El estado de salud de los cultivos ecológicos

E. Lázaro[1], C. Armero[1], J. Roselló[2], J. Serra[2], M.J. Muñoz[3], L. Rubio[2].

[1] Universitat de València
[2] Institut Valencià d'Investigacions Agràries
[3] Laboratorio de Virología del Servicio de Análisis Agroalimentario. Generalitat Valenciana.

I Jornadas Científicas de la SEB
19 y 20 de enero, 2015

Introducción

Objetivo

Datos

Análisis estadístico

Bibliografía

Introducción

- La simplificación de los agrosistemas como consecuencia de la intensificación de la agricultura ha aumentado considerablemente la **vulnerabilidad** de los cultivos a **plagas** y **enfermedades**.
- Las **virosis** son uno de los problemas fitosanitarios más **importantes** pudiendo llegar a provocar una pérdidas en la cosecha de entre el **15 y 65 %**.
- La compleja epidemiología de las virosis exige estudiar el problema desde una **perspectiva agroecológica**.
- El diseño de sistemas agrarios diversificados: sistemas en Producción Integrada y **Agricultura Ecológica**, se plantea como una estrategia para la **estabilización** de las virosis.

Objetivo

Valorar la **prevalencia** de **virosis** en cultivos ecológicos y no ecológicos de hortícolas en la Comunidad Valenciana.

> *"En Rothamsted (Fisher) establece las bases de la matemática estadística, los fundamentos de las técnicas modernas de diseño y análisis de experimentos y desarrolla un gran número de métodos originales como respuesta a los requerimientos de los investigadores que le consultan. Casi todo lo que hoy es la filosofía y la práctica en biometría, proviene de las ideas concebidas por Fisher" (Finney and Yates, 1981).*

Base de datos

- **30** parcelas en producción de hortícolas fueron valoradas en términos de:
  - **Prevalencia de virosis:** detección en 8 plantas por parcela de la presencia o ausencia de **ToMV, CMV y TSWV.**
  - **Sistema de producción:** ecológico y no ecológico.
  - **Caracterización de la parcela:**
    - **Altitud**
    - **Presencia de estructuras de invernadero**

Cantidades de interés y covariables

- **Cantidades de interés:**
  - probabilidad de infección una planta con al menos uno de los tres virus analizados, θ.
  - probabilidad de infección de una parcela para cada uno de los tres virus analizados, θ_1, θ_2 y θ_3.
- **Covariables:**
  - **Factor principal:** sistema de producción **ecológico** ó no ecológico *(Organic)*.
  - **Covariables:**
    - **Altitud de la parcela**, en metros y en escala logarítmica *(Alt)*.
    - Presencia o ausencia de**invernadero** *(Green)*.

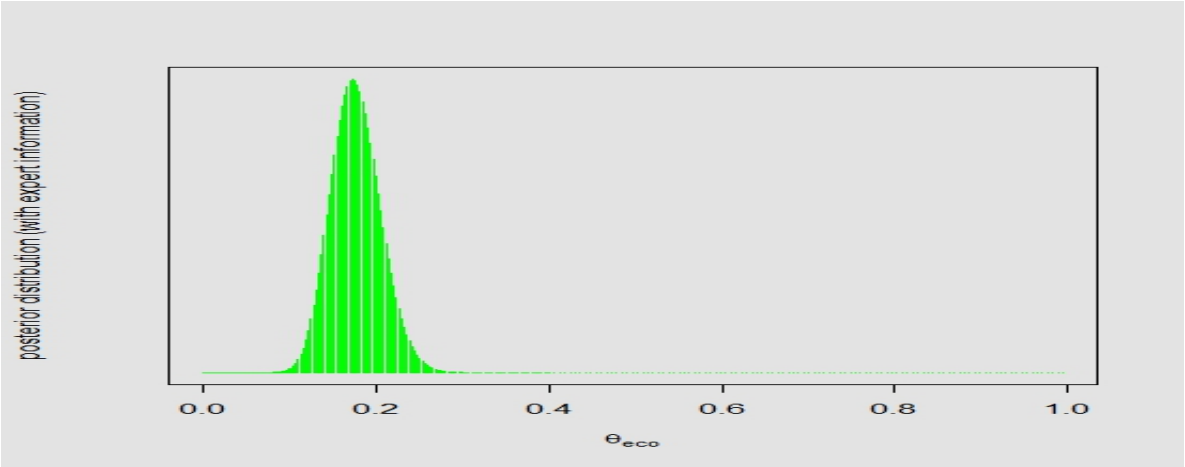
Modelización

Las características específicas de nuestro problema nos han permitido plantear el análisis estadístico como un **proceso de aprendizaje** progresivo de los procedimientos bayesianos:

