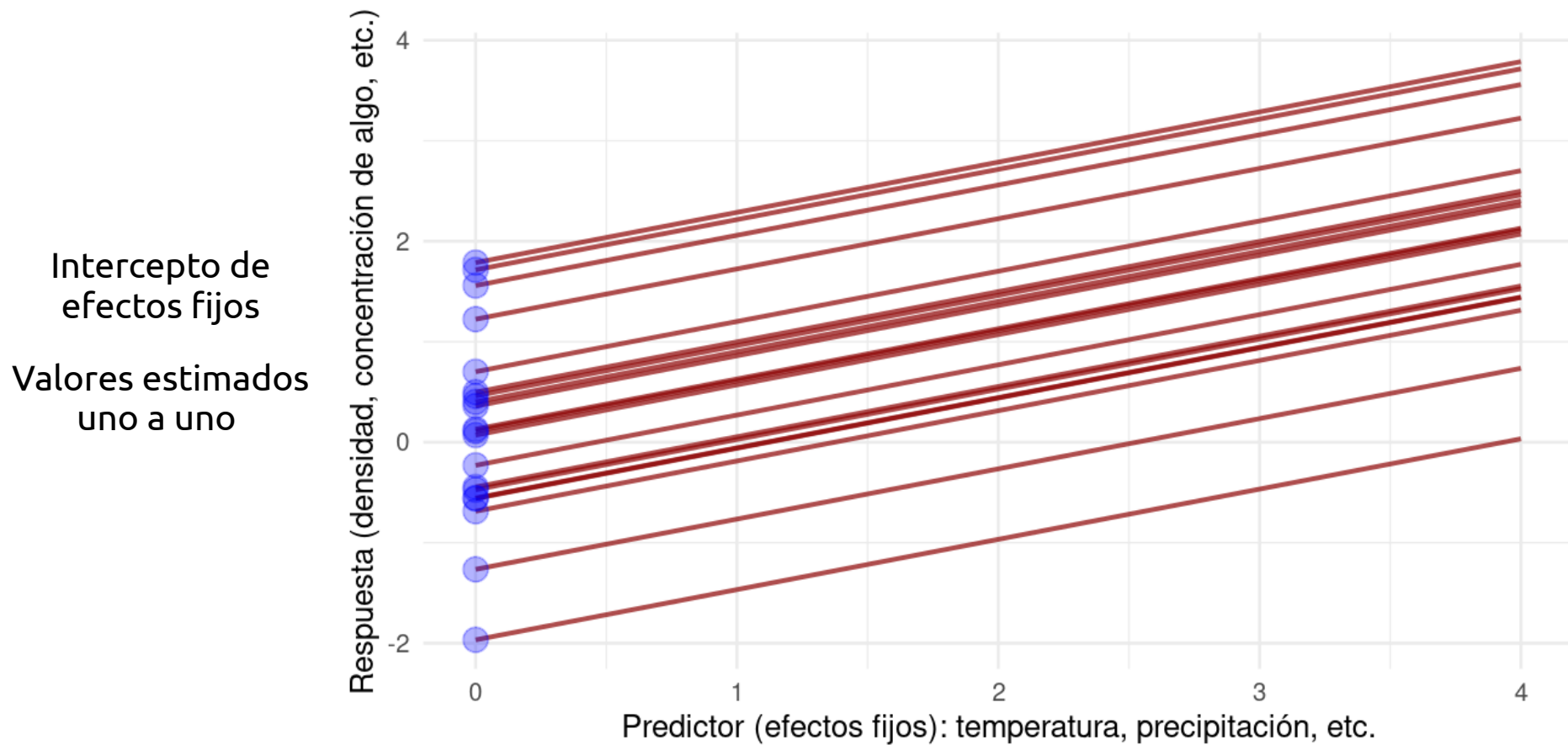
Interceptos distintos por grupo; pendiente común de tri_c



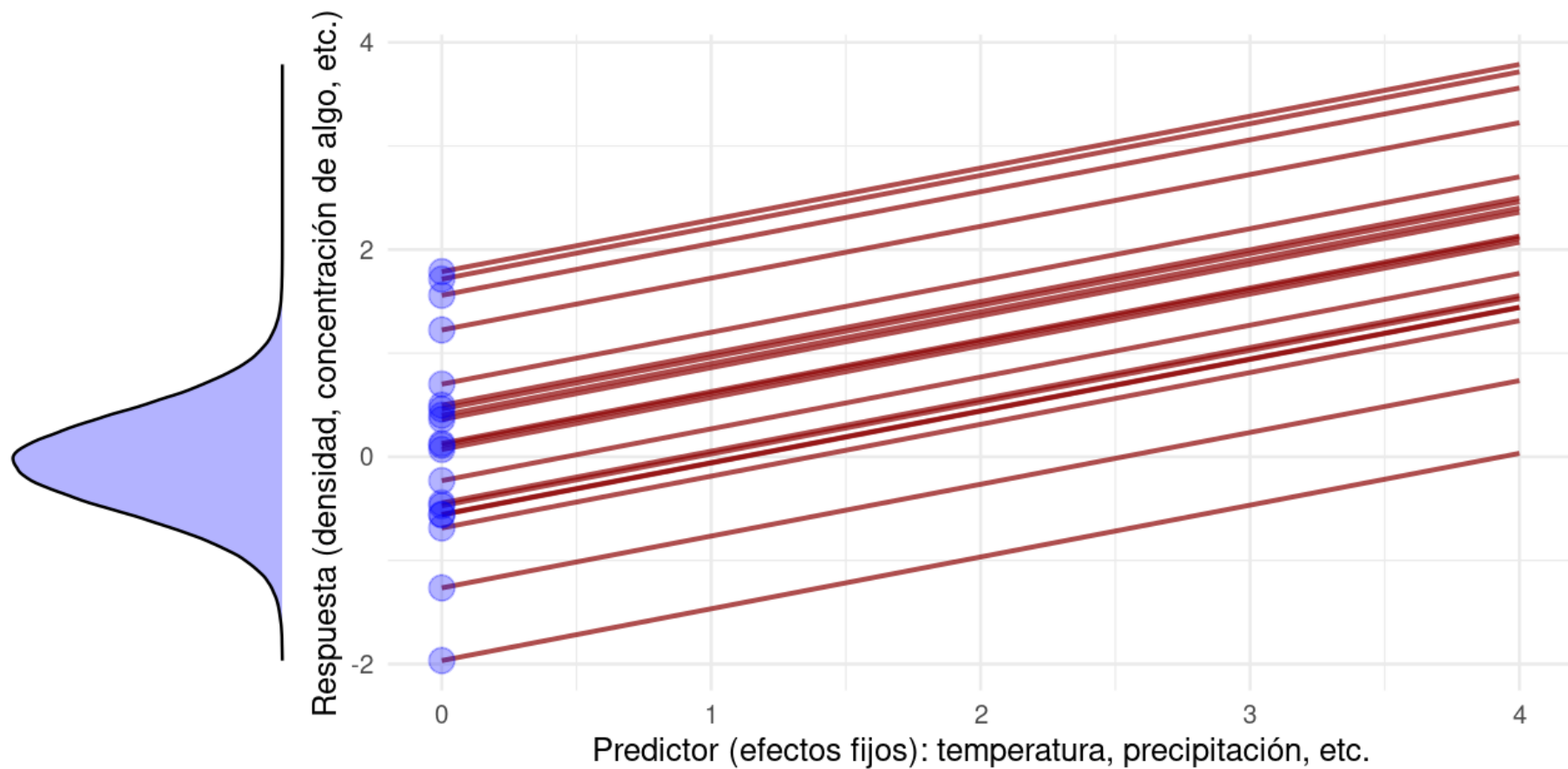
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



