

MODELOS NO LINEALES MIXTOS CON VARIABLES DE RESPUESTA CON DISTRIBUCIÓN BETA

Por

Yuri Marcela García Saavedra

Tesis sometida en cumplimiento parcial de los requerimientos para el grado de

MAESTRÍA EN CIENCIAS

en

MATEMÁTICAS (ESTADÍSTICA)

UNIVERSIDAD DE PUERTO RICO
RECINTO UNIVERSITARIO DE MAYAGÜEZ

2012

Aprobada por:

Dámaris Santana Morant, Ph.D.
Miembro, Comité Graduado

Fecha

Edgar Acuña, Ph.D.
Miembro, Comité Graduado

Fecha

Raúl E. Macchiavelli, Ph.D.
Presidente, Comité Graduado

Fecha

María Y. Canabal Torres, Ph.D.
Representante, Estudios Graduados

Fecha

Omar Colón, Ph.D.
Director del Departamento

Fecha

A Dios, a la virgencita, a mis padres Aracely y Antonio, a mis hermanos y sobrinos.
Ustedes son mi mayor inspiración y apoyo.

Agradecimientos

A Dios por siempre estar a mi lado y darme la fuerza, sabiduría y perseverancia.

Al Dr. Raúl Macchiavelli por sus valiosos aportes, apoyo y paciencia no sólo en la culminación de esta tesis sino en el transcurso de toda la maestría.

A los miembros del comité, Dr. Edgar Acuña y Dra. Dámaris Santana por su tiempo y aportes.

A los profesores del departamento de Ciencias Matemáticas por brindarme el espacio y las herramientas para llevar a cabo el fin de este trabajo.

A los profesores del departamento de Ciencias Agroambientales y de Cultivos, Dra. Linda Beaver por brindarme el espacio durante sus clases para el desarrollo de las simulaciones y Dr. José Chavarría por facilitarme los datos para las aplicaciones de la tesis.

Al Dr. Fernando Casanoves por sus constantes palabras de ánimo, apoyo incondicional, confianza e importantes aportes a mi formación académica y personal.

Al Dr. Jairo Armando Cardona por su valiosa y linda amistad, apoyo y confianza incondicional.

A mi gran amigo y compañero de batalla William Camacho por sus consejos, palabras de ánimo, confianza, paciencia y aportes a mi formación académica y personal .

A mis amigos y compañeros en Puerto Rico en especial a Glorimar, Geisel, Jairo, Yovani, Andrea, Cesar, Javier, Hector por los momentos lindos y especiales que vivimos, por el apoyo, confianza y la famosa frase “éxitos con tu tesis”.

A todos aquellos que confiaron en mi, incluso aquellos que me acompañan desde el cielo, mil gracias, su apoyo y buenos deseos fueron clave en la finalización de esta etapa.

Abstract of Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of Master of Science
NONLINEAR MIXED MODELS WITH RESPONSE VARIABLES
WITH BETA DISTRIBUTION

By

Yuri Marcela García Saavedra

2012

Adviser: Raúl E. Macchiavelli

Department: Mathematical Sciences

There are many situations in which random variables involving some kind of percentage, ratio, or fraction. Many studies have shown that linear regression models are not appropriate to model this type of data. However, the beta distribution is very useful for modeling data that are continuous and restricted to the interval $(0,1)$, and also can be explained by other variables through a regression structure. Due to this, Ferrari and Cribari-Neto (2004) proposed a beta linear regression model in which the response variable is distributed as a beta using a different parameterization of its density function, thus obtaining a regression structure for the mean of the response with a constant precision parameter. In this study we extend this theory, proposing a beta nonlinear mixed regression model, where the conditional distribution of observations is assumed beta and the distribution of the random effects is assumed normal. We study the induced marginal distribution and the properties of the model proposed by means of simulation, which are compared with those obtained by a nonlinear mixed regression model assuming normal distribution. The estimates are obtained using the maximum likelihood, until quasi-Newton technique for the optimization and Gaussian quadrature for the integration. Standard errors of the model parameters were estimated using the Hessian matrix. Finally, we apply these results to studies of disease severity (relative amount of affected tissue in a given time) in plantain crops in Puerto Rico (2006-2007), which usually estimates the percentage of leaf area affected. Nonlinear models yield a better fit of the disease progress. Furthermore, since the parameters of the nonlinear

curve vary, this variability is reflected with the inclusion of one or two random effects plant. This generates a correlation between observations from the same plant, so implicitly the correlation between repeated measures are incorporated into modeled.

Resumen de Tesis presentado en cumplimiento parcial de los
requerimientos para el grado de Maestría en Ciencias
MODELOS NO LINEALES MIXTOS CON VARIABLES DE RESPUESTA
CON DISTRIBUCIÓN BETA

By

Yuri Marcela García Saavedra

2012

Consejero: Raúl E. Macchiavelli

Departamento: Ciencias Matemáticas

Existen muchas situaciones en las que se involucran variables aleatorias con algún tipo de porcentaje, proporciones o fracciones. Muchos estudios han demostrado que los modelos de regresión lineal no son apropiados para modelar este tipo de datos. Sin embargo, la distribución beta es muy útil para modelar datos que se encuentran restringidos continuamente en el intervalo $(0,1)$ y que además pueden ser explicados por otras variables a través de una estructura de regresión. Debido a esto, Ferrari y Cribari-Neto (2004) propusieron un modelo de regresión beta lineal en el que la variable respuesta está distribuida como una beta usando una parametrización diferente de su función de densidad, obteniendo así, una estructura de regresión para la media de la respuesta con un parámetro de precisión constante. En este trabajo extendemos esta teoría, proponiendo un modelo de regresión beta no lineal mixto, donde la distribución condicional de las observaciones es asumida beta y la distribución de los efectos aleatorios se asume normal. Estudiamos la distribución marginal inducida y las propiedades del modelo propuesto mediante simulación, que se comparan con las obtenidas por un modelo de regresión no lineal mixto asumiendo distribución normal. Las estimaciones se realizan mediante máxima verosimilitud usando la técnica quasi-Newton para la optimización y la cuadratura Gaussiana para la integración. Los errores estándar de los parámetros del modelo fueron estimados mediante la matriz Hessiana. Finalmente, aplicamos estos resultados a estudios de severidad (cantidad relativa de tejido afectado en un momento dado) de enfermedades en cultivos de guineo en Puerto Rico (2006-

2007), donde se estima usualmente el porcentaje del área de la hoja afectada. Los modelos no lineales permitieron obtener un mejor ajuste de las curvas de progreso de la enfermedad. Además, como los parámetros de la curva no lineal varían, se reflejó esta variabilidad mediante la inclusión de uno y dos efectos aleatorios. Esto nos generará una correlación entre las observaciones provenientes de la misma planta, por lo que implícitamente se incorporan las correlaciones por mediciones repetidas en el modelo.

Copyright © 2012

By

Yuri Marcela García Saavedra

Índice general

1. Introducción	1
1.1. Antecedentes	1
1.2. Justificación	4
1.3. Objetivos	6
1.4. Organización del Trabajo	7
1.5. Implementación Computacional	7
2. Revisión de Literatura	8
2.1. Modelos Asumiendo Distribución Normal	8
2.1.1. Modelo Lineal	8
2.1.2. Modelo Lineal: Transformación de la Variable Respuesta	9
2.1.3. Modelo No Lineal	10
2.1.4. Modelo Normal Censurado o Modelo Tobit	11
2.2. Modelos Aplicando Regresión Robusta	12
2.3. Modelos Aplicando la Técnica de Cuasi-verosimilitud	13
2.4. Modelos Asumiendo Distribución Beta	15
2.5. Modelo de Regresión No Lineal	15

2.5.1.	Definición y Propiedades	15
2.5.2.	Estimación de Parámetros	17
2.6.	Modelo de Regresión No Lineal Mixto	20
3.	Modelos de Regresión Beta	26
3.1.	Distribución Beta	26
3.1.1.	Definición y Propiedades	26
3.1.2.	Estimación de Parámetros	28
3.2.	Modelo de Regresión Beta Lineal Generalizada	30
3.2.1.	Introducción	30
3.2.2.	Parametrización	31
3.2.3.	Definición	33
3.2.4.	Función de Enlace	33
3.2.5.	Estimación de Parámetros	36
3.3.	Modelo de Regresión Beta No Lineal	39
3.4.	Modelo de Regresión Beta No Lineal Mixto	40
3.4.1.	Introducción	40
3.4.2.	Definición del Modelo	41
3.4.3.	Estimación de Parámetros	41
4.	Simulación del Modelo Propuesto	44
4.1.	Proceso de Simulación	44
4.2.	Descripción Escenarios de Simulación	47

4.2.1. Modelo de Regresión No Lineal Mixto con Efecto Aleatorio de Intercepto . .	47
4.2.2. Modelo de Regresión No Lineal Mixto con Efecto Aleatorio de Pendiente . .	57
5. Aplicaciones: Estudio de Severidad de Enfermedades en Cultivos de Guineo en Puerto Rico	70
5.1. Enfermedad de la Sigatoka Negra	70
5.2. Descripción del Área Experimental y los Datos	71
5.3. Métodos de Análisis	74
5.4. Resultados	78
5.5. Conclusiones	84
6. Conclusiones Generales y Trabajos Futuros	86
6.1. Conclusiones Generales	86
6.2. Trabajos Futuros	88
Referencias	91
Anexos	92

Indice de Tablas

2.1. Datos observados para la variable respuesta longitud de circunferencia (y) en árboles de naranja.	22
2.2. Especificación del modelo para los datos de árbol	23
2.3. Criterios de ajuste.	24
2.4. Parámetros estimados	24
4.1. Parámetros de las distribuciones marginales y condicionales del modelo de regresión beta no lineal mixto con efecto aleatorio en β_0 para diferentes valores de σ_{b_0} y $\phi = 10$	49
4.2. Parámetros de las distribuciones marginales y condicionales del modelo de regresión beta no lineal mixto con efecto aleatorio en β_0 para diferentes valores de σ_{b_0} y $\phi = 15$	50
4.3. Parámetros de las distribuciones marginales y condicionales del modelo de regresión normal no lineal mixto con efecto aleatorio en β_0 para diferentes valores de σ_{b_0} y $\phi = 10$	51
4.4. Parámetros de las distribuciones marginales y condicionales del modelo de regresión normal no lineal mixto con efecto aleatorio en β_0 para diferentes valores de σ_{b_0} y $\phi = 15$	52
4.5. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 10$ y $\sigma_{b_0} = 0.22$	53

4.6. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 10$ y $\sigma_{b_0} = 0.27$	53
4.7. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 10$ y $\sigma_{b_0} = 0.33$	54
4.8. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 15$ y $\sigma_{b_0} = 0.22$	54
4.9. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 15$ y $\sigma_{b_0} = 0.27$	55
4.10. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 15$ y $\sigma_{b_0} = 0.33$	55
4.11. Prueba de McNemar para la convergencia de los modelos ajustados	57
4.12. Parámetros de las distribuciones marginales y condicionales del modelo de regresión beta no lineal mixto con efecto aleatorio en β_1 para diferentes valores de σ_{b_1} y $\phi = 10$	60
4.13. Parámetros de las distribuciones marginales y condicionales del modelo de regresión beta no lineal mixto con efecto aleatorio en β_1 para diferentes valores de σ_{b_1} y $\phi = 15$	61
4.14. Parámetros de las distribuciones marginales y condicionales del modelo de regresión normal no lineal mixto con efecto aleatorio en β_1 para diferentes valores de σ_{b_1} y $\phi = 10$	62
4.15. Parámetros de las distribuciones marginales y condicionales del modelo de regresión normal no lineal mixto con efecto aleatorio en β_1 para diferentes valores de σ_{b_1} y $\phi = 15$	63
4.16. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 10$ y $\sigma_{b_1} = 0.004$	64
4.17. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 10$ y $\sigma_{b_1} = 0.009$	64

4.18. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 10$ y $\sigma_{b_1} = 0.01$	65
4.19. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 10$ y $\sigma_{b_1} = 0.02$	65
4.20. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 15$ y $\sigma_{b_1} = 0.004$	66
4.21. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 15$ y $\sigma_{b_1} = 0.009$	66
4.22. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 15$ y $\sigma_{b_1} = 0.01$	67
4.23. Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 15$ y $\sigma_{b_1} = 0.02$	67
4.24. Prueba de McNemar para la convergencia de los modelos ajustados con efecto aleatorio de pendiente	69
5.1. Visualización de la base de datos para el análisis estadístico de la Sigatoka negra. . .	72
5.2. Criterios de ajuste para el modelo no lineal mixto con efecto aleatorio de intercepto. .	78
5.3. Criterios de ajuste para el modelo no lineal mixto con efecto aleatorio de pendiente. .	80
5.4. Criterios de ajuste para el modelo no lineal mixto con efecto aleatorio de intercepto y pendiente.	82

Índice de figuras

2.1. Gráfica de perfiles para los datos observados para el ejemplo de árboles de naranja.	22
2.2. Curvas ajustadas para los datos de arboles de naranja	24
3.1. Función de densidad beta para diferentes combinaciones de (p, q)	28
3.2. Función de densidad beta para diferentes combinaciones de (μ, ϕ)	32
4.1. Curva logística base para el proceso de simulación	45
5.1. Planta con síntomas característicos del ataque de Sigatoka negra (Marengo, 2010).	71
5.2. Grados de severidad de la enfermedad Sigatoka negra de acuerdo a la escala de Stover (Marengo, 2010).	73
5.3. Datos observados del IS vs. fecha en la enfermedad de Sigatoka negra.	74
5.4. Curvas ajustadas y gráficas de residuales del modelo no lineal mixto con efecto alea- torio de intercepto asumiendo distribución beta (-a- y -b-) y distribución normal (-c- y -d-).	80
5.5. Curvas ajustadas y gráfica de residuales del modelo no lineal mixto con efecto alea- torio de pendiente asumiendo distribución beta (-a- y -b-) y distribución normal (-c- y -d-)	82

5.6. Curvas ajustadas y gráficas de residuales del modelo no lineal mixto con efecto aleatorio de intercepto y pendiente asumiendo distribución beta (-a- y -b-) y distribución normal (-c- y -d-) 84

Introducción

1.1. Antecedentes

Los modelos lineales se utilizan en aquellas situaciones en las cuales la media puede ser descrita como una función lineal de un conjunto de parámetros desconocidos. Dicho en otras palabras, el modelo lineal puede explicar el comportamiento de la variable respuesta a partir de la o las variables predictoras correspondientes a un conjunto de datos, teniendo como base la existencia de y en el cual intervienen una serie de perturbaciones aleatorias que son ajenas al proceso, comúnmente llamados errores.

Nelder y Wedderburn (1972) presentan los Modelos Lineales Generalizados (GLM por sus siglas en inglés, Generalized Linear Models) como una nueva forma de investigación y modelación de datos cuando las observaciones son independientes y los supuestos del modelo lineal clásico no se cumplen. De esta manera, los GLM se formulan como una extensión a los modelos lineales los cuales permiten usar distribuciones que no son normales (por ejemplo, binomial, Poisson, Gamma) y varianzas no constantes.

Sea y la variable respuesta independiente generada a partir de una distribución de probabilidad que pertenece a la familia exponencial, las dos componentes que conforman a los GLM son:

- La componente aleatoria, que establece la distribución de probabilidad de la variable respuesta

y puede escribirse de la siguiente forma:

$$f(y|\theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} - c(y, \phi) \right\}, \quad (1.1)$$

donde θ es llamado parámetro natural, ϕ es el parámetro de dispersión y b y c son funciones conocidas correspondientes al tipo de distribución. El parámetro natural es una función de la media, que está determinada por la relación $\mu = b'(\theta)$. La varianza de y está descrita por la relación $Var(y) = \phi v(\mu)$ donde la función de varianza es determinada por $v(\mu) = b''(\theta)$.