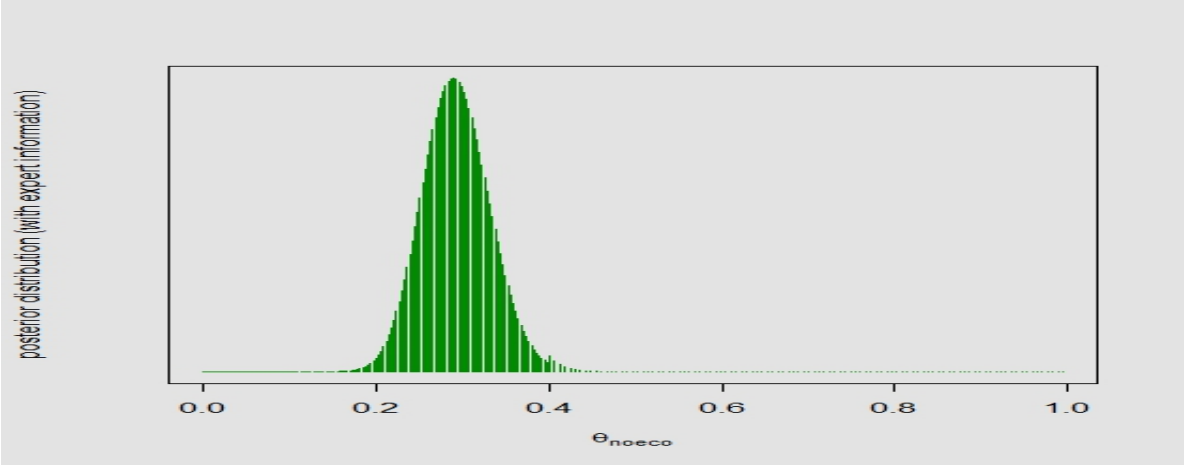
1. Modelo Beta-Binomial.
2. Modelo Lineal Generalizado para datos binomiales **(GLM)**.
3. Modelo Lineal Generalizado **Mixto** para datos binomiales **(GLMM)**.
4. Modelo Lineal Generalizado Mixto para datos binomiales **correlados (GLMM-correlados)**.

Modelo Beta-Binomial

- Distribución posterior cultivos ecológicos



- Distribución posterior cultivos no ecológicos



Modelización

Las características específicas de nuestro problema nos han permitido plantear el análisis estadístico como un **proceso de aprendizaje** progresivo de los procedimientos bayesianos:

1. Modelo Beta-Binomial.
2. **Modelo Lineal Generalizado para datos binomiales (GLM).**
3. Modelo Lineal Generalizado **Mixto** para datos binomiales **(GLMM)**.
4. Modelo Lineal Generalizado Mixto para datos binomiales **correlados (GLMM-correlados)**.