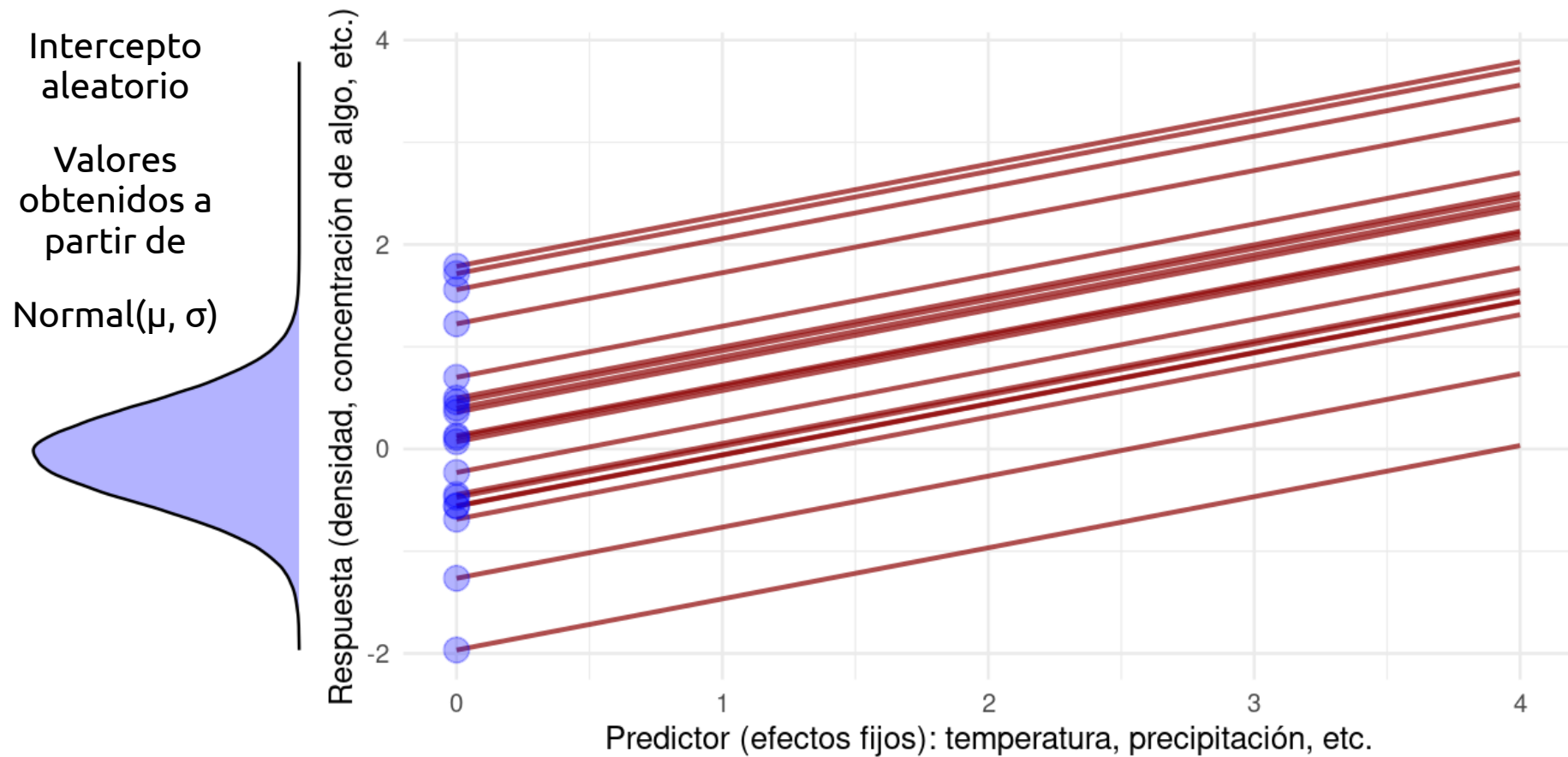
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



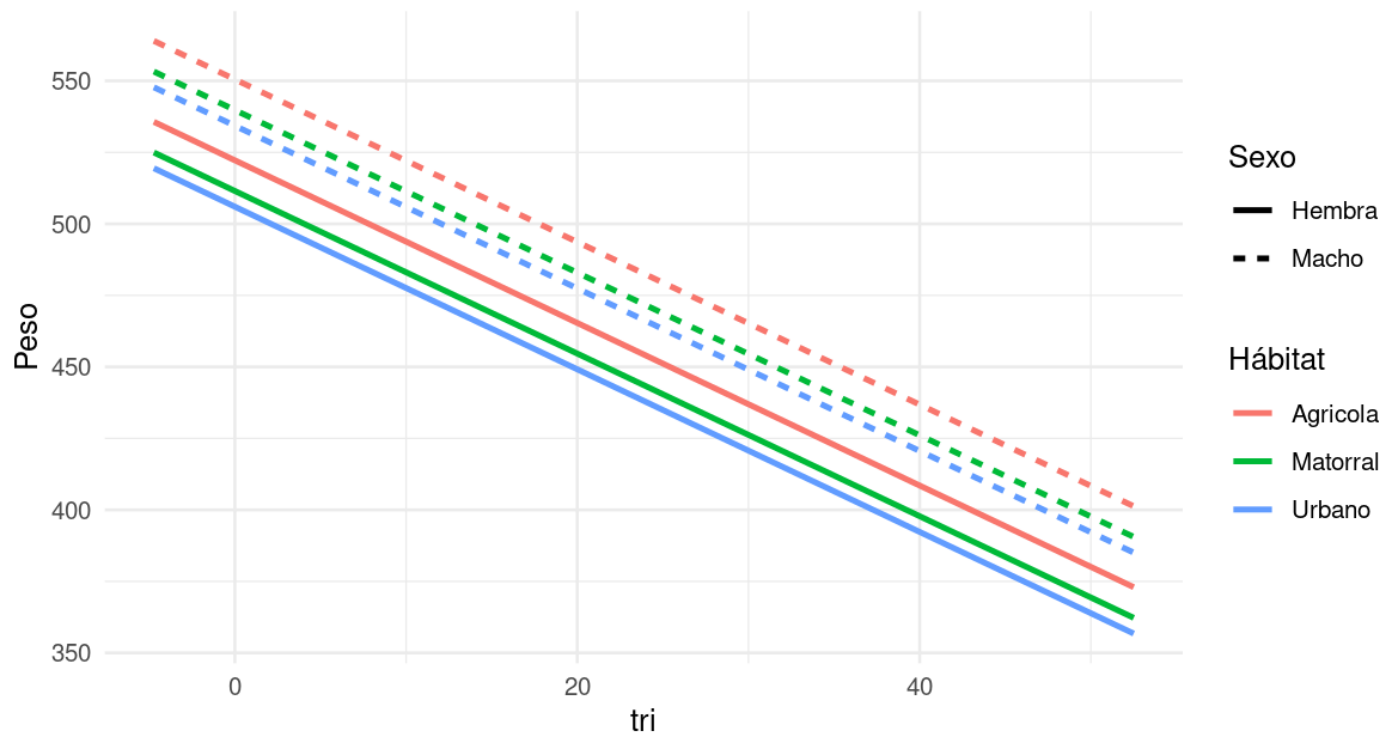
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\text{peso}_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 I(\text{sex}_i = \text{Macho}) + \beta_2 I(\text{hab}_i = \text{Matorral}) + \beta_3 I(\text{hab}_i = \text{Urbano}) + \beta_4 \text{tri}_i$$

Paso 3: Misma pendiente para las 6 rectas (modelo: $\text{peso} \sim \text{sex} + \text{hab} + \text{tri_c}$)

Interceptos distintos por grupo; pendiente común de tri_c



MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\text{n}^\circ \text{ garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{temperatura}_i + \beta_2 \text{cotoB}_i + \beta_3 \text{cotoC}_i + \dots + \beta_{27} \text{cotoZ}_i$$

Si usásemos el coto como un factor de efectos fijos, nuestro modelo tendría que **calcular un parámetro para cada coto de forma independiente...** por lo que necesitaríamos bastantes ciervos dentro de cada coto para tener estimas robustas.