- La componente sistemática que viene dada por:

$$g[E(y)] = g(\mu) = \eta = x^T \boldsymbol{\beta}, \quad (1.2)$$

donde η es el predictor lineal, $\boldsymbol{\beta}$ es el vector de parámetros de efectos fijos y g se denomina función de enlace la cual es monótona diferenciable (por ejemplo, función logit, probit y log-log complementario). La elección de la función de enlace se puede determinar por el ajuste que se le realice a los datos y por la facilidad en la interpretación de los parámetros en el predictor lineal.

La función de enlace que transforma la media al parámetro natural es llamada enlace canónico. Es decir, el enlace canónico relaciona directamente el parámetro natural con el predictor lineal de la forma

$$\theta = \theta(\mu) = \eta = x^T \boldsymbol{\beta}, \quad (1.3)$$

en otras palabras, $g(\mu) \equiv \theta(\mu)$.

En la práctica existen muchas situaciones en donde las observaciones procedentes de una misma unidad o sujeto (por ejemplo, planta, animal, parcela,...) tienden a parecerse más entre sí. Esto indica

falta de independencia entre las observaciones correspondientes al mismo sujeto, violando así los supuestos básicos en la solución de problemas de modelos lineales clásicos basados en la estimación de los parámetros por mínimos cuadrados ordinarios (OLS, por sus siglas en inglés, Ordinary Least Squares) y GLM basados en estimación por máxima verosimilitud. Los modelos mixtos inducen una estructura de correlación para las observaciones que consideran esta falta de independencia. Debido a casos como estos, surgen los modelos lineales generalizados mixtos (MLGM), en los cuales además de tener componentes o efectos fijos en el predictor lineal se incluyen efectos aleatorios independientes y en general, normalmente distribuidos. La especificación general de los MLGM viene dada por un vector de variables respuestas $\mathbf{y} = (y_1, y_2, \dots, y_t)$ con distribución condicional de la familia exponencial como:

$$g(E(\mathbf{y}|\mathbf{b})) = A\boldsymbol{\beta} + B\mathbf{b}, \quad (1.4)$$

con media condicional

$$E(\mathbf{y}|\mathbf{b}) = \boldsymbol{\mu}, \quad (1.5)$$

y varianza condicional

$$\text{var}(\mathbf{y}|\mathbf{b}) = \phi\nu(\boldsymbol{\mu}), \quad (1.6)$$

donde $\boldsymbol{\beta}$ es el vector de parámetros de efectos fijos, \mathbf{b} es un vector de efectos aleatorios independientes y A y B representan matrices de diseño asociadas a los efectos fijos y aleatorios, respectivamente. La diferencia entre estas matrices es que la matriz B no incluye una columna para el término constante.

Sin embargo, no todos los datos pueden ser caracterizados adecuadamente por modelos lineales o GLM. Muchas de las aplicaciones requieren modelos de regresión no lineal (MNL), en los cuales los parámetros ingresan a la ecuación de la esperanza en forma no lineal. El modelo de regresión no lineal

clásico para la t -ésima observación se define como:

$$E(y_t) = \mathbf{f}(x_t, \boldsymbol{\beta}), \quad (1.7)$$

donde \mathbf{f} es una función no lineal en los parámetros $\boldsymbol{\beta}$. En la especificación del modelo es necesario indicar la distribución de la variable respuesta y la independencia o dependencia entre los valores de la variable respuesta y .

Por otro lado, los supuestos de los MNL pueden ser generalizados de la misma forma como se hizo en los MLGM, incorporando una posible heteroscedasticidad (especificando una función de varianza) e incorporando uno o varios efectos aleatorios y que además no necesariamente estarían limitados a la distribución normal.

1.2. Justificación

Los modelos de regresión han sido una de las técnicas de mayor uso en el modelamiento de datos que de una u otra forma están relacionados a otra(s) variable(s). Particularmente, el modelo de regresión lineal ha sido un procedimiento muy usado en el modelamiento de datos medidos como una proporción continua en el intervalo (0,1) tales como probabilidades y proporciones (Paolino, 2001; Kieschnick y McCullough, 2003; Ferrari y Cribari-Neto, 2004; Santos de Oliveira, 2004; Salinas et al., 2006; Smithson y Verkuilen, 2006; Espinheira, 2007; Owen, 2008). Dicha metodología requiere del cumplimiento de los supuestos de linealidad, normalidad y homoscedasticidad (varianza constante), necesarios para la aplicación de la técnica de estimación por OLS. Sin embargo, este tipo de datos comúnmente no cumple con dichos supuestos, aunque en ocasiones los investigadores se apoyan en el argumento asintótico (presentado en muestras de tamaño lo suficientemente grande) para asegurar inferencias confiables.

Por otro lado, se ha recurrido a otras técnicas para contrarrestar la falta de cumplimiento de los supuestos; como por ejemplo aquellas que asumen distribución normal (transformación de la variable

respuesta, aplicación de un modelo normal censurado y función de respuesta no lineal). Estas metodologías de cierta manera ayudan a solucionar los problemas de linealidad, normalidad y homoscedasticidad. Sin embargo, es importante mencionar algunas falencias que dichas alternativas siguen presentando en el ajuste de estos datos. Algunas de ellas son:

- Los parámetros del modelo no pueden ser interpretados fácilmente en términos de la variable original.
- Al estar restringida la variable respuesta al intervalo $(0,1)$, se podrían generar valores ajustados que sobrepasen las cotas inferior y superior (Ferrari y Cribari-Neto, 2004).
- Las transformaciones no necesariamente estabilizan la varianza, por lo que no se garantizan varianzas constantes (Smithson y Verkuilen, 2006).
- Se requieren muestras grandes para garantizar inferencias correctas y confiables.

Otras técnicas como la regresión robusta y la cuasi-verosimilitud han sido aplicadas a datos restringidos al intervalo $(0,1)$. Por otro lado, la distribución beta es muy conocida por su flexibilidad para modelar variables en el intervalo $(0,1)$ (Casella y Berger, 2002). Dependiendo de los parámetros escogidos, la función de densidad de probabilidad puede tener forma unimodal, forma U, forma J o uniforme, lo que nos da una facilidad para modelar este tipo de datos ya que su distribución es típicamente asimétrica. Debido a esto, Ferrari y Cribari-Neto (2004) propusieron un modelo de regresión para variables que tienen distribución beta realizando una nueva parametrización de la función de densidad obteniendo así una estructura de regresión en términos de la media y un parámetro de precisión (dispersión) constante. Con esta parametrización se evita la necesidad de transformar la variable respuesta, ya que no impone el supuesto de homoscedasticidad y permite relaciones no lineales o algún tipo de asimetría.

Muchos estudios tornan alrededor de la propuesta de Ferrari y Cribari-Neto (2004). Por ejemplo, Cribari-Neto y Zeileis (2010) implementaron el paquete *betareg* en el software R (R-Team, 2010), Cribari-Neto y Souza (2011) trabajaron diferentes pruebas de inferencia en un modelo de regresión

lineal beta pero con parámetro de dispersión variable y, Ospina y Ferrari (2007 , 2012) generalizaron el modelo de regresión beta lineal para proporciones continuas conteniendo el 0 y el 1, es decir, en el intervalo $[0,1]$.

En este trabajo extendemos los modelos de regresión no lineal beta para datos de proporciones medidos en el intervalo continuo $(0,1)$ con parámetro de precisión constante, proponiendo un modelo de regresión beta no lineal mixto para así poder modelar datos longitudinales o correlacionados. Estudiamos la distribución marginal inducida, las propiedades del modelo propuesto y realizamos estimaciones por máxima verosimilitud usando métodos numéricos. Finalmente, aplicamos estos resultados a estudios de severidad (cantidad relativa de tejido afectado en un momento dado) de enfermedades en cultivos de guineo en Puerto Rico, donde se usa como variable de respuesta una estimación visual del porcentaje del área de la hoja afectada.

1.3. Objetivos

Objetivo General

Examinar las propiedades del modelo de regresión beta no lineal mixto.

Objetivos Específicos

- Examinar la distribución marginal inducida por el modelo de regresión beta no lineal mixto.
- Estudiar las propiedades del modelo de regresión beta no lineal mixto mediante simulación.
- Aplicar los resultados obtenidos a estudios de severidad de enfermedades en cultivos de guineo en Puerto Rico.
- Comparar los resultados obtenidos tanto en las simulaciones como en la aplicación de datos reales con los obtenidos asumiendo distribución normal para la variable dependiente y .

1.4. Organización del Trabajo

En el Capítulo 2 se sintetizan las diferentes metodologías utilizadas en el estudio de datos medidos como una proporción en el intervalo continuo $(0,1)$, se muestran las ventajas y desventajas de cada uno de ellos y, finalmente, se introduce la teoría de los modelos no lineales clásicos y modelos no lineales mixtos.

En el Capítulo 3 se presenta una descripción detallada de la distribución beta: su definición, propiedades e inferencia. De la misma manera, se especifican los modelos de regresión beta lineal y no lineal, finalizando con el modelo de regresión beta no lineal mixto.

En el Capítulo 4 se trabajan diferentes simulaciones para destacar algunas propiedades importantes en el modelo de regresión beta no lineal mixto. Entre las simulaciones se plantearon dos escenarios, uno de ellos es el ajuste de un modelo de regresión no lineal mixto con efecto aleatorio de intercepto.

En el Capítulo 5 se presentan las aplicaciones del modelo propuesto a datos reales en estudios de severidad de enfermedades en cultivos de guineo en Puerto Rico.

Finalmente, las conclusiones generales del trabajo e investigaciones futuros se tratan en el Capítulo 6.

1.5. Implementación Computacional

En este trabajo utilizaremos el procedimiento NLMIXED del software estadístico SAS (SAS Institute, 2003) el cual permite ajustar modelos no lineales mixtos. Este procedimiento permite realizar la estimación de los parámetros por máxima verosimilitud, análisis de diagnóstico y seleccionar diferentes aproximaciones a la función de verosimilitud (por ejemplo, la cuadratura gaussiana adaptativa o el método de primer orden) y algoritmos de optimización numérica (el método cuasi-newton, Newton-Raphson y gradiente conjugado, entre otros).

Revisión de Literatura

Diversas técnicas se pueden aplicar en el ajuste de datos medidos como una proporción continua en el intervalo (0,1). Sin embargo, no existe un consenso sobre cuál es el tipo de modelo que mejor ajusta este tipo de datos. Aquí se presenta una síntesis de los métodos y procedimientos comúnmente utilizados en la modelación de estos datos, presentando sus ventajas y desventajas. Finalmente, se hace una breve descripción del modelo clásico de regresión no lineal y del modelo de regresión no lineal mixto.

2.1. Modelos Asumiendo Distribución Normal

2.1.1. Modelo Lineal

Durante mucho tiempo la regresión lineal ha sido uno de los procedimientos más usados en el modelamiento de datos medidos como una proporción continua (Paolino, 2001; Kieschnick y McCullough, 2003; Salinas et al., 2006; Espinheira, 2007). La estimación de los parámetros es realizada por OLS y requiere del cumplimiento de los supuestos de linealidad, normalidad y homoscedasticidad. Sin embargo, sin importar si se cumplen o no los supuestos distribucionales los investigadores se apoyan constantemente en el argumento asintótico, ya que permite generar resultados válidos y confiables. Es importante mencionar, que en la inferencia esto no se cumple. Godfrey (1988) es más estricto al afir-

mar que cuando se realiza una prueba t o F se asume implícitamente que la distribución condicional es distribuida normalmente sin importar el tamaño de la muestra. De esta manera, se presentan varias falencias originadas con el uso de esta práctica. Algunas de ellas son:

- La varianza está en función de la media, la varianza se acerca a cero cuando la media se aproxima a los límites inferior y superior, lo que subestima los estimadores puntuales de los coeficientes, es decir, las varianzas son heterogéneas (Salinas et al., 2006).
- Al estar la variable respuesta medida en el intervalo acotado (0,1), la función de la esperanza debe ser no lineal, es decir, no se puede establecer una relación lineal entre la variable respuesta y la(s) variable(s) predictor(a)s).
- Al estar restringida la variable respuesta al intervalo (0,1), se podrían generar valores ajustados que sobrepasen las cotas inferior y superior (Ferrari y Cribari-Neto, 2004), ya que una función lineal toma valores en \mathbb{R} .

Al observar las desventajas que presenta esta técnica consideramos que el modelo de regresión lineal no es adecuado para datos de proporciones continuas.

2.1.2. Modelo Lineal: Transformación de la Variable Respuesta

Como su nombre lo indica, se transforma la variable respuesta para que asuma valores en \mathbb{R} y así solucionar los problemas de linealidad, normalidad y homoscedasticidad (supuestos necesarios para la aplicación de los OLS), además de remediar situaciones como los valores atípicos. Posteriormente se aplica un modelo de regresión lineal sobre la variable transformada usando la técnica OLS. Aunque existen varias transformaciones para datos de proporciones y porcentajes, las más usadas son la logit y la logarítmica. De hecho, Atkinson (1985) recomienda las dos porque generalmente logran linealizar la relación entre la variable respuesta y la(s) variable(s) predictor(a)s. La transformación logit se define como:

$$\text{logit}(y) = \ln\left(\frac{y}{1-y}\right) = x^T \boldsymbol{\beta} + \varepsilon,$$

donde $\ln(y/(1 - y))$ representa la transformación logit de la variable dependiente, la cual evita la posibilidad de producir valores esperados fuera de las cotas permitidas. La transformación logarítmica genera sólo valores no negativos y se define como:

$$-\ln(y) = x^T \boldsymbol{\beta} + \varepsilon$$

Los supuestos de linealidad, normalidad y homoscedasticidad deben de cumplirse para tener inferencias válidas y confiables. Estos enfoques pueden ser útiles si se cumplen las condiciones que ellos imponen. Sin embargo, las desventajas más relevantes de esta práctica en datos de proporciones continuas son:

- La transformación no necesariamente estabiliza la varianza, por lo que no garantiza varianzas constantes.
- Los parámetros del modelo no pueden ser interpretados fácilmente en términos de la variable original.
- Se requieren muestras grandes para garantizar un ajuste correcto y confiable.
- Puede haber valores de 0 ó 1 lo que es imposible en el contexto de proporciones en (0,1).

2.1.3. Modelo No Lineal

El uso de un modelo no lineal en el estudio de proporciones continuas consiste en ajustar un modelo de regresión no lineal aplicando mínimos cuadrados. En este caso, se tiene una estructura no lineal para una variable aleatoria continua y distribuida normalmente. Hermalin y Wallace (1972) usaron la

distribución normal acumulada como su función de esperanza condicional y estimaron los parámetros con el método de mínimos cuadrados. A diferencia, Kieschnick y McCullough (2003) usaron la función logística acumulada porque se enfoca más al supuesto distribucional; sin embargo, Smithson y Verkuilen (2006) manifiestan que la efectividad de este procedimiento depende de la estabilización de la varianza y la normalidad de los residuos.

Específicamente, Kieschnick y McCullough (2003) asumen que:

$$E(y_t) = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 x_t)\}},$$

donde

$$y_t \sim \mathbf{N}(E(y_t), \sigma^2)$$

La estimación de los parámetros la realizaron usando mínimos cuadrados. Debido a que en la aplicación de este trabajo se ajustan modelos no lineales, en la Sección 2.5 se tratarán aspectos complementarios de los modelos clásicos de regresión no lineal.