GLM

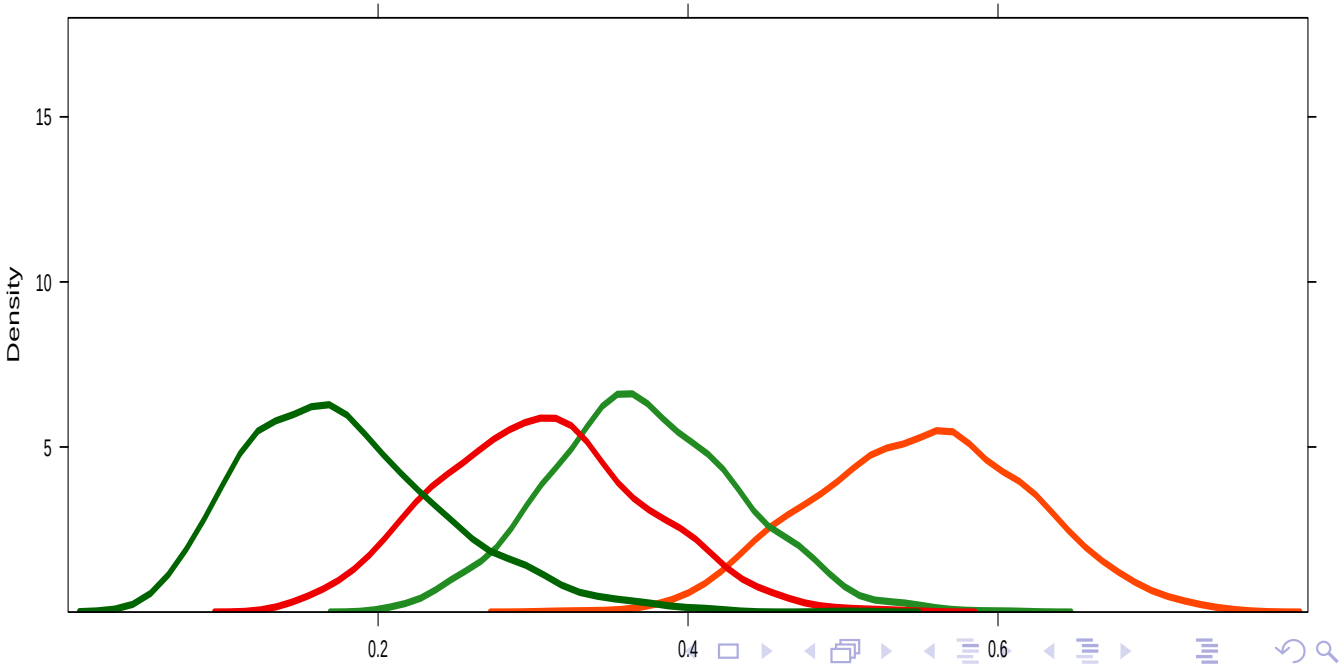
- θ_i: probabilidad de que una planta de la parcela i tenga al menos uno de los tres virus analizados.
- Y_i: número de plantas de la parcela i con al menos uno de los tres virus observados de un total de $n = 8$ analizadas.
- $(Y_i \mid \theta_i) \sim Binomial(n = 8, \theta_i), \;\; i = 1, \ldots, 30$
- $logit(\theta_i) = \beta_0 + \beta_1 \, I_{Organic(i)} + \beta_2 \; I_{Green(i)} + \beta_3 \, log(alt_i) + \beta_4 \, I_{Organic(i)} * log(alt_i)$
  - $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ son los coeficientes de regresión.
  - $I_{C(i)}$ es una variable indicadora que toma el valor 1 si la parcela i tiene la característica C y por el contrario, toma el valor 0 si la parcela no la tiene.
- Distribución previa, $\pi(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$
  - **Independencia** a priori
  - Distribución **no informativa**.

Resumen distribución a posteriori

Parámetros	Media	Desv.típica	$Q_{0,025}$	$Q_{0,5}$	$Q_{0,975}$
β_0	2.116	0.783	0.675	2.113	3.698
$\beta_1(organic)$	0.721	1.288	-1.779	0.689	3.397
$\beta_2(green)$	-1.083	0.418	-1.888	-1.075	-0.250
$\beta_3(log(alt))$	-0.538	0.187	-0.914	-0.535	-0.184
$\beta_4(log(alt) * organic)$	-0.416	0.309	-1.072	-0.410	0.136