Además, se podría decir que ***“se pierde información”***... estáis seguros de que el número de garrapatas que encontramos en el coto A no se va a parecer en nada a las del coto B, a las del coto C...? En palabras de Richard McElreath, estos modelos ***“tienen amnesia”***!

MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\begin{aligned} \text{n}^\circ \text{ garrapatas}_i &\sim \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \beta_{\text{coto}(j)} + \beta_1 \text{temperatura}_i \\ \beta_{\text{coto}(j)} &\sim \mathcal{N}(0, \sigma_{\text{coto}}) \end{aligned}$$

- $i = \text{ciervo}$
- $j = \text{coto}$

Podemos usar el **coto como un factor aleatorio**. Esto quiere decir que asumimos que todos esos coeficientes provienen de una misma distribución, una normal con media en 0 y una **desviación estándar que es el parámetro que tendremos que estimar**. Esto simplifica el modelo (estimamos menos parámetros), a costa de asumir que esos coeficientes provienen de la misma distribución. Es decir, el modelo asume que **se comparte información** entre los cotos. Estos modelos ***"tienen memoria"***!

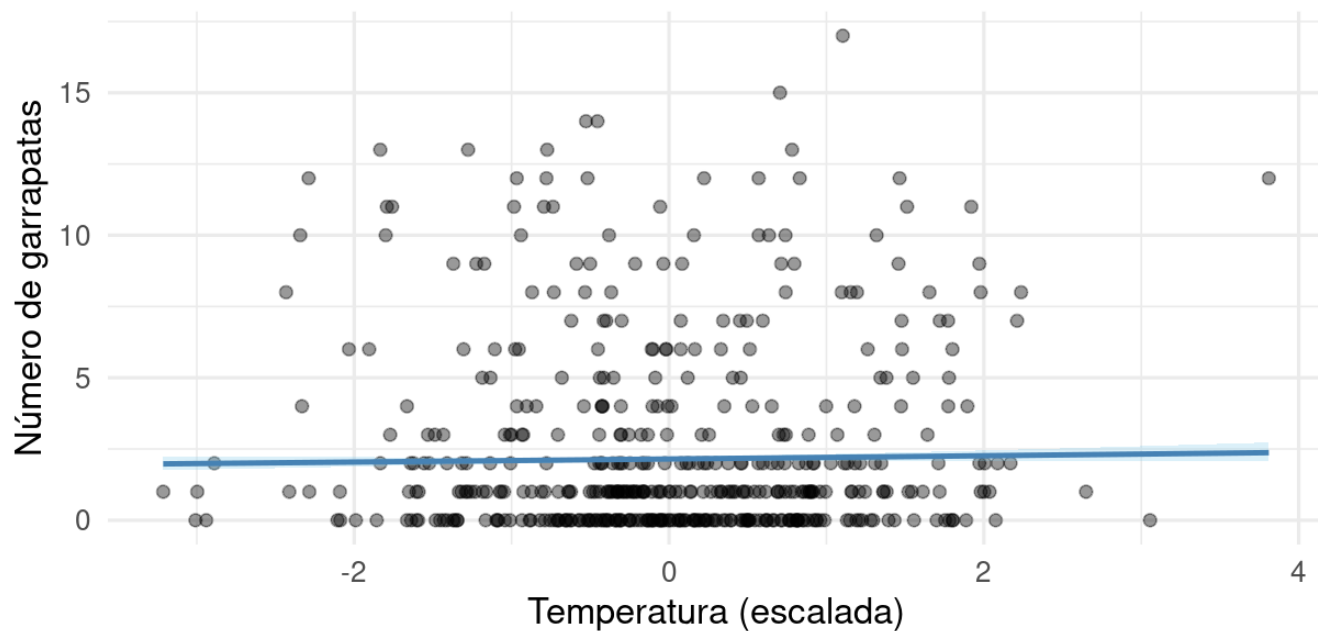
El número **efectivo** de parámetros no es exacto en este tipo de modelos, sino que varía entre 2 y 28.

MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



$$n^{\circ} \text{ garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{ temperatura}_i$$



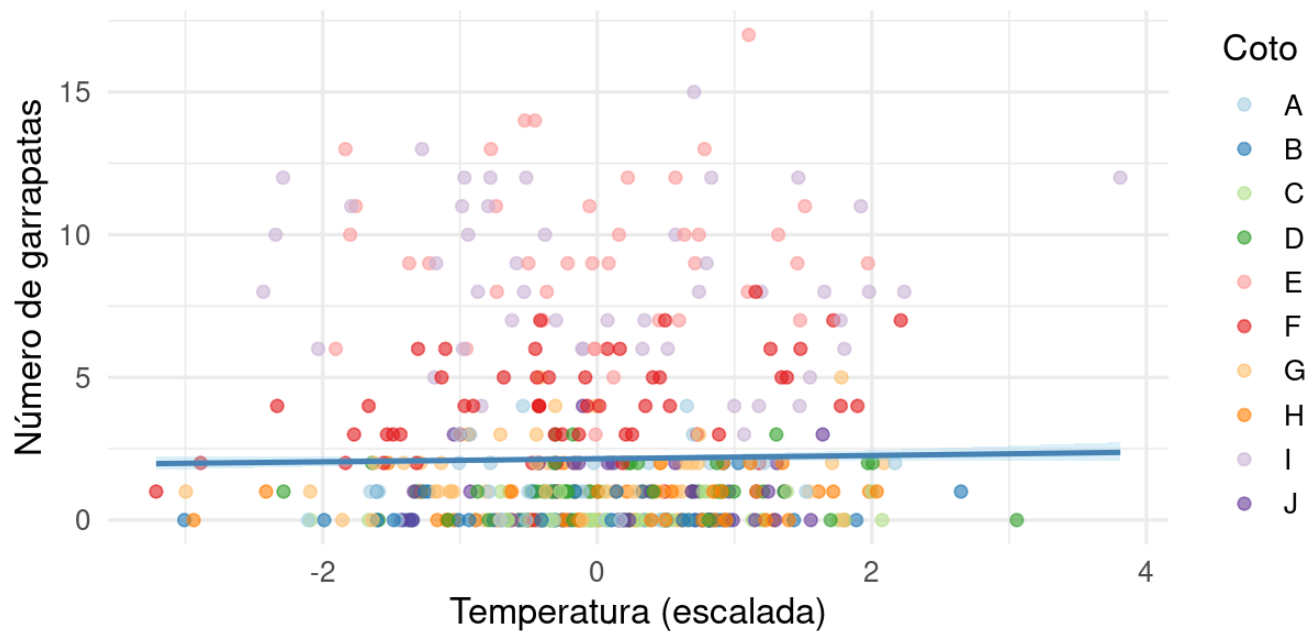
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



$$n^{\circ} \text{ garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{ temperatura}_i$$

?