2.1.4. Modelo Normal Censurado o Modelo Tobit

En algunas investigaciones se ha aplicado el modelo normal censurado o modelo Tobit para datos de proporciones continuas en el intervalo (0,1). Básicamente, este procedimiento asume que:

$$y_t^* = x_t^T \boldsymbol{\beta} + \varepsilon_t, \quad t = 1, 2, \dots, n$$

y

$$y_t = \begin{cases} 0, & \text{si } y_t^* \leq 0, \\ y_t^*, & \text{si } 0 < y_t^* < 1, \\ 1, & \text{si } y_t^* \geq 1. \end{cases}$$

donde los $\{\varepsilon_t\}$ son asumidos independientes e idénticamente distribuidos como una $N(0, \sigma^2)$. Algunos de los inconvenientes que tiene este procedimiento son:

- El supuesto de normalidad para los y_t^* , se realiza sólo para un rango específico, no se sabe que sucede con los valores fuera de $[0,1]$ ya que ellos están censurados. De esta manera, un modelo normal censurado es inapropiado para datos de proporciones (Kieschnick y McCullough, 2003).
- El modelo normal censurado es equivalente al modelo de regresión normal, por lo que las deficiencias de este enfoque están sujetas a las mismas del modelo de regresión normal.

2.2. Modelos Aplicando Regresión Robusta

Aunque es una técnica poco común, Smithson y Verkuilen (2006) hacen alusión al método de regresión robusta como una alternativa viable para contrarrestar aspectos importantes como la asimetría y heteroscedasticidad, comunes en datos de proporciones en un intervalo continuo. Inicialmente este enfoque fue usado para manejar problemas ocasionados por la presencia de outliers ¹ con el fin de reducir o quizás eliminar el efecto de los residuos grandes. La idea general de la regresión robusta para el contexto de proporciones continuas es estimar los parámetros usando algún tipo de recorte o ponderación de las observaciones más extremas. Sin embargo, en la comparación que hacen los autores con la regresión beta (técnica tratada más adelante) mencionan aspectos en los cuales hay que tener bastante cuidado en el momento de aplicar esta técnica. Dos de estos aspectos son que la regresión robusta tiene como premisa que la variable dependiente no es acotada y que la distribución

¹La presencia de outliers influyen de manera importante en el ajuste de datos por el método de mínimos cuadrados, presentando inferencias poco confiables en el análisis de los residuales.

subyacente es al menos simétrica. Además, sigue teniendo la influencia de datos atípicos (outliers). Para más detalles sobre regresión robusta ver, Rousseeuw y Leroy (1987).

2.3. Modelos Aplicando la Técnica de Cuasi-verosimilitud

En la mayoría de las investigaciones se conoce cuál es la distribución de probabilidad que tienen los datos; sin embargo, existen casos en que la información existente no es suficiente para establecer la forma de la distribución que siguen las observaciones. De esta manera no se podría construir la función de verosimilitud y, por ende, no se podrían aplicar métodos como la máxima verosimilitud. Inicialmente, Wedderburn (1974) plantea el enfoque de la cuasi-verosimilitud (investigado más en detalle por McCullagh y Nelder, 1989) como un método de cuantificación sin necesidad de especificar algún supuesto distribucional, y que aún así, sigue generando inferencias confiables. En esta técnica sólo se necesita especificar la relación entre la media y la varianza de las observaciones. De esta manera, la cuasi-verosimilitud es útil cuando dicha relación no está dada por una distribución de probabilidad estándar (Smithson y Verkuilen, 2006).

Un modelo cuasi-verosímil especifica únicamente el primer y segundo momento de la distribución condicional, donde la varianza de los datos observados está en función de la media. A su vez, la media es una función conocida de los parámetros desconocidos.

Sea $\mathbf{y} = (y_1, y_2, \dots, y_n)$ un vector de variables respuestas independientes con vector de medias $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$ y matriz de covarianza $\sigma^2 \mathbf{V}(\boldsymbol{\mu})$, donde σ^2 puede ser desconocida y $\mathbf{V}(\boldsymbol{\mu})$ es una matriz de funciones conocidas y es diagonal por la independencia de las variables respuesta, por lo tanto, $\mathbf{V}(\boldsymbol{\mu}) = \text{diag}\{V_1(\boldsymbol{\mu}), V_2(\boldsymbol{\mu}), \dots, V_n(\boldsymbol{\mu})\}$. Los parámetros de interés $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)$ relacionan la dependencia de $\boldsymbol{\mu}$ con las variables independientes $\mathbf{x} = (x_1, x_2, \dots, x_k)$. En base a esto, se define $E(\mathbf{y}) = \boldsymbol{\mu}$, $g(E(\mathbf{y})) = g(\boldsymbol{\mu}) = \mathbf{x}^T \boldsymbol{\beta}$, donde g es una función de enlace (ver Sección 3.2.4) y $\text{Var}(\mathbf{y}) = \sigma^2 \mathbf{V}(\boldsymbol{\mu})$. La estimación de los parámetros se halla maximizando la función de log cuasi-verosimilitud (en lugar de la función de log-verosimilitud), considerada como función de $\boldsymbol{\mu}$. La función de cuasi-verosimilitud está dada por el sistema de ecuaciones diferenciales

$$\frac{\partial Q(y_t, \mu_t)}{\partial \mu_t} = \frac{y_t - \mu_t}{\sigma^2 V(\mu_t)}$$

o equivalentemente

$$Q(y_t, \mu_t) = \int_{y_t}^{\mu_t} \frac{y_t - \tau}{\sigma^2 V(\tau)} d\tau$$

En el contexto de aplicaciones a datos de proporciones continuas en el intervalo (0,1), [Cox \(1996\)](#) investigó las siguientes cuatro especificaciones para los dos primeros momentos de la distribución condicional en dos muestras de datos:

- Par canónico: $\mu(\theta) = 1/(1 + \exp(-\theta))$ y $v(\mu) = \mu(1 - \mu)$,
- Par ortogonal: $\mu(\theta) = 1/(1 + \exp(-\theta))$ y $v(\mu) = \mu^2(1 - \mu)^2$,
- Par canónico: $\mu(\theta) = 1 - \exp\{-\exp(\theta)\}$ y $v(\mu) = (1 - \mu) \ln(1 - \mu)$,
- Par ortogonal: $\mu(\theta) = 1 - \exp\{-\exp(\theta)\}$ y $v(\mu) = (1 - \mu)^2 \ln^2(1 - \mu)$.

En ellas examina el uso de la función de enlace logit y log-log complementario con expresiones canónicas y ortogonales para la función de varianza, concluyendo que la combinación más adecuada para su conjunto de datos es la función logit con la función de varianza ortogonal.

Aunque el procedimiento de cuasi-verosimilitud ha sido una de las técnicas más usadas en la aplicaciones de datos reales, requiere un tamaño de muestra suficientemente grande y solamente modela los dos primeros momentos de la distribución. Por lo tanto, las posibles desventajas se dejan al criterio de cada investigador, ya que no hay un consenso de las posibles falencias.

2.4. Modelos Asumiendo Distribución Beta

La distribución beta es muy útil en el modelamiento de datos medidos como una proporción continua en el intervalo (0,1). Ferrari y Cribari-Neto (2004) propusieron un modelo de regresión para variables que tienen una distribución beta. Específicamente, realizaron una re-parametrización de la función de densidad, lo cual permitió definir la media y un parámetro de dispersión para el modelo. Ellos indican que la distribución beta tiene ventajas que se adecúan a este tipo de datos. Por ejemplo:

- No hay necesidad de transformar la variable dependiente.
- La varianza de la variable respuesta está en función de la media, por lo que se flexibiliza el supuesto de homoscedasticidad, permitiendo situaciones en que las varianzas son heterogéneas.
- Se pueden modelar diferentes formas de la distribución de acuerdo a los valores que se le asignen a sus parámetros, permitiendo así algún tipo de asimetría o relaciones no lineales.
- Estimaciones de los parámetros más precisas.

Debido a que el objetivo principal de este trabajo de tesis es examinar las propiedades del modelo de regresión beta, en el Capítulo 3 se realiza en forma detallada una descripción de la distribución beta tales como su definición, propiedades e inferencia.

2.5. Modelo de Regresión No Lineal

2.5.1. Definición y Propiedades

A diferencia de los modelos de regresión lineal en los cuales los parámetros ingresan linealmente en la ecuación, en los modelos no lineales los parámetros de regresión entran en forma no lineal. De esta manera, en este tipo de modelos se presenta la esperanza de la respuesta, como una función

no lineal de los parámetros. Usualmente en los modelos no lineales se usan menos parámetros que en un modelo lineal, dando así una descripción más parsimoniosa (menos parámetros) de los datos. Además, proveen predicciones más confiables para la variable respuesta fuera del rango observado de los datos, al contrario de los modelos polinomiales (Pinheiro y Bates, 2000).

En la especificación del modelo es necesario indicar la distribución de la variable respuesta y la independencia o dependencia entre los valores de la variable respuesta y . Así, el modelo no lineal clásico para la t -ésima observación está definido como:

$$E(y_t) = \mathbf{f}(x_t, \boldsymbol{\beta}), \quad (2.1)$$

donde y_t denota la respuesta tomada en el t -ésimo valor de la covariable x_t con $t = 1, 2, \dots, n$, los x_t son considerados a menudo como cantidades fijas y la función de regresión \mathbf{f} depende de $\boldsymbol{\beta}$ en forma no lineal, $\boldsymbol{\beta}$ es el vector de parámetros de regresión. De esta manera, en un modelo no lineal clásico los y_t son considerados independientes y normalmente distribuidos con:

$$E(y_t) = \mathbf{f}(x_t, \boldsymbol{\beta}) \quad y \quad Var(y_t) = \sigma^2. \quad (2.2)$$

Sin embargo, el supuesto de varianzas constantes es violado frecuentemente en la práctica. Por ejemplo, en datos de crecimiento donde se muestra el coeficiente de variación constante en vez de la varianza constante; en otras palabras, esto significa que la varianza es proporcional al cuadrado de la media de la variable respuesta (Davidian y Giltinan, 1995). En este caso, un supuesto más apropiado que el dado en la ecuación (2.2) podría ser:

$$E(y_t) = \mathbf{f}(x_t, \boldsymbol{\beta}) \quad y \quad Var(y_t) = \sigma^2 \{\mathbf{f}(x_t, \boldsymbol{\beta})\}^2, \quad (2.3)$$

donde el parámetro de escala σ es el coeficiente de variación. Para más detalles de varianzas heterogéneas ver Davidian y Giltinan (1995, sec. 2.2.1). Algunos ejemplos de modelos no lineales clásicos

son:

- Modelo Exponencial: $E(y) = \beta_0(1 - \exp(-\beta_1 x))$,
- Modelo bi-Exponencial: $E(y) = \beta_0 \exp(\beta_1 x) - \beta_2 \exp(-\beta_3 x)$,
- Modelo Gompertz: $E(y) = \exp(-\beta_0 \exp(-\beta_1 x))$,
- Modelo Richards: $E(y) = \frac{\beta_0}{[1 + \exp(-(\beta_1 + \beta_2 x))]^{\beta_3}}$,
- Modelo Monomolecular: $E(y) = \beta_0 - \beta_1 \exp(-\beta_2 x)$,
- Modelo Logístico: $E(y) = \frac{\beta_0}{1 + \beta_1 \exp(-\beta_2 x)}$.

2.5.2. Estimación de Parámetros

El principal interés de un modelo de regresión es la inferencia acerca de los parámetros de regresión β . Aquí, se realiza una breve descripción de los métodos de estimación más usuales en los modelos de regresión no lineales.

Método de Mínimos Cuadrados Ordinarios

El estimador mínimo cuadrado ordinario para β minimiza

$$\sum_{t=1}^n [y_t - \mathbf{f}(x_t, \beta)]^2, \quad (2.4)$$

equivalentemente, el estimador de mínimos cuadrados ordinarios β resuelve las p-ecuaciones de estimación

$$\sum_{t=1}^n [y_t - \mathbf{f}(x_t, \beta)] \frac{\partial \mathbf{f}(x_t, \beta)}{\partial \beta_k} = 0, \quad k = 1, 2, \dots, p, \quad (2.5)$$

Es importante resaltar que bajo los supuestos de independencia, normalidad y varianzas constantes, el estimador mínimo cuadrado ordinario es también el estimador máximo verosímil para β (Davidian y Giltinan, 1995, sec. 2.3.1).

Método de Mínimos Cuadrados Generalizados

En situaciones donde la varianza de la variable respuesta y no es constante, el método de mínimos cuadrados ordinarios de β puede ser ineficiente. El procedimiento de mínimos cuadrados puede ser modificado para varianzas heterogéneas si las varianzas de y_t son conocidas como una constante de proporcionalidad, de tal forma que,

$$E(y_t) = \mathbf{f}(x_t, \beta) \quad y \quad \text{Var}(y_t) = \frac{\sigma^2}{w_t}. \quad (2.6)$$

para alguna constante conocida w_t , $t = 1, 2, \dots, n$. Cabe resaltar, que con los supuestos de errores normales e independientes y bajo lo descrito en la ecuación (2.6), es relativamente sencillo mostrar que el estimador máximo verosímil de β es el valor que minimiza

$$\sum_{t=1}^n w_t [y_t - \mathbf{f}(x_t, \beta)]^2. \quad (2.7)$$

Equivalentemente, el estimador por mínimos cuadrados ponderado de β resuelve las ecuaciones de estimación

$$\sum_{t=1}^n w_t [y_t - \mathbf{f}(x_t, \beta)] \frac{\partial \mathbf{f}(x_t, \beta)}{\partial \beta_k} = 0. \quad (2.8)$$

El estimador mínimo cuadrático ponderado tiene una interpretación natural: la suma de cuadrados de las desviaciones es minimizada, pero cada desviación es ponderada por la proporción inversa de la magnitud de la respuesta asociada a la incertidumbre (Davidian y Giltinan, 1995, sec. 2.3.2). Finalmente, un estimador que toma en cuenta que la varianza es directamente proporcional a la media puede ser obtenido calculando las estimaciones ponderadas de la siguiente forma

$$\widehat{w}_t = \frac{1}{\mathbf{f}^2(x_t, \widehat{\beta}_{OLS})} \quad (2.9)$$

donde $\widehat{\beta}_{OLS}$ es el estimador de β por mínimos cuadrados ordinarios y resolviendo la ecuación (2.8) reemplazando \widehat{w}_i en lugar de w_i . Asumiendo lo dado en la ecuación (2.3) esta estimación es conocida en el contexto de los modelos de regresión no lineales como *mínimos cuadrados generalizados*. y debe ser obtenida de forma iterativa.