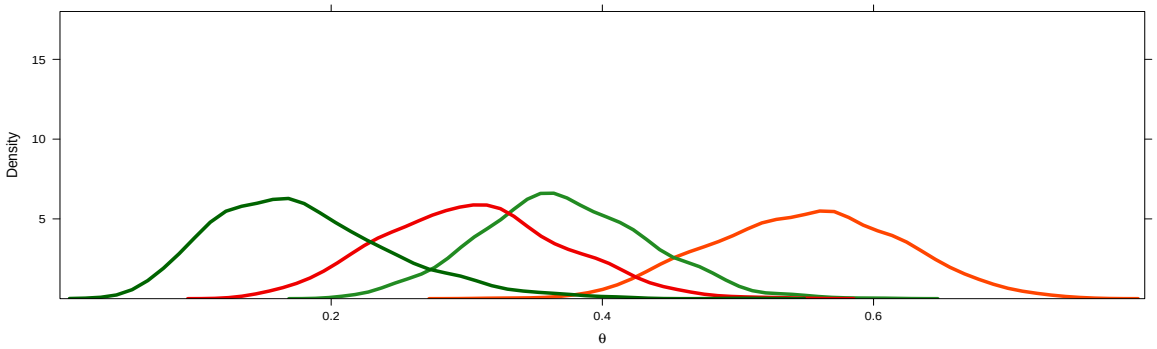
- ECO-INV
- ECO-NO INV
- NO ECO-INV
- NO ECO-NO INV