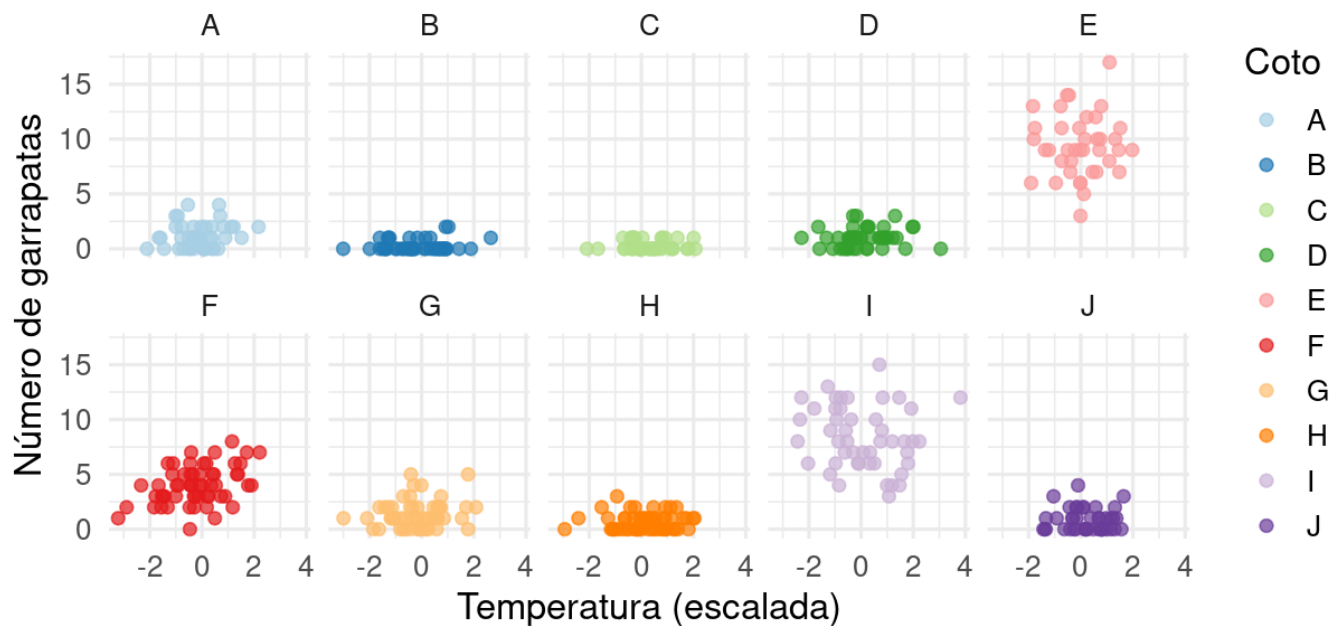
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



$$n^{\circ} \text{ garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{ temperatura}_i$$

?

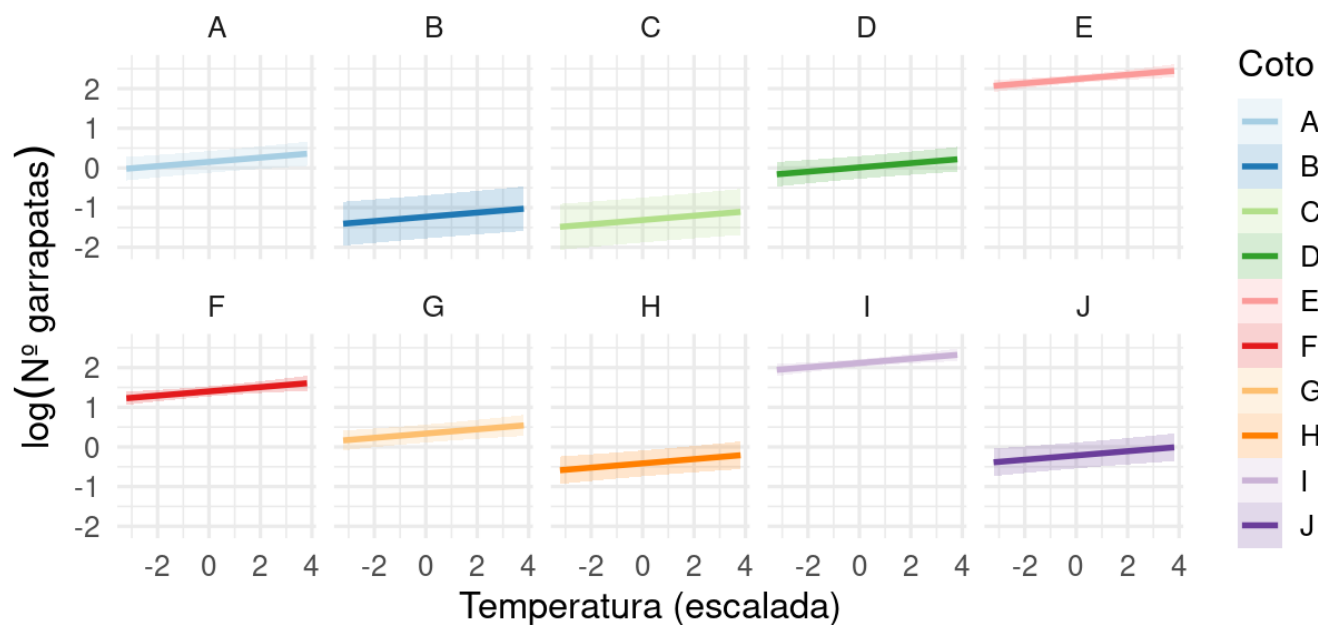


MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



$$n^{\circ} \text{ garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{temperatura}_i + \beta_2 \text{cotoB}_i + \beta_3 \text{cotoC}_i + \dots + \beta_{27} \text{cotoZ}_i$$

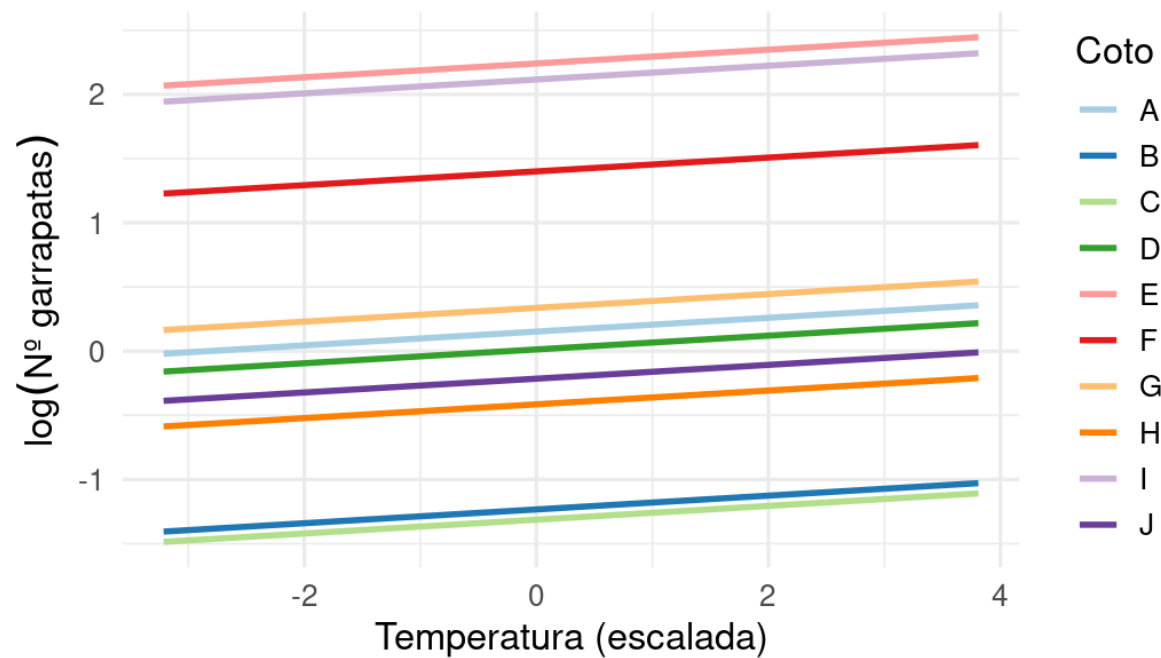


MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



$$n^{\circ} \text{ garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{temperatura}_i + \beta_2 \text{cotoB}_i + \beta_3 \text{cotoC}_i + \dots + \beta_{27} \text{cotoZ}_i$$