Método de Máxima Verosimilitud

En ocasiones la naturaleza de la variable dependiente impide la utilización de métodos tradicionales para la estimación de los parámetros, por lo tanto una opción viable es el método de máxima verosimilitud. Los estimadores máximo verosímiles son aquellos valores para los cuales la función de verosimilitud alcanza su máximo. Sea y_1, y_2, \dots, y_n una muestra aleatoria con función de densidad conjunta $p(y_1, y_2, \dots, y_n | \theta) = \prod_{i=1}^n p(y_i; \theta)$, donde θ es el vector de parámetros. La función de verosimilitud para un modelo no lineal está dada por

$$L(\beta, \sigma^2 | y_1, y_2, \dots, y_n) = \prod_{i=1}^n p(y_i; \beta, \sigma^2), \quad (2.10)$$

Asumiendo distribución normal y respuestas no correlacionadas, las ecuaciones de estimación se obtienen derivando la función de verosimilitud dada en la ecuación (2.10) respecto a los parámetros de interés. Para ilustrar el procedimiento de máxima verosimilitud en un modelo no lineal se muestra el siguiente ejemplo:

Sea p la función de densidad de la distribución normal

$$p(y) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(y - \mu)^2}{\sigma^2} \right\}, \quad y \in \mathbb{R}. \quad (2.11)$$

La función de verosimilitud para la ecuación (2.11) de un modelo no lineal queda determinada como:

$$L(\beta, \sigma^2 | y_1, y_2, \dots, y_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{f}(x_i, \beta))^2 \right\}. \quad (2.12)$$

Así el logaritmo de la función de verosimilitud dada en la ecuación (2.13) es:

$$\ell = \log L(\boldsymbol{\beta}, \sigma^2 | y_1, y_2, \dots, y_n) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \mathbf{f}(x_t, \boldsymbol{\beta}))^2. \quad (2.13)$$

Se escoge el modelo no lineal $E(y_t) = \mathbf{f}(x_t, \boldsymbol{\beta}) = 1 - \exp(-\exp(\beta_0 + \beta_1 x_t))$ y se sustituye en el logaritmo de la función de verosimilitud dada en la ecuación (2.13), se deriva e igualamos a cero para obtener las ecuaciones de estimación para cada parámetro.

Respecto a β_0

$$\frac{\partial \ell}{\partial \beta_0} = -\frac{1}{\sigma^2} \sum_{t=1}^n [(y_t - \exp(-\exp(\beta_0 + \beta_1 x_t))) \{\exp(-\exp(\beta_0 + \beta_1 x_t))\} \{\exp(\beta_0 + \beta_1 x_t)\}] = 0$$

Respecto a β_1

$$\frac{\partial \ell}{\partial \beta_1} = -\frac{1}{\sigma^2} \sum_{t=1}^n [x_t \{y_t - \exp(-\exp(\beta_0 + \beta_1 x_t))\} \{\exp(-\exp(\beta_0 + \beta_1 x_t))\} \{\exp(\beta_0 + \beta_1 x_t)\}] = 0$$

Respecto a σ^2

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{t=1}^n \{y_t - \exp(-\exp(\beta_0 + \beta_1 x_t))\}^2 = 0$$

Estas ecuaciones son no lineales en β_0 y β_1 y no existe una solución analítica sencilla. En general, se deben usar los métodos iterativos para encontrar el valor de las estimaciones de los parámetros β_0 y β_1 .

2.6. Modelo de Regresión No Lineal Mixto

En la Sección 2.5 se mostró que un modelo no lineal se puede definir como $E(y_t) = \mathbf{f}(x_t, \boldsymbol{\beta})$, donde $\boldsymbol{\beta}$ es el vector de parámetros de efectos fijos y \mathbf{f} es una función no lineal en los parámetros. En

la práctica existen situaciones en donde no se cumplen los supuestos de independencia y varianzas constantes básicos del modelo clásico de regresión no lineal. En este tipo de datos las observaciones procedentes de un mismo individuo (por ejemplo árbol, parcela) tienden a estar correlacionados, es decir, cierta medición puede depender de los resultados de alguna medición pasada; por ejemplo, los datos de crecimiento o desarrollo de una enfermedad. Esta falta de independencia necesita ser tomada en cuenta en el ajuste de los datos para obtener inferencias confiables y válidas. Lindstrom y Bates (1990) propusieron los modelos no lineales mixtos para datos longitudinales o correlacionados como una alternativa de solución. Una componente o efecto aleatorio es agregada al modelo. Este efecto ayuda a explicar la correlación o una posible heteroscedasticidad entre los individuos.

En un nivel, la j -ésima observación en cada uno de los t -grupos independientes es modelada como:

$$E(y_{tj}|\mathbf{b}_t) = \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}_t, x_{tj}), \quad t = 1, 2, \dots, m, \quad j = 1, 2, \dots, n_t, \quad (2.14)$$

donde $\boldsymbol{\beta}$ es el vector de parámetros de efectos fijos, \mathbf{b}_t es el vector de parámetros de efectos aleatorios, x_{tj} es el vector de covariables y \mathbf{f} es una función no lineal en $\boldsymbol{\beta}$.

Asumiendo normalidad los

$$\mathbf{b}_t \sim \mathbf{N}(0, \mathbf{G}),$$

donde \mathbf{G} es la matriz de varianzas-covarianzas.

Para ilustrar la formulación y ajuste de un modelo no lineal mixto asumiendo distribución normal para la variable dependiente se muestra el siguiente ejemplo.

Ejemplo (Draper y Smith, 1981)

Los datos que usaremos para ilustrar el ajuste de un modelo no lineal mixto consiste en un estudio longitudinal realizado a 5 árboles de naranja que fueron medidos durante 7 ocasiones. La variable respuesta observada fue la longitud de la circunferencia del tronco (ver Tabla 2.1).

Observación	árbol	días	y
1	1	118	30
2	1	484	58
3	1	664	87
4	1	1004	115
5	1	1231	120
6	1	1372	142
7	1	1582	145
8	2	118	33
9	2	484	69
10	2	664	111
11	2	1004	156
⋮	⋮	⋮	⋮
31	5	664	81
32	5	1004	125
33	5	1231	142
34	5	1372	174
35	5	1582	177

Tabla 2.1: Datos observados para la variable respuesta longitud de circunferencia (y) en árboles de naranja.

En la Figura 2.1 se muestra los perfiles de los datos observados para cada árbol a través del tiempo (días). Se puede observar que las curvas tienen una forma de S y varían de árbol a árbol, esta variabilidad se incrementa a medida que va pasando el tiempo.

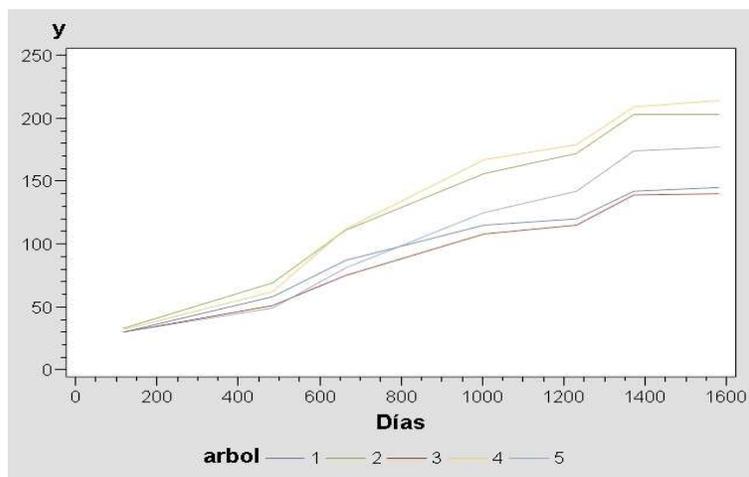


Figura 2.1: Gráfica de perfiles para los datos observados para el ejemplo de árboles de naranja.

Lindstrom y Bates (1990) proponen un modelo no lineal mixto usando una curva logística para ajustar estos datos. El efecto aleatorio, $\mathbf{b}_t = b_{t2}$, se adiciona a la asíntota. El modelo propuesto es:

$$E(y_{tj}|b_{t2}) = \frac{\beta_0 + b_{t2}}{1 + \exp[-(t_{ij} - \beta_1)/\beta_2]}, \quad (2.15)$$

donde

$$y_{tj}|b_{t2} \sim \mathbf{N}(E(y_{tj}|b_{t2}), \sigma_\varepsilon^2)$$

y el efecto aleatorio b_{t2} es distribuido como:

$$b_{t2} \sim \mathbf{N}(0, \sigma_{b_2}^2).$$

Los y_{tj} representa la j -ésima medición en el t -ésimo árbol, t_{ij} es el tiempo en días y $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ son los parámetros de efectos fijos.

Con este modelo se busca ajustar los datos teniendo en cuenta la variabilidad que existe de árbol a árbol, y la falta de independencia de las observaciones correspondientes al mismo sujeto (árbol). El modelo se ajustó con PROC NLMIXED de SAS (SAS Institute Inc., 2003), obteniendo los siguientes resultados.

Especificaciones del modelo	
Data Set	WORK.TREE
Variable dependiente	y
Distribución para la variable dependiente	Normal
Efecto aleatorio	b_2
Distribución para el efecto aleatorio	Normal
Sujeto	árbol
Técnica de optimización	Dual Quasi-Newton
Método de integración	Cuadratura Gaussiana Adaptativa

Tabla 2.2: Especificación del modelo para los datos de árbol

Criterios de ajuste	
-2 Log Likelihood	263.1
AIC (más pequeño es mejor)	273.1
AICC (más pequeño es mejor)	275.2
BIC (más pequeño es mejor)	271.2

Tabla 2.3: Criterios de ajuste.

Parámetros Estimados					
Parámetro	Estimación	Error Estándar	GL	valor t	Prob > t
β_0	192.05	15.6473	4	12.27	0.0003
β_1	727.9	35.2472	4	20.65	<.0001
β_2	348.07	27.079	4	12.85	0.0002
$\sigma_{b_1}^2$	999.88	647.44	4	1.54	0.1974
σ_{ε}^2	61.5139	15.8831	4	3.87	0.0179

Tabla 2.4: Parámetros estimados

En la Figura 2.2 se muestran las curvas ajustadas con el modelo no lineal mixto asumiendo distribución normal para la variable respuesta. Como se puede observar el ajuste a los datos observados es bastante bueno, esto nos indica que el modelo no lineal mixto parece ser una alternativa adecuada para el estudio de datos longitudinales.

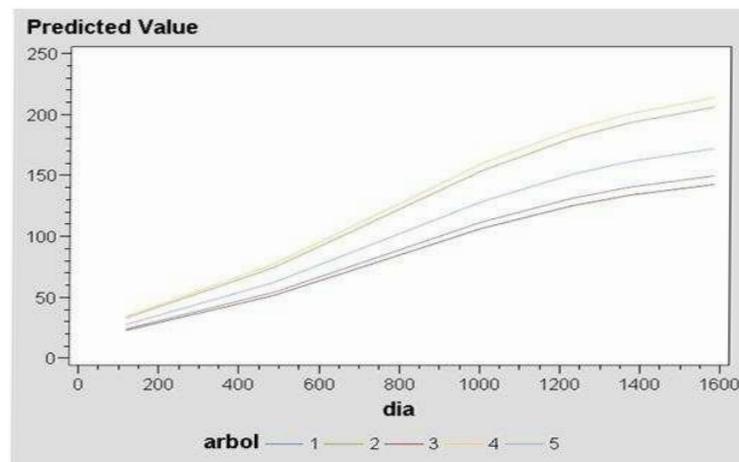


Figura 2.2: Curvas ajustadas para los datos de arboles de naranja .

En los Capítulos 3, 4 y 5 se analiza en detalle este tipo de modelos en situaciones en que la distribución de la variable respuesta dado el efecto aleatorio no es normal. Se describen aspectos como los modelos de regresión beta, se realizan simulaciones y se aplican los modelos de regresión no lineal mixtos a datos reales.

Modelos de Regresión Beta

En este Capítulo se presenta una descripción detallada de la distribución beta: su definición, propiedades e inferencia. De la misma manera, se especifican los modelos de regresión beta lineal y no lineal, finalizando con el modelo de regresión beta no lineal mixto.

3.1. Distribución Beta

3.1.1. Definición y Propiedades

La distribución beta es una distribución de probabilidad continua con dos parámetros p y q definida en el intervalo $(0,1)$ (Casella y Berger, 2002). La función de densidad está dada por:

$$f(y|p, q) = \frac{1}{B(p, q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1, \quad p > 0, \quad q > 0,$$

donde $B(p, q)$ representa la función beta :

$$B(p, q) = \int_0^1 y^{p-1} (1-y)^{q-1} dy.$$

La función beta está relacionada a la función gamma a través de la siguiente identidad:

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)}.$$

en donde $\Gamma(p)$ es una función gamma evaluada en el punto p , es decir,

$$\Gamma(p) = \int_0^{\infty} y^{p-1} e^{-y} dy.$$

De esta manera, la función de densidad se puede escribir como :

$$f(y; p, q) = \frac{\Gamma(p + q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1 - y)^{q-1}, \quad 0 < y < 1, \quad p > 0, \quad q > 0. \quad (3.1)$$

Así; la media, la varianza y la moda (que existe cuando p y q son mayores que 1) de variables aleatorias que siguen una distribución beta como en la ecuación (3.1) están dadas por:

$$E(y) = \frac{p}{p + q}, \quad (3.2)$$

$$Var(y) = \frac{pq}{(p + q)^2(p + q + 1)}, \quad (3.3)$$

y

$$moda(y) = \frac{p - 1}{p + q - 2}.$$

La distribución beta es muy conocida por su flexibilidad para modelar variables medidas en el intervalo (0,1). Dependiendo de los valores asignados a los dos parámetros, la función de densidad de probabilidad puede tener forma unimodal, forma U, forma J o uniforme (cuando $p = q = 1$), facilitando el modelamiento de datos que involucran tasas y proporciones, situaciones en que la respuesta se encuentra entre 0 y 1 y su distribución es típicamente asimétrica. Algunas de estas formas se muestran en la Figura 3.1. Las partes (a) y (b) representan funciones de densidad simétricas ($p = q$), con forma unimodal en (a) y U ($p, q < 1$) en (b); (c), (d) y (e) representan funciones de densidad asimé-

tricas ($p \neq q$), con forma U en (c) y unimodal en (d) y (e). Finalmente en (f) se tiene una función de densidad de forma J (Santos de Oliveira, 2004).