Altitud: 34 m

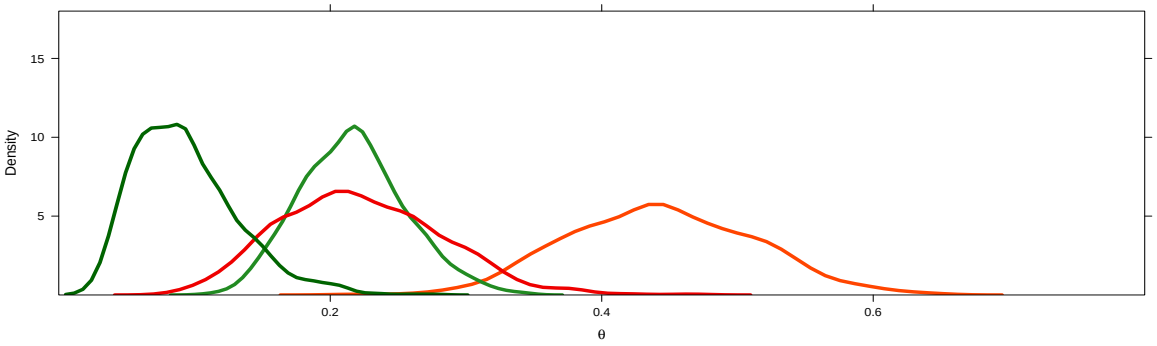


Distribución a posteriori de la probabilidad de infección de una planta

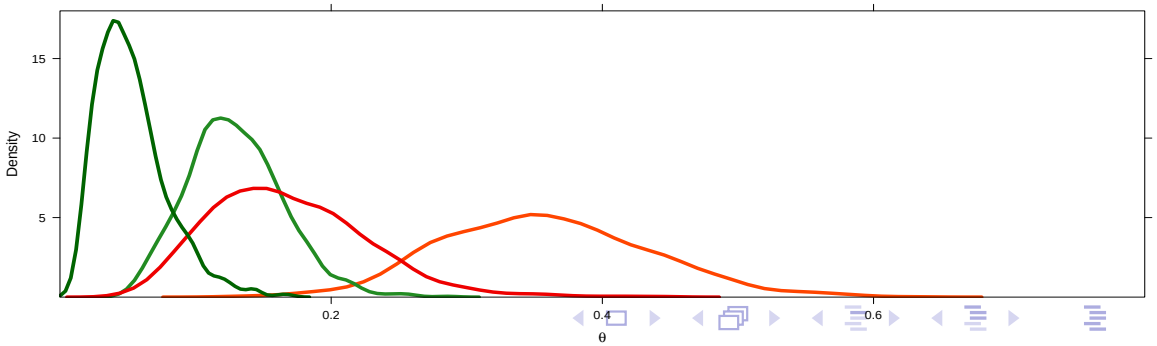
- Altitud: 34 m



- Altitud: 76 m

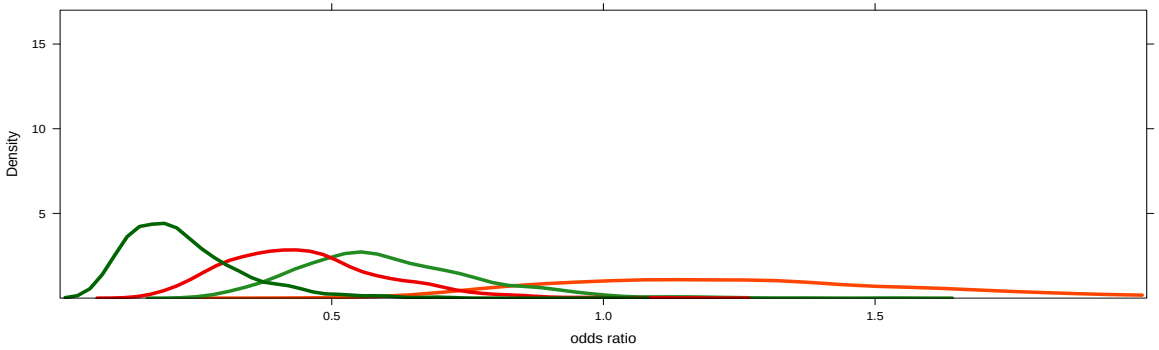


- Altitud: 151 m

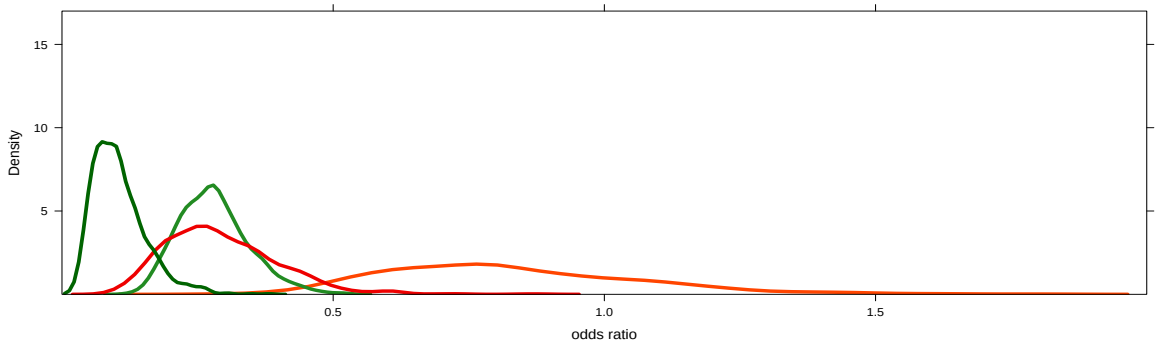


Distribución a posteriori de los odds ratio

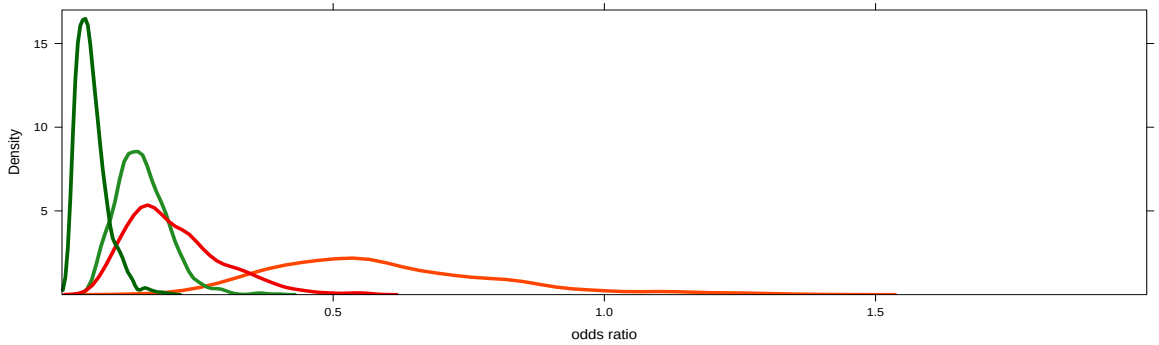
- **Altitud: 34 m**



- **Altitud: 76 m**



- **Altitud: 151 m**



Modelización

Las características específicas de nuestro problema nos han permitido plantear el análisis estadístico como un **proceso de aprendizaje** progresivo de los procedimientos bayesianos:

1. Modelo Beta-Binomial.
2. Modelo Lineal Generalizado para datos binomiales **(GLM)**.
3. **Modelo Lineal Generalizado Mixto para datos binomiales (GLMM)**.
4. Modelo Lineal Generalizado Mixto para datos binomiales **correlados (GLMM-correlados)**.