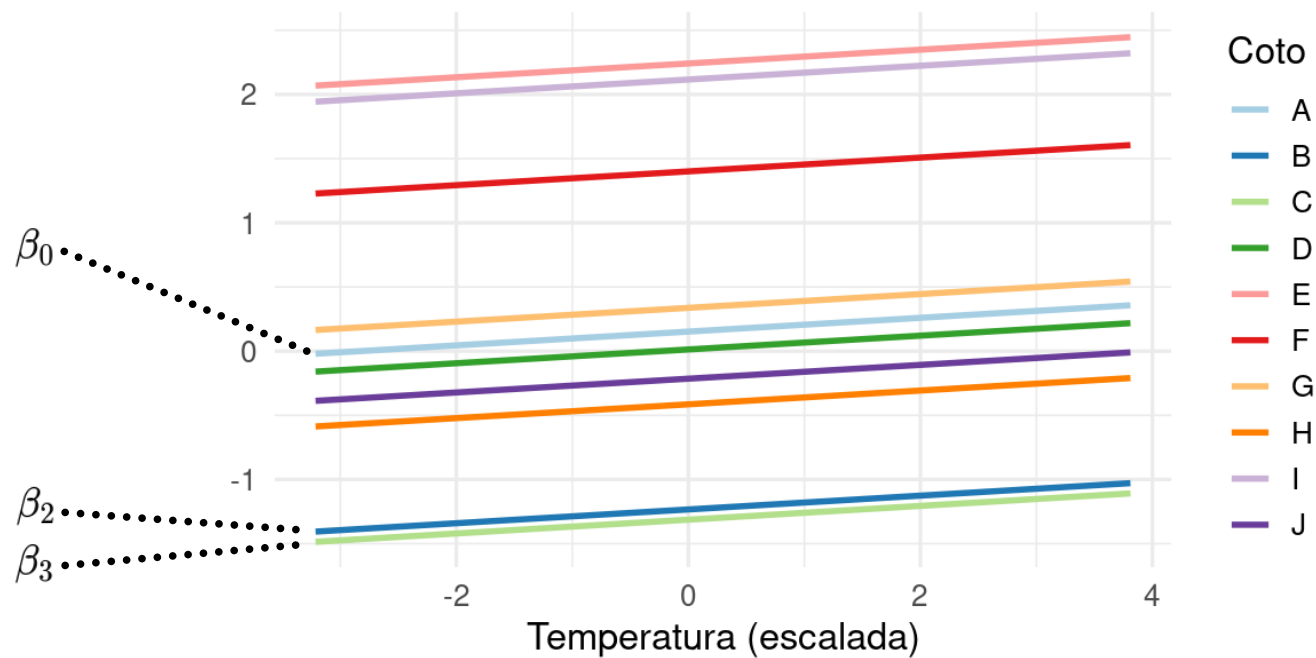


MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



$$n^{\circ} \text{ garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{temperatura}_i + \beta_2 \text{cotoB}_i + \beta_3 \text{cotoC}_i + \dots + \beta_{27} \text{cotoZ}_i$$



MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

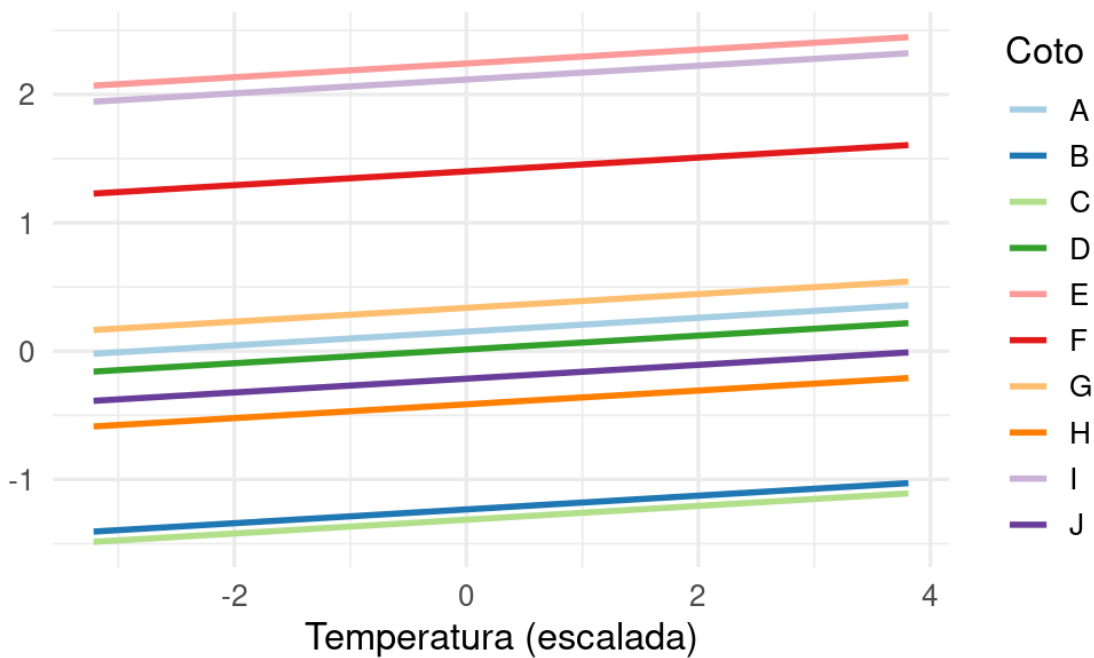


$$n^{\circ} \text{ garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_{\text{coto}(j)} + \beta_1 \text{temperatura}_i$$