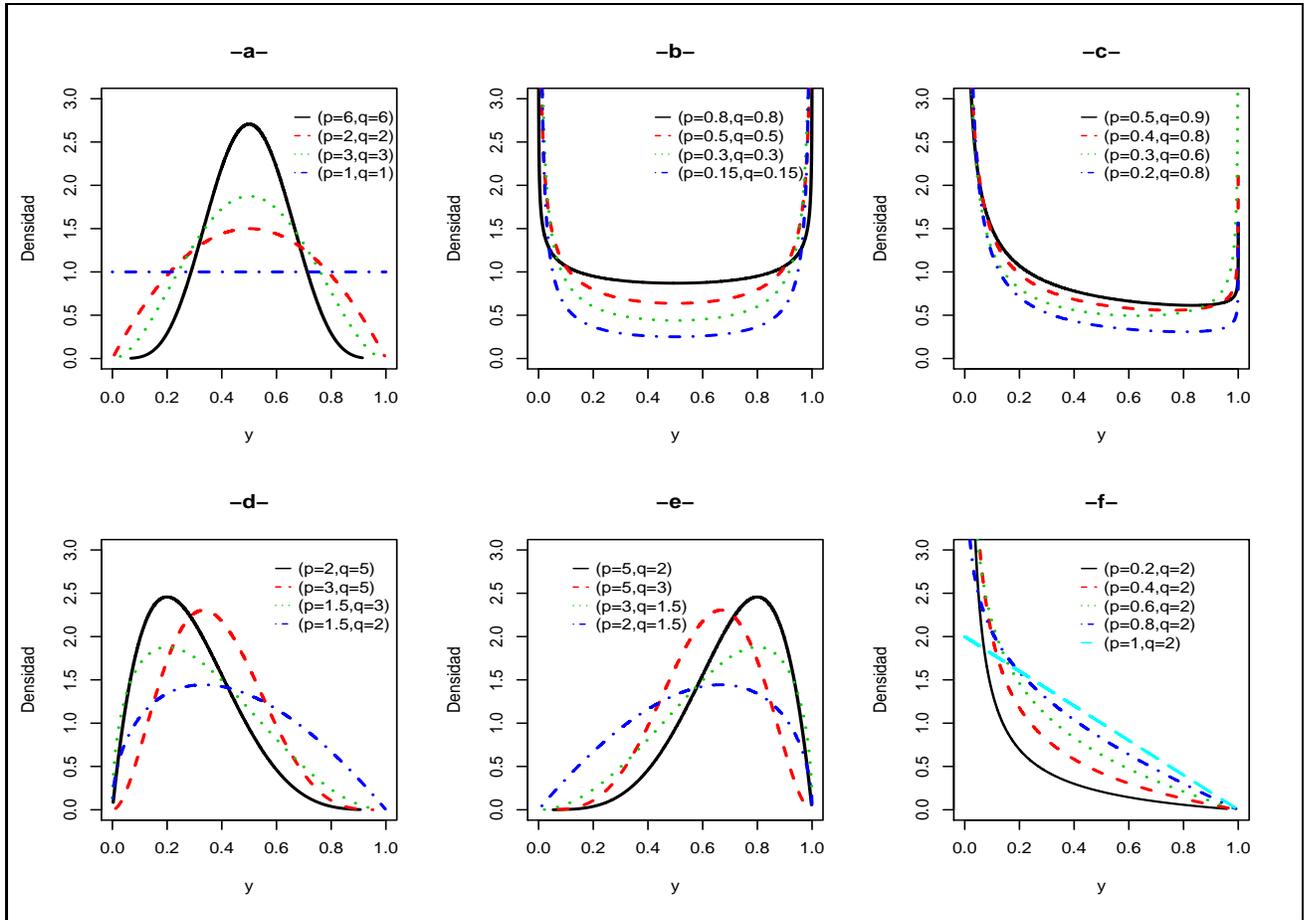


Figura 3.1: Función de densidad beta para diferentes combinaciones de (p, q)

3.1.2. Estimación de Parámetros

Los métodos más comunes de estimación del vector de parámetros β son el método de máxima verosimilitud y el método de momentos. Particularmente, se describe la estimación por máxima verosimilitud. Sea y_1, y_2, \dots, y_n una muestra aleatoria de tamaño n con función de densidad $p(y, \theta)$, se define la función de verosimilitud como:

$$L(\theta|y_1, y_2, \dots, y_n) = \prod_{i=1}^n p(y_i|\theta), \quad (3.4)$$

La función de verosimilitud para una muestra observada con función de densidad como en la ecuación (3.1) está dada por:

$$L(p, q|y_i) = \prod_{i=1}^n \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y_i^{p-1} (1-y_i)^{q-1} = \left[\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \right]^n \prod_{i=1}^n y_i^{p-1} \prod_{i=1}^n (1-y_i)^{q-1}. \quad (3.5)$$

Aplicando el logaritmo de la función de verosimilitud dada en la ecuación (3.5) se tiene:

$$\begin{aligned} \ell(p, q) = \log L(p, q|y_i) &= n \left\{ \log \left[\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \right] \right\} + (p-1) \sum_{i=1}^n \log(y_i) + (q-1) \sum_{i=1}^n \log(1-y_i) \\ &= n[\log \Gamma(p+q) - \log \Gamma(p) - \log \Gamma(q)] + (p-1) \sum_{i=1}^n \log(y_i) \\ &\quad + (q-1) \sum_{i=1}^n \log(1-y_i). \end{aligned} \quad (3.6)$$

Las ecuaciones de estimación o función score se obtienen derivando la ecuación (3.6) respecto a cada parámetro e igualando a cero como:

$$\begin{aligned} \frac{\partial \ell(p, q)}{\partial p} &= n(\psi(p+q) - \psi(p)) + \sum_{i=1}^n \log(y_i) = 0, \\ \frac{\partial \ell(p, q)}{\partial q} &= n(\psi(p+q) - \psi(q)) + \sum_{i=1}^n \log(1-y_i) = 0. \end{aligned}$$

Despejando, se tiene

$$\begin{aligned} \psi(p) - \psi(p+q) &= \frac{1}{n} \sum_{i=1}^n \log(y_i), \\ \psi(q) - \psi(p+q) &= \frac{1}{n} \sum_{i=1}^n \log(1-y_i), \end{aligned}$$

donde $\psi(\cdot)$ es la función digamma y está definida como:

$$\psi(z) = \frac{\partial \log \Gamma(z)}{\partial z} = \frac{\Gamma'(z)}{\Gamma(z)}, \quad z > 0. \quad (3.7)$$

La matriz de información de Fisher para p y q se obtiene hallando las segundas derivadas del logaritmo de la función de verosimilitud dada en la ecuación (3.6) respecto a cada parámetro. Esta matriz puede ser escrita como:

$$K(p, q) = n \begin{pmatrix} \psi'(p) - \psi'(p+q) & -\psi'(p+q) \\ -\psi'(p+q) & \psi'(q) - \psi'(p+q) \end{pmatrix} \quad (3.8)$$

donde $\psi'(\cdot)$ es la función trigamma y está definida como :

$$\psi'(z) = \frac{\partial \psi(z)}{\partial z} = \frac{\partial^2 \log \Gamma(z)}{\partial z^2} = \frac{\Gamma''(z)\Gamma(z) - [\Gamma'(z)]^2}{[\Gamma(z)]^2}, \quad z > 0. \quad (3.9)$$

Cabe resaltar que las ecuaciones de máxima verosimilitud de la distribución beta no tienen forma cerrada y no es sencillo encontrar soluciones analíticas a estas ecuaciones; debido a esto, los estimadores de los parámetros son obtenidos por medio de métodos numéricos iterativos de maximización como el método de Newton Raphson.

3.2. Modelo de Regresión Beta Lineal Generalizada

3.2.1. Introducción

En muchas situaciones es de interés identificar el comportamiento de una variable en función de una u otras variables. En este tipo de casos, los modelos de regresión son comúnmente usados para identificar una posible relación entre una variable dependiente y una o más variables independientes. En el contexto del uso de la distribución beta para situaciones en las que la variable respuesta se distribuye continuamente en el intervalo (0,1) y que además puede ser explicada por otras variables a través de la estructura de regresión, se encuentra el modelo de regresión beta lineal generalizado propuesto por Ferrari y Cribari-Neto (2004). En esta propuesta se realizó una re-parametrización de la función de densidad de la distribución beta dada en la ecuación (3.1). Estos autores indican que

este enfoque tiene ventajas deseables: primero, no se necesita transformar la variable dependiente, lo que permite una interpretación directa de los coeficientes de regresión sobre el valor esperado de la respuesta; segundo, la varianza de la variable de respuesta está en función de la media, por lo que no se impone el supuesto de homoscedasticidad y finalmente, los parámetros de la distribución beta permiten modelar diferentes formas de la distribución dependiendo del valor que le asigne a dichos parámetros, en especial aquellas que muestran algún tipo de asimetría.

Ferrari y Cribari-Neto (2004) manifiestan que la distribución beta no pertenece a la familia exponencial uni-parámetrica debido a que su función de densidad no puede ser escrita en forma canónica con un parámetro θ . Sin embargo, está relacionada a los GLM para modelos conjuntos de media y dispersión. Esto permite aplicar técnicas desarrolladas para los GLM como pruebas de bondad de ajuste y análisis de residuales. En forma más detallada McCullagh y Nelder (1989) y Agresti (2002) desarrollan conceptos y procedimientos correspondientes a la familia exponencial y a los GLM.

3.2.2. Parametrización

Para definir el modelo de regresión beta lineal, Ferrari y Cribari-Neto (2004) propusieron un modelo para variables aleatorias distribuidas beta realizando, una parametrización diferente de la función de densidad de la distribución beta dada en la ecuación (3.1) para obtener así una estructura de regresión en términos de la media y un parámetro de precisión (dispersión) constante.

Considerando $\mu = \frac{p}{p+q}$ y $\phi = p+q$, se obtiene que $p = \mu\phi$ y $q = \phi(1-\mu)$. De esta forma, la media y la varianza dadas en las ecuaciones (3.2) y (3.3) respectivamente, quedan determinadas bajo la nueva parametrización como:

$$E(y) = \frac{\mu\phi}{\phi} = \mu, \quad y \quad Var(y) = \frac{\mu(1-\mu)}{\phi+1},$$

Donde μ es la media de la variable respuesta, ϕ es interpretado como el parámetro de precisión, en el sentido en que, para valores fijos de μ , la varianza se hace mas pequeña cuando éste se hace más

grande y $\mu(1 - \mu)$ es una función de varianza. De esta manera, la función de densidad dada en (3.1) queda determinada bajo la nueva parametrización como:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma[(1 - \mu)\phi]} y^{\mu\phi-1} (1 - y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad 0 < \mu < 1, \quad \phi > 0, \quad (3.10)$$

y el logaritmo de la función de densidad como:

$$\log f(y; \mu, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu\phi) - \log \Gamma[(1 - \mu)\phi] + (\mu\phi - 1) \log(y) + [(1 - \mu)\phi - 1] \log(1 - y)$$

La función de densidad dada en (3.10) origina gráficas de diferentes formas dependiendo de los valores de los parámetros μ y ϕ . En particular, es simétrica cuando $\mu = 1/2$. En la Figura 3.2 se muestra las densidades de esta distribución bajo diferentes combinaciones de (μ, ϕ) . Una distribución uniforme se obtiene haciendo $\mu = 0.5$ y $\phi = 2$.

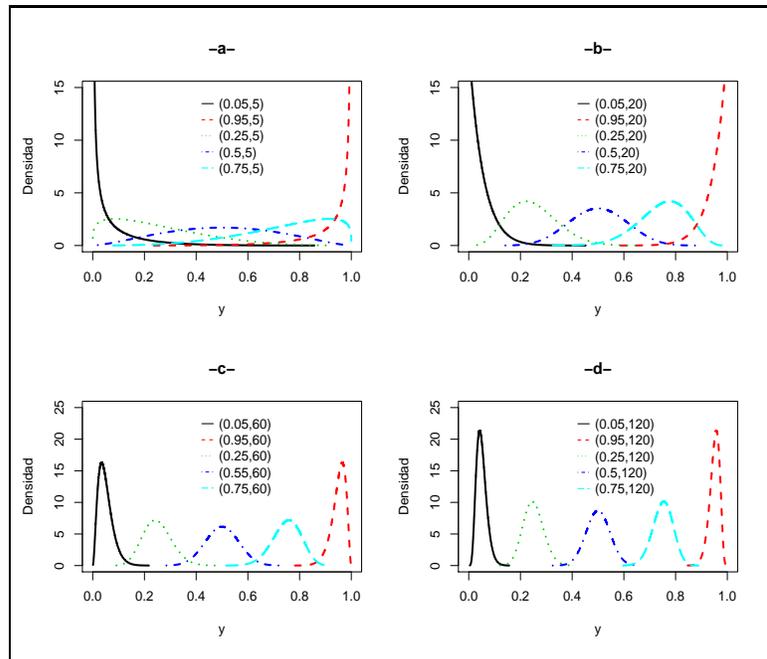


Figura 3.2: Función de densidad beta para diferentes combinaciones de (μ, ϕ)

Aunque se asume que la variable respuesta está restringida al intervalo $(0, 1)$, el modelo propuesto

por Ferrari y Cribari-Neto (2004) puede ser utilizado en formas más generales, por ejemplo, cuando la variable respuesta está restringida a un intervalo (a, b) , donde a y b son constantes conocidas, con $(a < b)$. En este caso, se modela $(y - a)/(b - a)$ en vez de modelar a y directamente.

3.2.3. Definición

Considerando y_1, y_2, \dots, y_n variables aleatorias independientes tal que cada y_t ($t = 1, 2, \dots, n$) sigue una función de densidad como en la ecuación (3.10) con media μ_t y parámetro de precisión desconocido y constante para todo t , se obtuvo el modelo admitiendo que la media de y_t podría ser escrita como

$$g(\mu_t) = x_t^T \boldsymbol{\beta} = \eta_t \quad (3.11)$$

o consecuentemente,

$$\mu_t = g^{-1}(\eta_t), \quad (3.12)$$

siendo $\eta_t = x_t^T \boldsymbol{\beta}$ el predictor lineal, donde $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)$ es el vector de parámetros desconocidos, $\boldsymbol{\beta} \in \mathbb{R}^k$; $x_t^T = (x_{t1}, x_{t2}, \dots, x_{tk})$ son las observaciones de las k covariables ($k < n$), las cuales son asumidas conocidas y fijas y $g(\cdot)$ es una función de enlace. La varianza de y_t está en función de la media y como consecuencia de los valores de las variables explicativas. Cabe resaltar que las variables respuesta y_t con varianzas no constantes son naturalmente acomodadas en el modelo.

3.2.4. Función de Enlace

Una función de enlace $g(\cdot)$ relaciona el predictor lineal η con el valor esperado de la variable respuesta, $E(y_t|x_t) = \mu_t$, a través de la función $g(\mu_t) = \eta_t$. La función g es conocida, estrictamente monótona y dos veces diferenciable. Como consecuencia, la media μ_t se obtiene fácilmente de la inversa de $g(\mu_t)$ como: $\mu_t = g^{-1}(\eta_t)$. Cabe resaltar que existen muchas posibilidades para su elección. Sin embargo,

como en los modelos de regresión beta la variable respuesta está definida en el intervalo continuo (0,1) y por ende su media también, se hace un énfasis especial a las funciones que realizan la transformación del intervalo (0,1) a \mathbb{R} , como:

- Función Logit:

$$g_1(\mu_t) = \log\left(\frac{\mu_t}{1 - \mu_t}\right); \quad \mu_t = \frac{\exp(\eta_t)}{1 + \exp(\eta_t)}.$$

- Función Probit:

$g_2(\mu_t) = \Phi^{-1}(\mu_t)$, donde $\Phi^{-1}(\cdot)$ es la función de distribución acumulada de una variable aleatoria que se distribuye normal; $\mu_t = \Phi(\eta_t)$.

- Función Log-Log :

$$g_3(\mu_t) = -\log[-\log(\mu_t)]; \quad \mu_t = \exp[-\exp(-\eta_t)].$$

- Función Log-Log Complementario:

$$g_4(\mu_t) = \log[-\log(1 - \mu_t)]; \quad \mu_t = 1 - \exp[-\exp(\eta_t)].$$

Las cuatro funciones se pueden obtener como las inversas de funciones de distribución acumulada conocidas teniendo como soporte todo el eje real. Las dos primeras funciones son simétricas en el sentido que $g_1(\mu_t) = -g_1(1 - \mu_t)$. Las dos últimas no son simétricas en este sentido, pero están relacionadas por $g_4(\mu_t) = -g_3(1 - \mu_t)$. Las cuatro funciones son continuas y crecientes en el intervalo (0,1). Adicionalmente, McCullagh y Nelder (1989, sec.4.3.1) realizaron una comparación gráfica de la función logit contra las funciones probit, log- log complementario y log-log para valores de μ_t en el rango 0.01 a 0.99.