GLMM

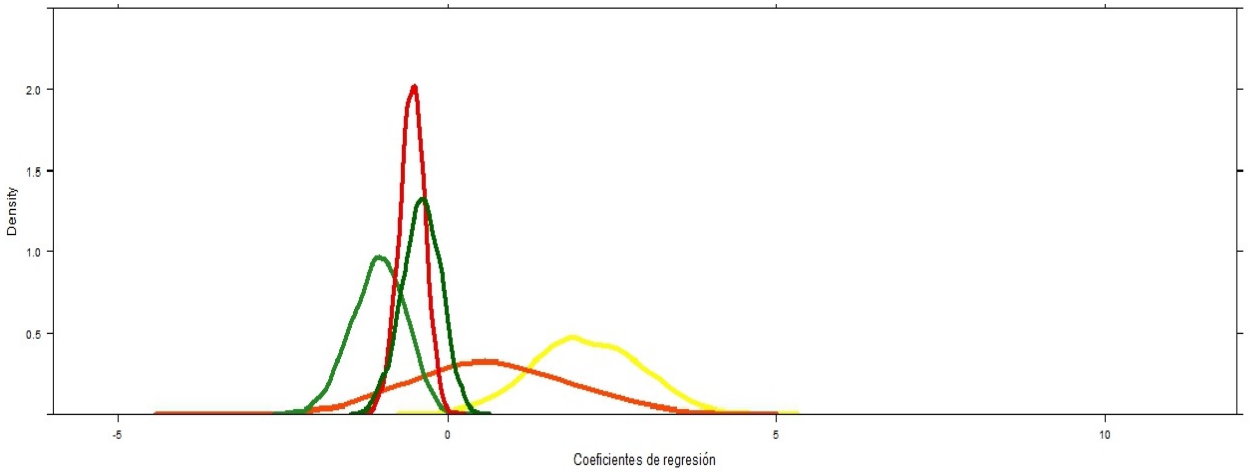
- θ_i: probabilidad de que una planta de la parcela i tenga al menos uno de los tres virus analizados.
- Y_i: número de plantas de la parcela i con al menos uno de los tres virus observados de un total de $n = 8$ analizadas.
- $(Y_i \mid \theta_i) \sim Binomial(n = 8, \theta_i), \;\; i = 1, \ldots, 30$
- $logit(\theta_i) = \beta_0 + \beta_1 \, I_{Organic(i)} + \beta_2 \; I_{Green(i)} + \beta_3 \, log(alt_i) + \beta_4 \, I_{Organic(i)} * log(alt_i) + b_i$
  - $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ son los coeficientes de regresión.
  - $I_{C(i)}$ es una variable indicadora que toma el valor 1 si la parcela i tiene la característica C y por el contrario, toma el valor 0 si la parcela no la tiene.
  - $(b_i \mid \sigma_b^2) \sim \mathbf{N}(0, \sigma_b^2)$ **que es el efecto aleatorio independiente asociado a la parcela** i
- Distribución previa, $\pi(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \sigma_b)$
  - **Independencia** a priori.
  - Distribución **no informativa**.