$$\beta_{\text{coto}(j)} \sim \mathcal{N}(0, \sigma_{\text{coto}})$$

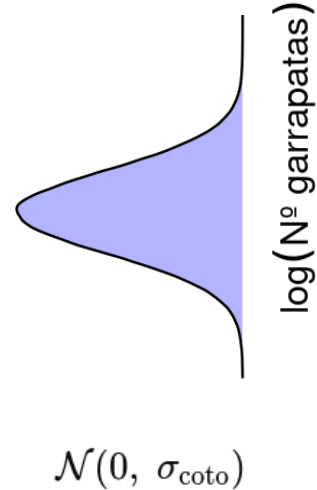
$\log(N^{\circ} \text{ garrapatas})$



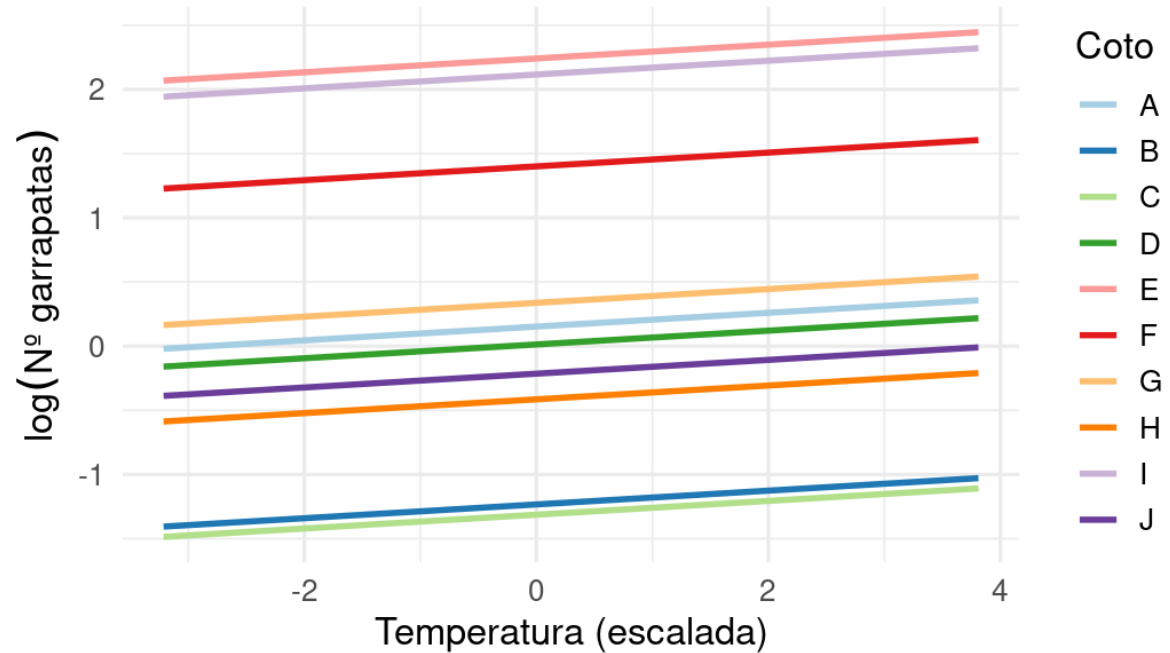
$\mathcal{N}(0, \sigma_{\text{coto}})$

MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

El efecto “*shrinkage to the mean*” (contracción hacia la media global) estabiliza las estimas incluso cuando tenemos pocas muestras por grupo porque “comparte información entre grupos” (tiene memoria!)



$$\begin{aligned} \text{n}^\circ \text{ garrapatas}_i &\sim \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \beta_{\text{coto}(j)} + \beta_1 \text{temperatura}_i \\ \beta_{\text{coto}(j)} &\sim \mathcal{N}(0, \sigma_{\text{coto}}) \end{aligned}$$

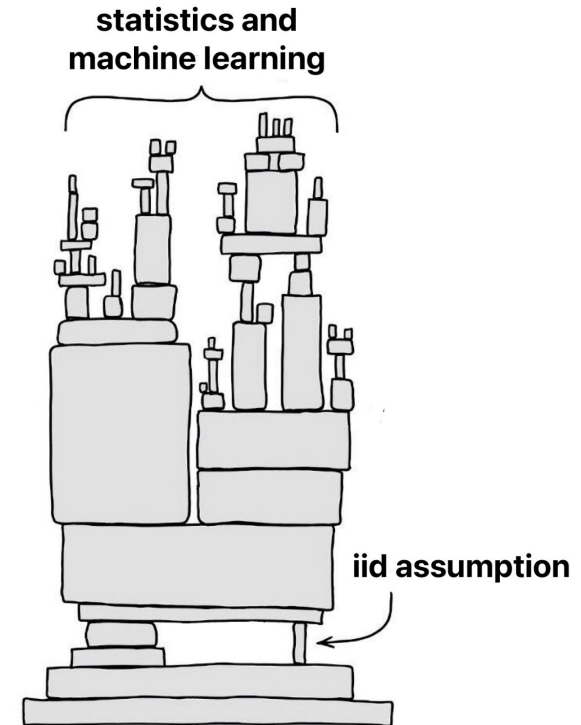


BONDAD DE AJUSTE EN GLM

- En este tipo de modelos, es muy común que la mayor parte de la variación de nuestros datos venga explicado por los efectos aleatorios, sobre todo cuando trabajamos con datos fuera del laboratorio con mucho “ruido”.
- Por eso, es común diferenciar entre la proporción de la varianza explicada por los efectos fijos (en los cuales estamos realmente interesados, la temperatura en nuestro caso, **R^2 marginal**), de la proporción total de varianza explicada por el modelo (efectos fijos + efectos aleatorios, **R^2 condicional**)
- El R^2 de **Nakagawa** sirve a este propósito.

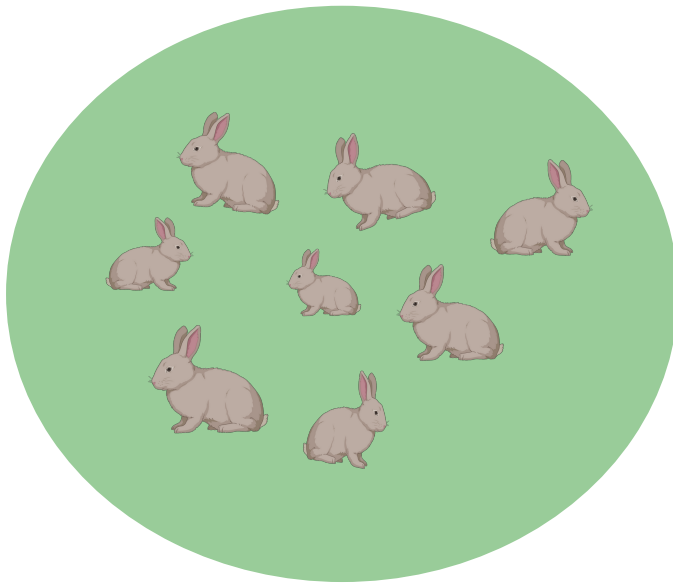
MODELOS MIXTOS Y MEDIDAS REPETIDAS

- La inclusión de efectos aleatorios puede verse como una manera de informar al modelo sobre la estructura de no-independencia en nuestros datos.

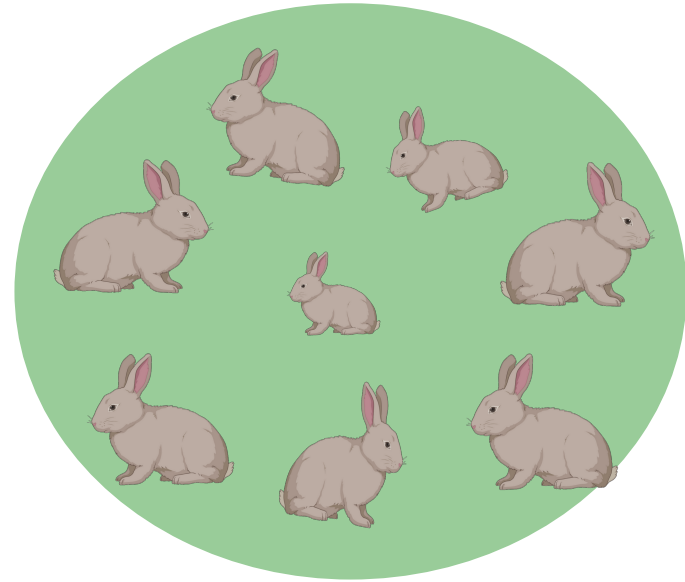


MODELOS MIXTOS Y MEDIDAS REPETIDAS

- La inclusión de efectos aleatorios puede verse como una manera de informar al modelo sobre la estructura de no-independencia en nuestros datos.