Por lo general se elige la función de enlace que mejor establezca la relación entre el predictor lineal y la media de la variable dependiente. Sin embargo, Ferrari y Cribari-Neto (2004) afirman que en

el modelamiento de datos medidos como una proporción, estas cuatro funciones pueden presentar resultados similares. En este contexto la más usada es la logit debido a que facilita la interpretación de los parámetros de regresión. De la misma manera bajo la técnica de cuasi-verosimilitud, [Cox \(1996\)](#) encontró que la función de enlace logit resultó mejor que las otras funciones consideradas en el ajuste de los modelos de regresión. Para más detalles sobre la función logit, ver [Agresti \(2002, sec.4.2.3\)](#). A continuación se describe la función logit en detalle

$$g(\mu_t) = \log\left(\frac{\mu_t}{1 - \mu_t}\right) = x_t^T \boldsymbol{\beta}, \quad t = 1, 2, \dots, n.$$

Análogamente puede ser escrita como

$$\frac{\mu_t}{1 - \mu_t} = \exp(x_t^T \boldsymbol{\beta}),$$

de lo que se obtiene

$$\mu_t = \frac{\exp(x_t^T \boldsymbol{\beta})}{1 + \exp(x_t^T \boldsymbol{\beta})}$$

donde $x_t^T = (x_{t1}, x_{t2}, \dots, x_{tk})$. Esta expresión es la función inversa de $g(\mu_t)$. Cuando la función logit es usada, los parámetros de regresión pueden ser interpretados en términos del cociente de chances (odds ratio). Suponiendo que el valor de la k -ésima variable regresora ha aumentado c unidades, mientras que las demás variables independientes permanecen constantes, y sea μ^\dagger la media de y bajo el nuevo valor de la covariable mientras que μ es la media de y bajo los valores originales de las covariables, se obtiene

$$\frac{\mu^\dagger}{1 - \mu^\dagger} = \exp[x_{t1}\beta_1 + x_{t2}\beta_2 + \dots + (x_{ti+} + c)\beta_i + \dots + x_{tk}\beta_k],$$

así, el cociente de chances queda determinado por:

$$\exp(c\beta) = \frac{\frac{\mu^\dagger}{1 - \mu^\dagger}}{\frac{\mu}{1 - \mu}}.$$

3.2.5. Estimación de Parámetros

Ferrari y Cribari-Neto (2004) proponen expresiones para maximizar la función de verosimilitud y encontrar los estimadores máximo verosímiles de los parámetros β y ϕ por medio de maximización numérica, usando un proceso iterativo como algún algoritmo de optimización no lineal. Algunos de estos algoritmos son el método de Newton (Newton Raphson, Fisher Scoring, etc.) y el algoritmo de quasi-Newton (BFGS). Este proceso de estimación ya se encuentra implementado en el paquete *betareg* disponible en el software estadístico R (R-Team, 2010).

Una función de log-verosimilitud basada en una muestra de n observaciones independientes está dada por:

$$\ell(\beta, \phi) = \sum_{t=1}^n \ell_t(\mu_t, \phi), \quad (3.13)$$

donde,

$$\begin{aligned} \ell_t(\mu_t, \phi) = & \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma[(1 - \mu_t)\phi] + (\mu_t \phi - 1) \log(y_t) \\ & + [(1 - \mu_t)\phi - 1] \log(1 - y_t) \end{aligned} \quad (3.14)$$

y $\mu_t = g^{-1}(\eta_t)$ como está definido en la ecuación (3.12).

La función score se obtiene derivando la función de log-verosimilitud respecto a los parámetros desconocidos e igualando a cero. De la ecuación (3.14) se obtiene

$$\frac{\partial \ell(\beta, \phi)}{\partial \beta_k} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_k}, \quad (3.15)$$

donde $\frac{d\mu_t}{d\eta_t} = \frac{1}{g'(\mu_t)}$, $\frac{\partial\eta_t}{\partial\beta_k} = x_{tk}$ y de la ecuación (3.13) se obtiene

$$\frac{\partial\ell_t(\mu_t, \phi)}{\partial\mu_t} = \phi \left\{ \log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t\phi) - \psi((1-\mu_t)\phi)] \right\} \quad (3.16)$$

donde $\psi(\cdot)$ es la función digamma definida en la ecuación (3.7). Definiendo $y_t^* = \log\left(\frac{y_t}{1-y_t}\right)$ y $\mu_t^* = \psi(\mu_t\phi) - \psi((1-\mu_t)\phi)$, la expresión dada en (3.16) puede ser escrita como:

$$\frac{\partial\ell_t(\mu_t, \phi)}{\partial\mu_t} = \phi(y_t^* - \mu_t^*), \quad (3.17)$$

recordando que el valor esperado de las derivadas de la ecuación (3.13) se igualan a cero, entonces $\mu_t^* = E(y_t^*)$. Por lo tanto, la ecuación dada en (3.15) queda determinada por:

$$\frac{\partial\ell(\boldsymbol{\beta}, \phi)}{\partial\beta_k} = \phi \sum_{t=1}^n (y_t^* - \mu_t^*) \frac{1}{g'(\mu_t)} x_{tk}. \quad (3.18)$$

De la misma forma, la derivada del logaritmo de la función de verosimilitud respecto a ϕ está dada por:

$$\frac{\partial\ell(\boldsymbol{\beta}, \phi)}{\partial\phi} = \sum_{t=1}^n \frac{\partial\ell_t(\mu_t, \phi)}{\partial\phi}, \quad (3.19)$$

donde

$$\frac{\partial\ell_t(\mu_t, \phi)}{\partial\phi} = \mu_t \left[\log\left(\frac{y_t}{1-y_t}\right) - \psi(\mu_t\phi) + \psi[(1-\mu_t)\phi] \right] + \log(1-y_t) - \psi[(1-\mu_t)\phi] + \psi(\phi). \quad (3.20)$$

Por lo tanto una *función score* para el parámetro $\boldsymbol{\beta}$, o vector columna $U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi)$ de dimensión k , puede ser escrita de forma matricial como:

$$U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi) = \phi X^T T(y^* - \mu^*), \quad (3.21)$$

donde X es una matriz de dimensión $n \times k$ cuya t -ésima fila es x_t^T , $T = \text{diag} \left\{ \frac{1}{g'(\mu_1)}, \frac{1}{g'(\mu_2)}, \dots, \frac{1}{g'(\mu_n)} \right\}$, $y^* = (y_1^*, y_2^*, \dots, y_n^*)^T$ y $\mu^* = (\mu_1^*, \mu_2^*, \dots, \mu_n^*)^T$ y para el parámetro de precisión ϕ se tiene:

$$U_\phi(\boldsymbol{\beta}, \phi) = \sum_{t=1}^n \{ \mu_t (y_t^* - \mu_t^*) + \log(1 - y_t) - \psi[(1 - \mu_t)\phi] + \psi(\phi) \}. \quad (3.22)$$

siendo $U_\phi(\boldsymbol{\beta}, \phi)$ un escalar.

Para determinar la variabilidad de las estimaciones de los parámetros del modelo de regresión beta, Ferrari y Cribari-Neto (2004) obtuvieron la matriz de información de Fisher como se muestra a continuación:

Sea $W = \text{diag}(w_1, w_2, \dots, w_n)$, con

$$w_t = \phi \{ \psi'(\mu_t \phi) + \psi'[(1 - \mu_t)\phi] \} \frac{1}{[g'(\mu_t)]^2},$$

$c = (c_1, c_2, \dots, c_n)^T$, con $c_t = \phi \{ \psi'(\mu_t \phi) \mu_t - \psi'[(1 - \mu_t)\phi] (1 - \mu_t) \}$, donde $\psi'(\cdot)$ es la función trigamma definida como en (3.9); además, $D = \text{diag}(d_1, d_2, \dots, d_n)$, donde $d_t = \psi'(\mu_t \phi) \mu_t^2 + \psi'((1 - \mu_t)\phi) (1 - \mu_t)^2 - \psi'(\phi)$.

Por lo tanto, la matriz de información de Fisher puede ser escrita como:

$$K = K(\boldsymbol{\beta}, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}$$

donde $K_{\beta\beta} = \phi X^T W X$, $K_{\beta\phi} = K_{\phi\beta}^T = X^T T c$ y $K_{\phi\phi} = \text{tr}(D)$, donde $\text{tr}(\cdot)$ denota la traza de una matriz cuadrada. A diferencia de lo que sucede en los Modelos Lineales Generalizados, los parámetros $\boldsymbol{\beta}$ y ϕ no son ortogonales, es decir, $K_{\beta\phi} \neq 0$.

3.3. Modelo de Regresión Beta No Lineal

Como extensión al modelo de regresión beta lineal dado en la ecuación (3.10) y teniendo en cuenta lo expuesto en la Sección 2.5, se considera el modelo de regresión beta no lineal. Para este modelo, la expresión dada en (3.12) puede ser escrita como:

$$\mu_t = \mathbf{f}(x_t^T, \boldsymbol{\beta}), \quad (3.23)$$

donde \mathbf{f} es una función no lineal en los parámetros, los parámetros $\boldsymbol{\beta} \in \mathbb{R}^k$ y $\phi > 0$ es constante.

La estimación de los parámetros de regresión del modelo beta no lineal se realiza de forma similar a como se desarrolló en la Sección 3.2.5. La diferencia más importante está en que en este modelo la relación dada en la ecuación (3.11) no se satisface debido a la no linealidad en los parámetros.

Sea $\mu_t = \mathbf{f}(x_t^T, \boldsymbol{\beta})$. El logaritmo de la función de verosimilitud dada en la ecuación (3.13) queda definido como :

$$\ell(\boldsymbol{\beta}, \phi) = \sum_{t=1}^n \ell_t(\mathbf{f}(x_t^T, \boldsymbol{\beta}), \phi), \quad (3.24)$$

así,

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_k} = \sum_{t=1}^n \frac{\partial \ell_t(\mathbf{f}(x_t^T, \boldsymbol{\beta}), \phi)}{\partial \mathbf{f}(x_t^T, \boldsymbol{\beta})} \frac{\partial \mathbf{f}(x_t^T, \boldsymbol{\beta})}{\partial \beta_k}, \quad (3.25)$$

donde

$$\frac{\partial \ell_t(\mathbf{f}(x_t^T, \boldsymbol{\beta}), \phi)}{\partial \mathbf{f}(x_t^T, \boldsymbol{\beta})} = \phi(y_t^* - \mathbf{f}(x_t^T, \boldsymbol{\beta})^*), \quad (3.26)$$

por lo tanto,

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_k} = \sum_{t=1}^n \phi(y_t^* - \mathbf{f}(x_t^T, \boldsymbol{\beta})^*) \frac{\partial \mathbf{f}(x_t^T, \boldsymbol{\beta})}{\partial \beta_k}. \quad (3.27)$$

Respecto al parámetro ϕ , la derivada del logaritmo de la función de verosimilitud está dada por:

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \phi} = \sum_{i=1}^n \frac{\partial \ell_i(\mathbf{f}(x_i^T, \boldsymbol{\beta}), \phi)}{\partial \phi}. \quad (3.28)$$

Las ecuaciones de estimación son en general no lineales y tienen que ser resueltas iterativamente. Cabe resaltar que la complejidad de la estimación va a estar ligada al término $\partial \mathbf{f}(x_i^T, \boldsymbol{\beta}) / \partial \beta_k$.

3.4. Modelo de Regresión Beta No Lineal Mixto

3.4.1. Introducción

Existen muchas situaciones en las que el modelo para la media, además de sus efectos fijos (el mismo para todas las observaciones), incorpora efectos aleatorios (dependen de los individuos o sub-grupos de la población) que son normalmente distribuidos (aunque no siempre). Así, el modelo tiene dos tipos de parámetros: parámetros globales que corresponden a los efectos fijos y parámetros que varían con la población, asociados a los efectos aleatorios. Las curvas de crecimiento no lineal y datos binomiales con sobre-dispersión son dos casos comunes. Por ejemplo, es posible que algunos de los parámetros de una curva de crecimiento no lineal varíen de sujeto a sujeto, permitiendo así reflejar esta variabilidad mediante la incorporación de efectos aleatorios. En este contexto, la función que se está modelando expresa la esperanza condicional de la variable respuesta y dado el efecto aleatorio \mathbf{b} , es decir, $E(y|\mathbf{b})$. Esta formulación permite incorporar uno o varios efectos aleatorios y no está limitada a datos con distribución normal. La presencia de efectos aleatorios origina correlación entre las observaciones provenientes de la misma unidad, la cual varía dependiendo de los valores de x involucrados ([Macchiavelli, 2010](#)).

3.4.2. Definición del Modelo

Sea $y_t = (y_{t1}, y_{t2}, \dots, y_{tm})$ un vector con n_t observaciones del t -ésimo sujeto, se define el modelo no lineal mixto beta

$$y_{tj} | \mathbf{b}_t \sim \text{beta}(\mu_{tj}, \phi), \quad \mu_{tj} = E(y_{tj} | x_{tj}, \mathbf{b}_t) = \mathbf{f}_t(x_{tj}, \boldsymbol{\beta}, \mathbf{b}_t), \quad t = 1, 2, \dots, m, \quad j = 1, 2, \dots, n_t, \quad (3.29)$$

donde \mathbf{b}_t es el vector de efectos aleatorios independientes y se distribuye

$$\mathbf{b}_t \sim \mathbf{N}(0, \mathbf{G}), \quad (3.30)$$

\mathbf{G} es la matriz de varianzas-covarianzas de los efectos aleatorios y $\boldsymbol{\beta}$ es el vector de parámetros de efectos fijos.

Al integrar a través de la distribución de los efectos aleatorios \mathbf{b}_t se obtiene la esperanza marginal inducida dada por:

$$E(y_{tj} | x_{tj}) = \int \mathbf{f}_t(x_{tj}, \boldsymbol{\beta}, \mathbf{b}_t) q(\mathbf{b}_t) d\mathbf{b}_t, \quad (3.31)$$

donde $q(\mathbf{b}_t)$ es la distribución de los efectos aleatorios.

3.4.3. Estimación de Parámetros

El principal objetivo en el análisis de un modelo no lineal mixto es la estimación de $\boldsymbol{\beta}$ y \mathbf{G} . Esta estimación se realiza en base a la maximización de la función de densidad marginal (Lindstrom y Bates, 1990; Davidian y Giltinan, 1995; Pinheiro y Bates, 2000). La función de verosimilitud está definida como:

$$L(\boldsymbol{\beta}, \mathbf{G}; y_{ij}) = \prod_{t=1}^m \int \prod_{j=1}^{n_t} p(y_{tj} | \mathbf{b}_t; \boldsymbol{\beta}) q(\mathbf{b}_t, \mathbf{G}) d\mathbf{b}_t, \quad (3.32)$$

donde p es una densidad de la distribución beta, q es una densidad de la distribución normal y \mathbf{G} es la matriz de varianzas-covarianzas de los efectos aleatorios \mathbf{b}_t de dimensión $k \times k$.

Encontrar la estimación máximo verosímil significa resolver las ecuaciones score para $\boldsymbol{\beta}$ y \mathbf{G} . Recordemos que la integral dada en la ecuación (3.32) se realiza a través de la distribución de los efectos aleatorios \mathbf{b}_t , lo que hace necesario aplicar una estrategia de aproximación numérica ya que en general se hace imposible resolverlas analíticamente. Sin embargo, cuando el número de efectos aleatorios es muy grande, la complejidad computacional aumenta considerablemente. La solución a este problema ha sido estudiada bajo diferentes aproximaciones. Algunos métodos consisten en tomar la expansión de Taylor de primer orden alrededor del valor esperado del efecto aleatorio o usar cuadratura Gaussiana.

Pinheiro y Bates (2000, sec:7.2.1) describen en forma detallada tres métodos para la aproximación de la función de verosimilitud en los modelos no lineales mixtos. El primero, el Algoritmo Lindstrom y Bates, aproxima la función de densidad marginal por la verosimilitud de un modelo lineal de efectos mixtos. En el segundo método se utiliza una aproximación Laplaciana de la función de verosimilitud y en el tercero, se usa la cuadratura Gaussiana adaptativa que mejora la aproximación Laplaciana.

El método de aproximación por defecto del procedimiento NLMIXED de SAS es el de cuadratura Gaussiana. A continuación se describe brevemente el método de aproximación Cuadratura Gaussiana Adaptativa.