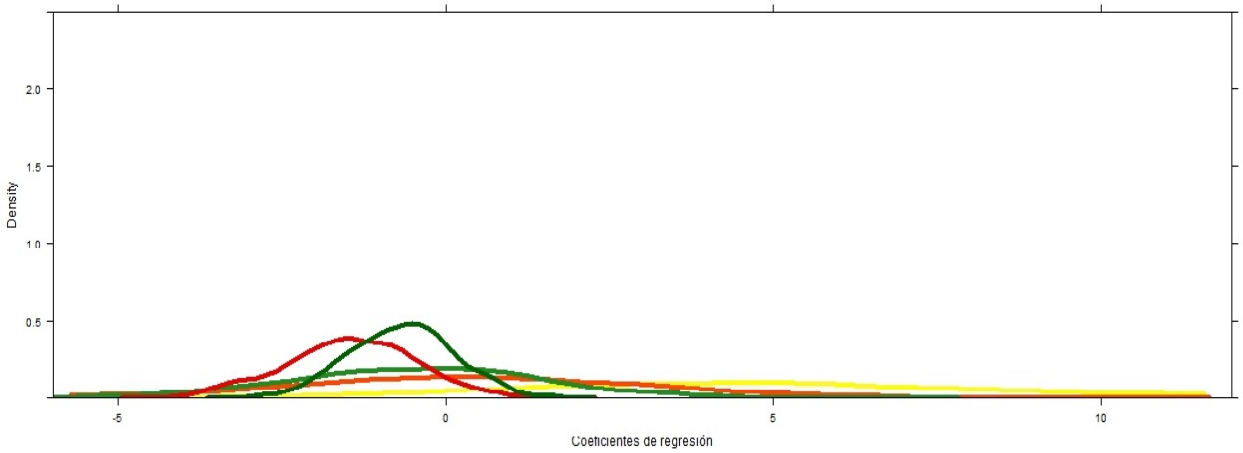
Resumen distribución a posteriori

Parámetros	Media	Desv.típica	$Q_{0.025}$	$Q_{0.5}$	$Q_{0,975}$
β_0	4.789	4.037	-2.816	4.718	13.046
$\beta_1(organic)$	0.357	2.888	-5.479	0.309	6.147
$\beta_2(green)$	-0.436	2.030	-4.521	-0.369	3.482
$\beta_3(log(alt))$	-1.444	1.034	-3.531	-1.426	0.609
$\beta_4(log(alt) * organic)$	-0.669	0.818	-2.305	-0.649	0.895
σ_b	4.889	0.747	3.266	4.973	5.933

Distribución a posteriori GLM vs. GLMM

- β_0
- $\beta_1(organic)$
- $\beta_2(green)$
- $\beta_3(log(alt)$
- $\beta_4(log(alt) * organic)$





Modelización

Las características específicas de nuestro problema nos han permitido plantear el análisis estadístico como un **proceso de aprendizaje** progresivo de los procedimientos bayesianos:

1. Modelo Beta-Binomial.
2. Modelo Lineal Generalizado para datos binomiales **(GLM)**.
3. Modelo Lineal Generalizado **Mixto** para datos binomiales **(GLMM)**.
4. **Modelo Lineal Generalizado Mixto para datos binomiales correlados (GLMM-correlados).**

GLMM correlados

- θ_{ij}: **probabilidad de que alguna planta de la parcela i tenga el virus j.**
- Y_{ij}: **presencia de alguna planta en la parcela parcela i con el virus j.**
- El **subíndice i** hace referencia al número de parcelas.
- El **subíndice j** hace referencia al virus analizado, correspondiéndose:
  - $j = 1$, con el **ToMV**
  - $j = 2$, con el **CMV**
  - $j = 3$, con el **TSWV**
- $(Y_{ij} \mid \theta_{ij}) \sim Bernoulli(\theta_{ij}),\ i = 1, \ldots, 30\ j = 1, \ldots, 3$
- $logit(\theta_{ij}) = \beta_j + b_i$
  - $(b_i \mid \sigma_b^2) \sim \mathbf{N}(0, \sigma_b^2)$ es el efecto aleatorio independiente asociado a la parcela i y hace referencia a la susceptibilidad de la parcela i frente a las virosis j
  - β_j es el coeficiente de regresión y hace referencia a la tasa de frecuencia del virus j.
- Distribución previa, $\pi(\beta_1, \beta_2, \beta_3, \sigma_b)$
  - **Independencia** a priori.
  - Distribución **no informativa**.

Albert, J. (2007). *Bayesian computation with R*. Springer.

Agresti A. (2002). *Categorical data analysis.* John Wiley Sons, New York.

Gelman, A. Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.

Kéry, M. (2010). *Introduction to WinBUGS for Ecologists: A Bayesian Approach to Regression, ANOVA and Related Analyses*. Academic Press.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Wiley Series in Computational Statistics, Hoboken, USA.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology, 321-334. En *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV*, University of California Press, Berkeley, California.