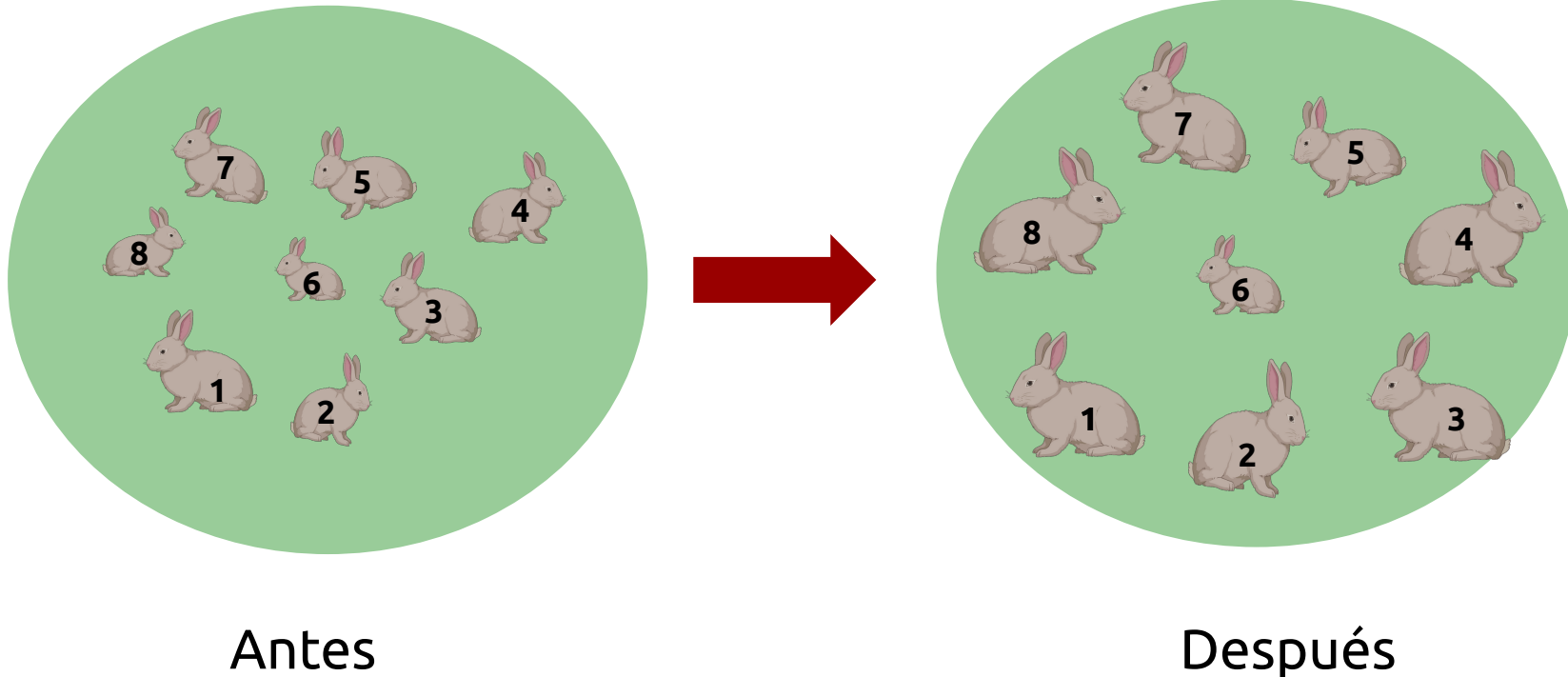
Antes



Después

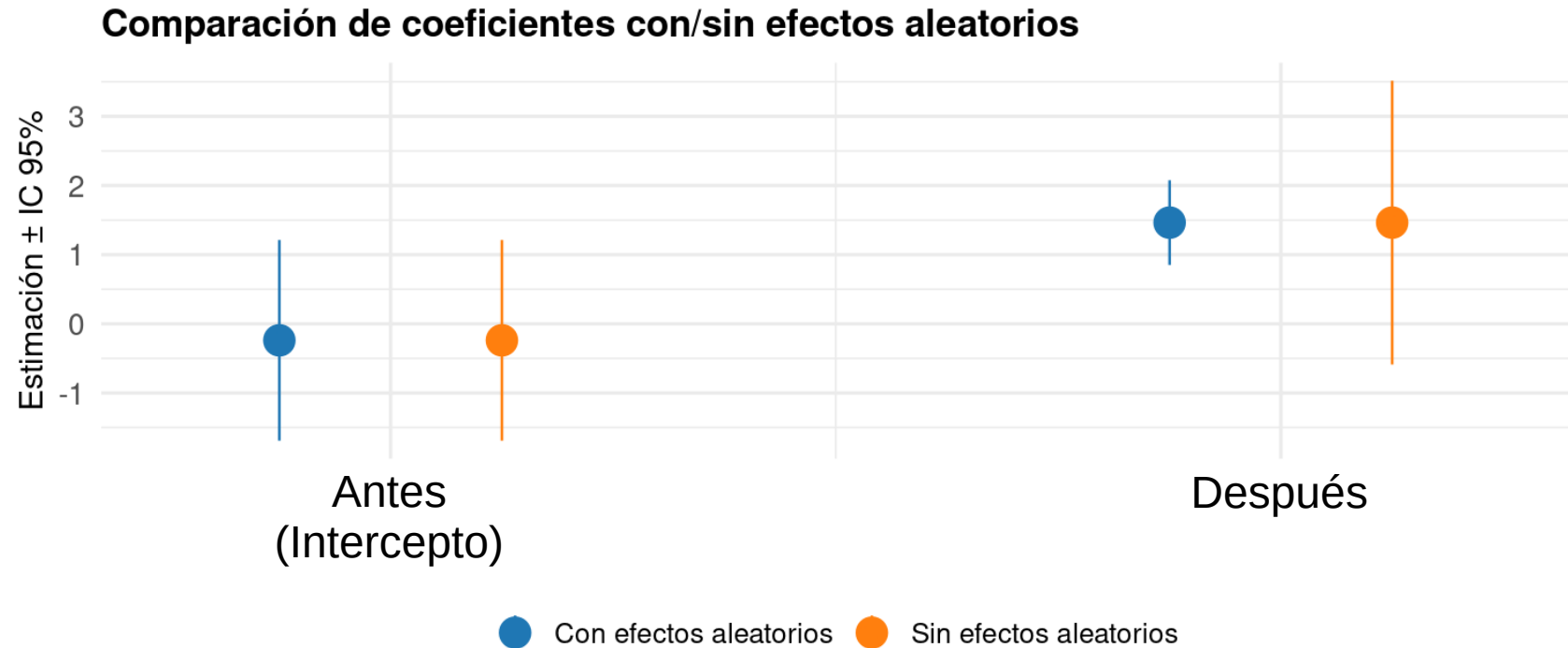
MODELOS MIXTOS Y MEDIDAS REPETIDAS

- La inclusión de efectos aleatorios puede verse como una manera de informar al modelo sobre la estructura de no-independencia en nuestros datos.