Cuadratura Gaussiana Adaptativa

El método de cuadratura Gaussiana aproxima una integral dada por medio de una suma ponderada sobre las abscisas (coordenada horizontal en un eje de coordenadas) definidas por los efectos aleatorios. Una buena aproximación se puede obtener con un número adecuado de puntos de cuadratura o la

ampliación de las abcisas. La cuadratura Gaussiana adaptativa para la integral sobre los b_t se centra en la integral de las estimaciones empíricas de Bayes de los b_t , definida como el vector \widehat{b}_t que minimiza

$$- \log[p(y_{tj}|x_{tj}, \boldsymbol{\beta}, b_t)q(b_t, \mathbf{G})]$$

con $\boldsymbol{\beta}$ y \mathbf{G} igual al conjunto de sus estimaciones actuales. Supongamos que z_q y w_q , $q = 1, 2, \dots, Q$, denota las abcisas y los pesos estándar de Gauss-Hermite y θ el vector de parámetros. De esta manera, la aproximación de la integral por medio de la cuadratura Gaussiana adaptativa está dada por

$$\int p(y_{tj}|x_{tj}, \boldsymbol{\beta}, b_t)q(b_t, \mathbf{G})db_t \approx 2^{r/2}|\Gamma(x_{tj}, \theta)|^{-1/2} \sum_{q1=1}^Q \dots \sum_{qr=1}^Q \left[p(y_{tj}|x_{tj}, \boldsymbol{\beta}, a_{q1, \dots, qr})q(a_{q1, \dots, qr}|\mathbf{G}) \prod_{k=1}^r w_{qk} \exp z_{qk}^2 \right]$$

donde r es la dimensión de b_t , $\Gamma(x_{tj}, \theta)$ es la matriz Hessiana ¹ desde la minimización empírica de Bayes, $z_{q1, q2, \dots, qr}$ es un vector con elementos $z_{q1}, z_{q2}, \dots, z_{qr}$ y $a_{q1, \dots, qr} = \widehat{b}_t + \sqrt{2}\Gamma(x_{tj}, \theta)^{-1/2}z_{q1, q2, \dots, qr}$.

¹ **Matriz Hessiana:** La matriz de segundas derivadas parciales se conoce como la matriz Hessiana. La inversa de la matriz Hessiana se aproxima a la matriz de varianzas-covarianzas de los parámetros estimados. Intuitivamente debe haber una relación entre las segundas derivadas de los parámetros y sus errores estándar. Si el cambio de la pendiente es muy fuerte en torno al mínimo entonces las derivadas de segundo orden serán muy grandes. Sin embargo las estimaciones de los parámetros serán muy estables en el sentido de que el mínimo con respecto al parámetro es claramente identificable. La matriz Hessiana y los errores estándar asintóticos para los parámetros se pueden calcular a través de aproximaciones.

Simulación del Modelo Propuesto

En este capítulo estudiamos el modelo de regresión beta no lineal mixto detallado en la Sección 3.4 y lo comparamos con un modelo de regresión no lineal mixto asumiendo distribución normal para la variable dependiente. El proceso se realizó mediante simulación para evaluar las estimaciones de los parámetros (máxima verosimilitud), momentos marginales y valores predichos de la media condicional para cuando los efectos aleatorios son cero.

4.1. Proceso de Simulación

Para estudiar las propiedades del modelo beta no lineal mixto, se considera la distribución marginal inducida de las observaciones del sujeto t como:

$$\int \prod_{j=1}^{n_t} \left\{ \frac{\Gamma(\phi)}{\Gamma(\mu_{tj}\phi)\Gamma\{(1-\mu_{tj})\phi\}} y_{tj}^{\mu_{tj}\phi-1} (1-y_{tj})^{(1-\mu_{tj})\phi-1} \right\} \left\{ \frac{\exp\left(-\frac{1}{2}\mathbf{b}'_t \mathbf{G}^{-1} \mathbf{b}_t\right)}{|\mathbf{G}|^{\frac{1}{2}} (2\pi)^{\frac{k}{2}}} \right\} d\mathbf{b}_t. \quad (4.1)$$

Los escenarios de simulación planteados para el proceso de simulación son:

- Modelo de regresión no lineal mixto con efecto aleatorio de intercepto.
- Modelo de regresión no lineal mixto con efecto aleatorio de pendiente.

Para la generación de datos distribuidos beta (datos entre 0 y 1) se fijaron los parámetros $\beta_0 = -1.22$, $\beta_1 = 0.08$ y $\phi = 10$ y $\phi = 15$. Estos parámetros son aproximadamente los que corresponde a un GLM con los datos del problema de la aplicación, que se estudiarán en detalle en el Capítulo 5. El proceso se realizó para 10 sujetos, cada uno con 11 mediciones. En cada escenario se usa la curva logística para media del modelo, variando únicamente en la parte donde se adiciona el o los efectos aleatorios. En la Figura 4.1 se muestra la curva con efecto aleatorio cero que se usará como base para el proceso de simulación. En el eje horizontal vemos las 11 mediciones iniciando en el tiempo 0 y finalizando en el tiempo 50, en intervalos de 5 en 5. En el eje vertical se observa la variable dependiente y que se distribuye beta.

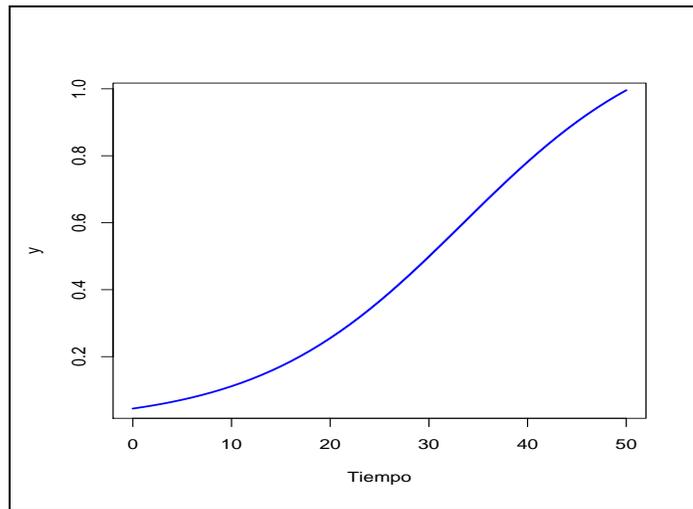


Figura 4.1: Curva logística base para el proceso de simulación

El proceso de simulación se llevó a cabo en el software estadístico SAS (v. 9.1.3). Los efectos aleatorios fueron generados con la función **rannor** y la variable dependiente y con la función **rand('beta', p, q)**, donde p y q son los parámetros de la distribución. Para nuestro modelo se usa la parametrización $p = \mu\phi$ y $q = \phi(1 - \mu)$ dada por Ferrari y Cribari-Neto (2004). Luego de tener los datos generados en cada uno de los escenarios se utilizó el procedimiento NLMIXED para ajustar los modelos correspondientes asumiendo distribución beta o normal para la variable dependiente y. Estos modelos de explicarán más en detalle en las siguientes secciones.

En cada modelo ajustado se estudian las siguientes propiedades: las distribuciones condicionales se analizan particularmente para cuando los efectos aleatorios son cero y se calcula su media. En las distribuciones marginales se calcula la media, la desviación estándar, el cuartil 25, la mediana, el cuartil 75, el percentil 95 y la asimetría. Finalmente, se hacen comparaciones relevantes entre las medias marginales, medias condicionales, mediana y asimetría. A continuación se presenta el algoritmo utilizado para las simulaciones en general.

Algoritmo

1. Generar datos.

1.1 Inicia ciclo para cada sujeto específico $t = 1, 2, 3, \dots, 10$.

i. Generar el efecto aleatorio para cada sujeto, con $\mathbf{b}_t \sim N(0, \sigma_{\mathbf{b}})$.

ii. Crear la variable fija $tiempo_j = 0, 5, 10, 15, \dots, 50$.

iii. Calcular la media condicional μ_{tj} .

iv. Calcular los parámetros de la distribución beta p_{tj} y q_{tj} .

v. Generar la variable aleatoria $y_{tj} \sim beta(\mu_{tj}, \phi)$.

1.2 Fin del ciclo.

2. Análisis de datos generados.

2.1 Estimación de parámetros.

2.2 Calcular el valor predicho de la media condicional cuando el efecto aleatorio es cero.

3. Repetir los pasos 1 y 2, 1000 veces.

4. Estimar parámetros de interés de las distribuciones marginales: estimación de los parámetros, media, desviación estándar, cuartiles, percentiles y asimetría.

4.2. Descripción Escenarios de Simulación

4.2.1. Modelo de Regresión No Lineal Mixto con Efecto Aleatorio de Intercepto

Al modelo de regresión no lineal mixto le adicionamos el efecto aleatorio al intercepto, $\mathbf{b}_t = b_{t0}$, el cual indica que existe una variabilidad entre sujetos que debe ser explicada por el efecto aleatorio asociado a cada sujeto. Una vez generados los datos se realiza el ajuste asumiendo distribución beta o normal para la variable respuesta y . Así, el modelo de regresión no lineal mixto beta con efecto aleatorio de intercepto se define como:

$$y_{tj}|b_{t0} \sim \text{beta}(\mu_{tj}, \phi), \quad (4.2)$$

y el modelo de regresión no lineal mixto normal se define como:

$$y_{tj}|b_{t0} \sim \mathbf{N}(\mu_{tj}, \sigma_\varepsilon^2), \quad (4.3)$$

donde la variable y_{tj} para los dos modelos tiene media condicional μ_{tj} :

$$\mu_{tj} = \frac{\exp(\beta_0 + b_{t0} + \beta_1 * \text{tiempo}_{tj})}{1 + \exp(\beta_0 + b_{t0} + \beta_1 * \text{tiempo}_{tj})} \quad (4.4)$$

y distribución para el efecto aleatorio b_{t0} :

$$b_{t0} \sim \mathbf{N}(0, \sigma_{b_0}^2). \quad (4.5)$$

Los valores iniciales de los parámetros β_0 y β_1 fueron calculados realizando una regresión lineal generalizada (con enlace logit) aplicando un GLM con el procedimiento GLIMMIX de SAS (v.9.1.3). Los parámetros de dispersión (ϕ para la beta y σ_ε para la normal) y todas las varianzas fueron para-

metrizadas en la escala \log para evitar problemas de acotamiento como:

- $\phi = \exp(-\log\phi_i)$, donde $\log\phi_i = \ln(\text{escala})$. El valor inicial para el parámetro de escala es obtenido de la regresión lineal aplicando un GLM, tal como lo sugiere Smithson y Verkuilen (2006).
- $\sigma_\varepsilon = \exp(\log\sigma_\varepsilon)$
- $\sigma_{b_0} = \exp(\log\sigma_{b_0})$

cuyos valores iniciales para $\log\sigma_\varepsilon$ y $\log\sigma_{b_0}$ fueron asignados teniendo en cuenta las características de los datos generados, de tal manera que la convergencia del modelo fuese la mejor posible.

Resultados

A continuación se muestran los resultados de las distribuciones marginales y condicionales (cuando el efecto aleatorio $b_0 = 0$) asumiendo distribución beta y distribución normal únicamente para cuando la variable tiempo toma los valores de 0, 15, 30 y 50 (para los demás valores se encontraron resultados similares). El parámetro σ_{b_0} varía entre 0.22 y 0.33.

ϕ	σ_{b_0}	Tiempo	Media condicional $b_0 = 0$	Q_1	Mediana	Media marginal	Q_3	P_{95}	Desviación Estándar	Asimetría
10	0.22	0	0.2274613	0.1971774	0.2259962	0.2296882	0.2579297	0.3165002	0.0495144	0.535829
		15	0.4935456	0.4544633	0.4931779	0.4930897	0.5303330	0.6009313	0.0646820	0.0227182
		30	0.7637634	0.7346818	0.7639293	0.7607756	0.7903838	0.8348579	0.047915	-0.524911
		50	0.9408765	0.9306626	0.9412699	0.9393824	0.9507871	0.9636545	0.0172026	-0.9795504
	0.27	0	0.2292304	0.2006237	0.2288655	0.2312315	0.2578700	0.3140162	0.0476953	0.440334
		15	0.4945070	0.4565935	0.4944309	0.4940924	0.5317328	0.5974020	0.0625889	-0.05544112
		30	0.7632737	0.7343885	0.7637837	0.7605511	0.7901679	0.8321875	0.0467176	-0.5383506
		50	0.9402388	0.9295642	0.9406240	0.9388733	0.9503983	0.9625762	0.0168773	-0.869143
	0.33	0	0.2277453	0.1981986	0.2264182	0.2298110	0.2575424	0.3156954	0.0487925	0.5767187
		15	0.4936157	0.4550854	0.4921177	0.4931784	0.5306698	0.5986391	0.0635843	0.0361417
		30	0.7635357	0.7349448	0.7632440	0.7607395	0.7911708	0.8337099	0.0473115	-0.4968190
		50	0.9406667	0.9301335	0.9411402	0.9392720	0.9508819	0.9636843	0.0171350	-0.895743

Tabla 4.1: Parámetros de las distribuciones marginales y condicionales del modelo de regresión beta no lineal mixto con efecto aleatorio en β_0 para diferentes valores de σ_{b_0} y $\phi = 10$.

ϕ	σ_{b_0}	Tiempo	Media condicional $b_0 = 0$	Q_1	Mediana	Media marginal	Q_3	P_{95}	Desviación Estándar	Asimetría
15	0.22	0	0.2281073	0.1978449	0.2272609	0.2304555	0.2586867	0.3163468	0.0487143	0.498117
		15	0.4951411	0.4545261	0.4945317	0.4947302	0.5342538	0.6054522	0.0649719	0.018783
		30	0.7652408	0.7557376	0.7652888	0.7622191	0.7928807	0.8361301	0.0479026	-0.5059
		50	0.9418830	0.9314953	0.9418437	0.9400782	0.9510694	0.9632204	0.0164987	-0.92381
	0.27	0	0.2281376	0.1985816	0.2275138	0.23052167	0.2592751	0.3167145	0.0489216	0.477575
		15	0.4934711	0.4529912	0.4924705	0.4930902	0.5346718	0.6020584	0.065433	-0.02152
		30	0.7628048	0.7322302	0.7624782	0.7597343	0.7920903	0.8343274	0.048663	-0.5142
		50	0.9403331	0.9296186	0.9406868	0.938712	0.9501023	0.9626943	0.0169041	-0.8820
	0.33	0	0.229040	0.1990352	0.2282922	0.2313884	0.2600501	0.3166635	0.0489122	0.5281
		15	0.4951606	0.4540031	0.4942216	0.4947291	0.5759725	0.6040189	0.065150	0.04067
		30	0.764292	0.7331564	0.7639301	0.7612474	0.7922479	0.8357981	0.0480908	-0.426886
		50	0.9408863	0.930244	0.9410806	0.9393663	0.9506942	0.9633030	0.0166578	-0.80783

Tabla 4.2: Parámetros de las distribuciones marginales y condicionales del modelo de regresión beta no lineal mixto con efecto aleatorio en β_0 para diferentes valores de σ_{b_0} y $\phi = 15$.