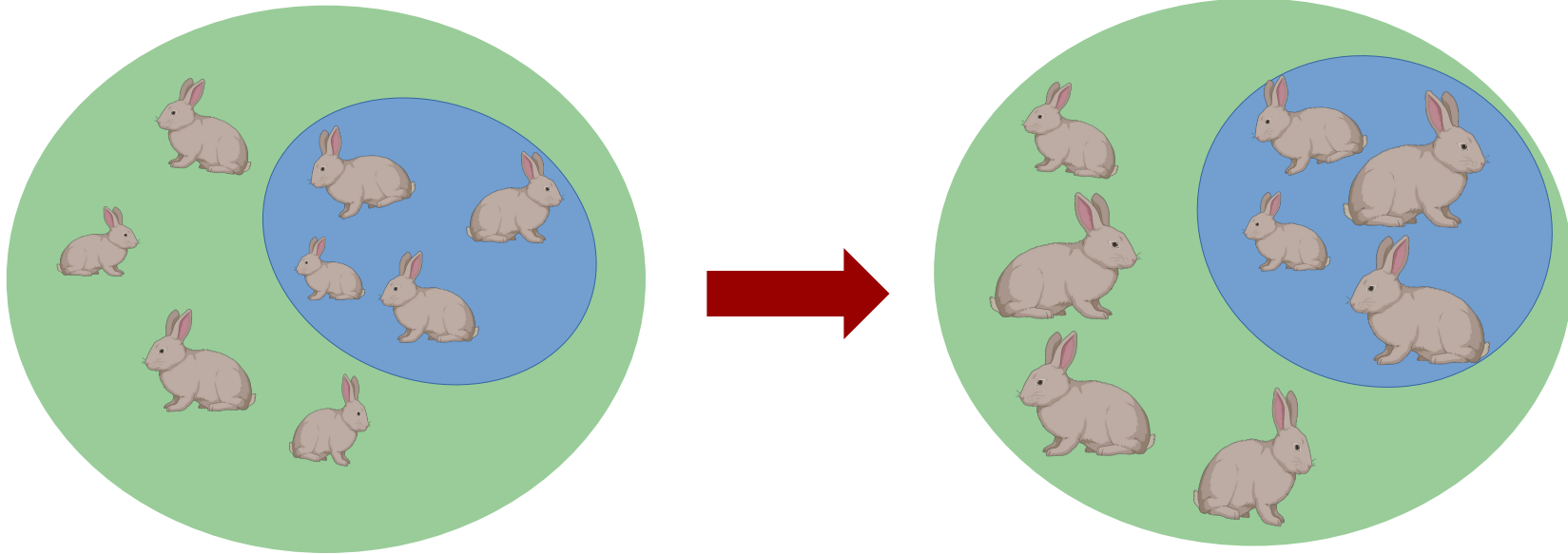
MODELOS MIXTOS Y MEDIDAS REPETIDAS

- La inclusión de efectos aleatorios puede verse como una manera de informar al modelo sobre la estructura de no-independencia en nuestros datos.



MODELOS MIXTOS, MEDIDAS REPETIDAS Y CASO/CONTROL

- Las interacciones entre factores nos ayudan a evaluar si los efectos de un factor varían entre los niveles del otro. Ejemplo: efecto de un fármaco vitamínico



Antes

Después

UN FLUJO DE TRABAJO RECOMENDADO

1) Pensar sobre nuestra realidad/pregunta y establecer hipótesis (fase “expansiva”). Objetivo del modelo ¿Predicción o explicación?

UN FLUJO DE TRABAJO RECOMENDADO

1) Pensar sobre nuestra realidad/pregunta y establecer hipótesis (fase “expansiva”). Objetivo del modelo ¿Predicción o explicación?

2) Identificar y estudiar posibles elementos (variables) del modelo (fase “simplificadora”): rangos, distribuciones de probabilidad, función vínculo... esbozar el álgebra.

respuesta ~ predictores continuos fijos + factores fijos + factores aleatorios...

UN FLUJO DE TRABAJO RECOMENDADO

1) Pensar sobre nuestra realidad/pregunta y establecer hipótesis (fase “expansiva”). Objetivo del modelo ¿Predicción o explicación?

2) Identificar y estudiar posibles elementos (variables) del modelo (fase “simplificadora”): rangos, distribuciones de probabilidad, función vínculo... esbozar el álgebra.

respuesta ~ predictores continuos fijos + factores fijos + factores aleatorios...

3) Estadística descriptiva y visualización de datos (histogramas, dispersión, cajas, etc.). Identificación de problemas, outliers, missing data, tamaño muestral vs predictores, etc.

UN FLUJO DE TRABAJO RECOMENDADO

1) Pensar sobre nuestra realidad/pregunta y establecer hipótesis (fase “expansiva”). Objetivo del modelo ¿Predicción o explicación?

2) Identificar y estudiar posibles elementos (variables) del modelo (fase “simplificadora”): rangos, distribuciones de probabilidad, función vínculo... esbozar el álgebra.

respuesta ~ predictores continuos fijos + factores fijos + factores aleatorios...

3) Estadística descriptiva y visualización de datos (histogramas, dispersión, cajas, etc.). Identificación de problemas, outliers, missing data, tamaño muestral vs predictores, etc.

4) Estudiar colinealidad entre predictores (corrplot, boxplot, posterior al modelo VIF). Toma de decisiones.

UN FLUJO DE TRABAJO RECOMENDADO

1) Pensar sobre nuestra realidad/pregunta y establecer hipótesis (fase “expansiva”). Objetivo del modelo ¿Predicción o explicación?

2) Identificar y estudiar posibles elementos (variables) del modelo (fase “simplificadora”): rangos, distribuciones de probabilidad, función vínculo... esbozar el álgebra.

respuesta ~ predictores continuos fijos + factores fijos + factores aleatorios...

3) Estadística descriptiva y visualización de datos (histogramas, dispersión, cajas, etc.). Identificación de problemas, outliers, missing data, tamaño muestral vs predictores, etc.

4) Estudiar colinealidad entre predictores (corrplot, boxplot, posterior al modelo VIF). Toma de decisiones.

5) Pensar sobre posibles relaciones no lineales. Toma de decisiones.

UN FLUJO DE TRABAJO RECOMENDADO

1) Pensar sobre nuestra realidad/pregunta y establecer hipótesis (fase “expansiva”). Objetivo del modelo ¿Predicción o explicación?

2) Identificar y estudiar posibles elementos (variables) del modelo (fase “simplificadora”): rangos, distribuciones de probabilidad, función vínculo... esbozar el álgebra.

respuesta ~ predictores continuos fijos + factores fijos + factores aleatorios...

3) Estadística descriptiva y visualización de datos (histogramas, dispersión, cajas, etc.). Identificación de problemas, outliers, missing data, tamaño muestral vs predictores, etc.

4) Estudiar colinealidad entre predictores (corrplot, boxplot, posterior al modelo VIF). Toma de decisiones.

5) Pensar sobre posibles relaciones no lineales. Toma de decisiones.

6) Intentar simular una base de datos y ajustar un modelo con datos simulados. Ojo, simplificar mucho.

7) Ajustar el modelo con los datos reales e interpretar resultados: betas, errores y significancia, etc.

- 7) Ajustar el modelo con los datos reales e interpretar resultados: betas, errores y significancia,
- 8) Comprobar asunciones del modelo: normalidad en residuos, homocedasticidad, datos influyentes, independencia, etc.

UN FLUJO DE TRABAJO RECOMENDADO

- 7) Ajustar el modelo con los datos reales e interpretar resultados: betas, errores y significancia,
- 8) Comprobar asunciones del modelo: normalidad en residuos, homocedasticidad, datos influyentes, independencia, etc.
- 9) Evaluar la bondad de ajuste del modelo (observados vs predichos).

UN FLUJO DE TRABAJO RECOMENDADO

- 7) Ajustar el modelo con los datos reales e interpretar resultados: betas, errores y significancia,
- 8) Comprobar asunciones del modelo: normalidad en residuos, homocedasticidad, datos influyentes, independencia, etc.
- 9) Evaluar la bondad de ajuste del modelo (observados vs predichos).
- 10) Evaluar el modelo con datos externos (si se dispone de ellos).

UN FLUJO DE TRABAJO RECOMENDADO

- 7) Ajustar el modelo con los datos reales e interpretar resultados: betas, errores y significancia,
- 8) Comprobar asunciones del modelo: normalidad en residuos, homocedasticidad, datos influyentes, independencia, etc.
- 9) Evaluar la bondad de ajuste del modelo (observados vs predichos).
- 10) Evaluar el modelo con datos externos (si se dispone de ellos).

* Selección de modelos (predictores)

- Debe prevalecer el criterio del profesional (vosotrxs). Tener en cuenta el objetivo del modelo.
- Importante entender las consecuencias de la posible violación de asunciones.
- Correlación no implica causa-efecto.

OTROS TAKE-HOME MESSAGES

- Preparar un "*road-map*" tentativo.
- Intentar simular el proceso, aunque sea de forma muy sencilla o simplificada. Siempre se aprende algo nuevo y se practica el álgebra.
- Siempre, siempre, desde lo más sencillo hacia lo más complejo: número de predictores, estructura del modelo, distribuciones de probabilidad, etc.
- Comunicar el modelo, compartirlo con los compañerxs.
- No será ÉL modelo, será UN modelo.