ϕ	σ_{b_0}	Tiempo	Media condicional $b_0 = 0$	Q_1	Mediana	Media marginal	Q_3	P_{95}	Desviación Estándar	Asimetría
10	0.22	0	0.2254503	0.1935259	0.2240789	0.2282649	0.2593850	0.3242776	0.0533898	0.5468683
		15	0.4934508	0.4504071	0.4921594	0.4930451	0.5345572	0.6107687	0.0701065	0.0412107
		30	0.7656922	0.7335590	0.7659060	0.7621754	0.7950525	0.8431925	0.0521095	-0.5103675
		50	0.9421200	0.9304343	0.9426160	0.9403832	0.9530830	0.9666462	0.0187882	-0.9844653
	0.27	0	0.2281789	0.1971877	0.2282051	0.2306894	0.2589968	0.3210606	0.0514210	0.445761
		15	0.4943320	0.4526197	0.4942601	0.4939721	0.5346744	0.6073045	0.0675465	-0.0438244
		30	0.7640933	0.7312485	0.7639451	0.7609140	0.7938952	0.8396871	0.0508105	-0.5166609
		50	0.9407115	0.9284792	0.9408802	0.9391168	0.9521407	0.9654919	0.0185970	-0.8118874
	0.33	0	0.2262905	0.1944970	0.2250343	0.2288930	0.2589563	0.3216684	0.052549	0.5484083
		15	0.4935343	0.4514343	0.4930028	0.4931513	0.5350035	0.6076905	0.0687145	0.0259113
		30	0.7649260	0.7324333	0.7646303	0.7616409	0.7938820	0.8406585	0.0514700	-0.487394
		50	0.9414990	0.9295760	0.9418981	0.9398695	0.9525709	0.9668129	0.0188704	-0.8870335

Tabla 4.3: Parámetros de las distribuciones marginales y condicionales del modelo de regresión normal no lineal mixto con efecto aleatorio en β_0 para diferentes valores de σ_{b_0} y $\phi = 10$.

ϕ	σ_{b_0}	Tiempo	Media condicional $b_0 = 0$	Q_1	Mediana	Media marginal	Q_3	P_{95}	Desviación Estándar	Asimetría
10	0.22	0	0.227144	0.195325	0.2265255	0.229954	0.259745	0.322644	0.05159	0.499881
		15	0.495068	0.451727	0.494941	0.494734	0.537271	0.610106	0.068933	0.02892
		30	0.766094	0.73361	0.765702	0.762713	0.795428	0.841236	0.051015	-0.46682
		50	0.942107	0.930939	0.942295	0.940425	0.952644	0.965453	0.01769	-0.87298
	0.27	0	0.227247	0.195073	0.22707	0.230076	0.260763	0.319836	0.051772	0.48145
		15	0.493374	0.449493	0.493053	0.493076	0.536705	0.606813	0.06911	-0.01048
		30	0.763597	0.730556	0.763582	0.760188	0.79482	0.838508	0.51547	-0.50003
		50	0.940838	0.929133	0.941262	0.939102	0.9114	0.964425	0.018047	-0.87112
	0.33	0	0.228465	0.196327	0.22813	0.231223	0.261368	0.321646	0.051624	0.518498
		15	0.495136	0.451164	0.494251	0.494782	0.537217	0.609631	0.068670	0.044692
		30	0.764844	0.731236	0.764347	0.761497	0.79518	0.839563	0.050877	-0.42045
		50	0.941196	0.929529	0.941355	0.939517	0.951781	0.965082	0.017802	-0.78875

Tabla 4.4: Parámetros de las distribuciones marginales y condicionales del modelo de regresión normal no lineal mixto con efecto aleatorio en β_0 para diferentes valores de σ_{b_0} y $\phi = 15$.

A continuación se muestran los resultados de los promedios de los parámetros estimados tanto para el modelo beta no lineal mixto como para el modelo normal no lineal mixto.

Efecto aleatorio en β_0 , $\phi = 10$ y variando σ_{b_0}

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.2343	0.154110	-1.2447	0.157924
β_1	0.0803	0.004472	0.0811	0.005477
<i>Logphi</i>	-2.32487	0.141633	-	-
<i>Logse</i>	-	-	-2.1056	0.0707
Logsb_0	-1.3118	0.46496	-1.2419	0.394575
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	10.2682	1.466120	-	-
σ_e	-	-	0.12256	0.0089
σ_{b_0}	0.26058	0.135388	0.28384	0.11305

Tabla 4.5: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 10$ y $\sigma_{b_0} = 0.22$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.22005	0.151921	-1.226990	0.15446
β_1	0.07991	0.004472	0.08034	0.00547
<i>Logphi</i>	-2.31169	0.141456	-	-
<i>Logse</i>	-	-	-2.10339	0.0707
Logsb_0	-1.34808	0.441021	-1.29695	0.4184
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	10.2660	1.462928	-	-
σ_e	-	-	0.12266	0.00894
σ_{b_0}	0.2498	0.121614	0.27083	0.124498

Tabla 4.6: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 10$ y $\sigma_{b_0} = 0.27$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.2279	0.1528	-1.2368	0.1557
β_1	0.0802	0.0044	0.0808	0.0055
$Logphi$	-2.3263	0.1416	-	-
$Logse$	-	-	-2.1052	0.0708
$Logsb_0$	-1.3335	0.4485	-1.2826	0.4149
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	10.2834	1.4678	-	-
σ_e	-	-	0.1224	0.0089
σ_{b_0}	0.2518	0.1229	0.2754	0.1114

Tabla 4.7: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 10$ y $\sigma_{b_0} = 0.33$.

Efecto aleatorio en β_0 , $\phi = 15$ y variando σ_{b_0}

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.22488	0.136124	-1.23066	0.137076
β_1	0.08035	0.004123	0.08072	0.004690
$Logphi$	-2.73483	0.141067	-	-
$Logse$	-	-	-2.29868	0.070710
$Logsb_0$	-1.35585	0.3850	-1.31524	0.3634
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	15.50201	2.207903	-	-
σ_e	-	-	0.10092	0.00707
σ_{b_0}	0.29935	0.09407	0.27500	0.09348

Tabla 4.8: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 15$ y $\sigma_{b_0} = 0.22$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.22429	0.136930	-1.23117	0.137840
β_1	0.07989	0.004123	0.08025	0.004690
$Logphi$	-2.72287	0.140996	-	-
$Logse$	-	-	-2.29009	0.070724
$Logsb_0$	-1.33747	0.37468	-1.29953	0.368842
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	15.3225	2.171382	-	-
σ_e	-	-	0.101711	0.007211
σ_{b_0}	0.26381	0.093487	0.278620	0.093198

Tabla 4.9: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 15$ y $\sigma_{b_0} = 0.27$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.220182	0.136102	-1.223447	0.136773
β_1	0.080023	0.004123	0.080262	0.004690
$Logphi$	-2.733891	0.141028	-	-
$Logse$	-	-	-2.295201	0.070731
$Logsb_0$	-1.347177	0.385516	-1.307088	0.419240
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	15.51881	2.210692	-	-
σ_e	-	-	0.101195	0.022583
σ_{b_0}	0.26202	0.095655	0.276019	0.094503

Tabla 4.10: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_0 , $\phi = 15$ y $\sigma_{b_0} = 0.33$.

Conclusiones

En las Tablas 4.1 y 4.2 correspondientes al ajuste con la distribución beta se observa lo siguiente:

- La media marginal es mayor a la media condicional en el tiempo cero. Esto se debe a que en el

modelo marginal le estamos adicionando el efecto aleatorio al intercepto el cual se presenta en los primeros valores de la variable tiempo.

- A medida que se aumenta la desviación estándar del efecto aleatorio también aumenta la media marginal en el tiempo cero. Notemos que las desviaciones estándar están relacionadas directamente a la media marginal, de esta manera se espera que esto suceda.
- La asimetría es mayor cuando la desviación estándar del efecto aleatorio es mayor. Sin embargo, disminuye a medida que aumenta el valor de la variable tiempo, esto debido a que se le adicionó un efecto aleatorio de intercepto al modelo.
- Las medias condicionales y marginales se encuentran entre el percentil 25 y el percentil 95, lo cual permite ver que la curva típica o curva mediana (media condicional $b_0 = 0$) y las curvas marginales para cada sujeto están acotadas por las curvas de estos percentiles.

En las Tablas 4.3 y 4.4 correspondientes al ajuste con la distribución normal se observa lo siguiente:

- La media marginal es mayor a la media condicional en el tiempo 0.
- La asimetría es mayor cuando la desviación estándar del efecto aleatorio es mayor. Sin embargo, disminuye a medida que aumenta el valor de la variable tiempo. La distribución normal es simétrica por definición, y como se observa en la tablas se obtuvieron valores que se pueden considerar altos, por lo que este ajuste no parece ser el adecuado para datos de que se encuentran entre 0 y 1.
- Las medias condicionales y marginales se encuentran entre el percentil 25 y el percentil 95.

En las tablas correspondientes a las estimaciones de los parámetros (Tabla 4.5 a Tabla 4.10) se observa que los errores estándar de los parámetros estimados β_0 y β_1 son menores en el ajuste con la distribución beta que con la distribución normal.

Respecto a la convergencia se puede decir que existe diferencia significativa entre las proporciones de convergencia en ambos modelos. Esto se puede verificar con la prueba no paramétrica de McNemar

(Kvam y Vidakovic, 2007). A continuación se muestra la prueba de McNemar sólo para los datos de la simulación correspondiente a la Tabla 4.5 y no hacer muy extensos los resultados. Las demás pruebas muestran resultados similares.

		BETA		
		Converge	No Converge	Total
NORMAL	Converge	881	39	920
	No Converge	19	61	80
	Total	900	100	1000

Tabla 4.11: Prueba de McNemar para la convergencia de los modelos ajustados

La hipótesis que se plantea es:

- H_0 : La proporción de convergencias es igual en ambos modelos.
- H_1 : La proporción de convergencias es diferente en en ambos modelos.

El estadístico se calcula como:

$$\chi^2_{calculado} = \frac{(39 - 19)^2}{(19 + 39)} = 6.8965 \quad (4.6)$$

con una distribución $\chi^2_{1,0.95} = 3.8414$ y un valor $p = 0.0086$. Por lo tanto se rechaza la hipótesis nula con un nivel de significancia del 0.05 y se concluye que sí hay diferencia significativa en la proporción de convergencias de los dos modelos.

Cabe resaltar que a pesar de que el número de convergencias para el modelo que asume distribución normal es mayor, el modelo que asume distribución beta alcanza la convergencia más rápido.

4.2.2. Modelo de Regresión No Lineal Mixto con Efecto Aleatorio de Pendiente

Al modelo de regresión no lineal mixto le adicionamos el efecto aleatorio a la pendiente, el cual definimos $\mathbf{b}_t = b_{t1}$. Una vez generados los datos se realiza el ajuste asumiendo distribución beta

ó normal para la variable respuesta y . Así, el modelo de regresión no lineal mixto beta con efecto aleatorio de pendiente se define como:

$$y_{tj}|b_{t1} \sim \text{beta}(\mu_{tj}, \phi), \quad (4.7)$$

y el modelo de regresión no lineal mixto normal se define como:

$$y_{tj}|b_{t1} \sim \mathbf{N}(\mu_{tj}, \sigma_\varepsilon^2), \quad (4.8)$$

donde la variable y_{tj} para los dos modelos tiene media condicional μ_{tj} :

$$\mu_{tj} = \frac{\exp(\beta_0 + (\beta_1 + b_{t1}) * \text{tiempo}_{tj})}{1 + \exp(\beta_0 + (\beta_1 + b_{t1}) * \text{tiempo}_{tj})} \quad (4.9)$$

y distribución para el efecto aleatorio b_{t1} :

$$b_{t1} \sim \mathbf{N}(0, \sigma_{b1}^2). \quad (4.10)$$

Igualmente a los modelos de la sección anterior, los valores iniciales de los parámetros β_0 y β_1 fueron calculados realizando una regresión lineal generalizada (con enlace logit) aplicando un GLM con el procedimiento GLIMMIX de SAS (v.9.1.3). Los parámetros de dispersión (ϕ para la beta y σ_ε para la normal) y todas las varianzas fueron parametrizadas en la escala *log* para evitar problemas de acotamiento como:

- $\phi = \exp(-\log\phi_i)$, donde $\log\phi_i = \ln(\text{escala})$. El valor inicial para el parámetro de escala es obtenido de la regresión lineal aplicando un GLM, tal como lo sugiere Smithson y Verkuilen (2006).
- $\sigma_\varepsilon = \exp(\log\sigma_\varepsilon)$

- $\sigma_{b_1} = \exp(\log sb_1)$

cuyos valores iniciales para $\log se$ y $\log sb_1$ fueron asignados teniendo en cuenta las características de los datos generados, de tal manera que la convergencia del modelo fuese lo mejor posible.

Resultados

A continuación se muestran los resultados de las distribuciones marginales y condicionales (cuando el efecto aleatorio $b_1 = 0$), asumiendo distribución beta o distribución normal únicamente para cuando la variable tiempo toma los valores de 0, 15, 30 y 50 (para los demás valores se encontraron resultados similares). El parámetro σ_{b_1} varía entre 0.004 y 0.02.

ϕ	σ_{b_1}	Tiempo	Media condicional $b_1 = 0$	Q_1	Mediana	Media marginal	Q_3	P_{95}	Desviación Estándar	Asimetría
10	0.004	0	0.2285414	0.2136114	0.2283783	0.2285414	0.2427662	0.2652006	0.0212951	0.1302581
		15	0.4944838	0.4717554	0.4936741	0.4932603	0.5153784	0.5507496	0.0353243	-0.0315848
		30	0.7635565	0.7349135	0.7622100	0.7593997	0.7887268	0.8302264	0.0465960	-0.5057904
		50	0.9400589	0.9242454	0.9462744	0.9355777	0.9530624	0.9698560	0.0273127	-1.5379001
	0.009	0	0.2270800	0.2136124	0.2266974	0.2270800	0.2402699	0.2610107	0.0202853	0.0904034
		15	0.4942615	0.4720166	0.4930041	0.4930475	0.5146668	0.5511611	0.0353371	-0.0591482
		30	0.7646550	0.7357498	0.7633219	0.7605004	0.7895226	0.8326879	0.0469421	-0.6023098
		50	0.9409280	0.9249411	0.9410398	0.9364433	0.9541233	0.9705823	0.0276467	-1.8916127
	0.01	0	0.2277823	0.2134356	0.2274930	0.2277823	0.2416976	0.2630291	0.0206533	0.1039190
		15	0.4944678	0.4715620	0.4932700	0.4932598	0.5155662	0.5506067	0.0348122	-0.0366121
		30	0.7642814	0.7362179	0.7622912	0.7602012	0.7890063	0.8315211	0.0463384	-0.6269812
		50	0.9406069	0.9247711	0.9403528	0.9361768	0.9535982	0.9706314	0.0276176	-2.0101821
	0.02	0	0.2288081	0.2131499	0.2283331	0.2288081	0.2433187	0.2665926	0.0216519	0.0501860
		15	0.4959151	0.4735302	0.4947513	0.4947069	0.5166629	0.5511407	0.0342035	-0.0227524
		30	0.7654317	0.7360330	0.7645139	0.7614190	0.7901354	0.8309771	0.0452292	-0.5272148
		50	0.9409815	0.9251868	0.9414534	0.9367023	0.9538654	0.9705600	0.0265039	-1.5184922

Tabla 4.12: Parámetros de las distribuciones marginales y condicionales del modelo de regresión beta no lineal mixto con efecto aleatorio en β_1 para diferentes valores de σ_{b_1} y $\phi = 10$.

ϕ	σ_{b_1}	Tiempo	Media condicional $b_1 = 0$	Q_1	Mediana	Media marginal	Q_3	P_{95}	Desviación Estándar	Asimetría
15	0.004	0	0.2278699	0.2155950	0.2270898	0.2278699	0.2391270	0.2592865	0.0179316	0.2138650
		15	0.4947238	0.4726913	0.4938619	0.4935234	0.5945041	0.5500642	0.0348232	-0.0713160
		30	0.7644567	0.7348625	0.7632178	0.7601599	0.7905692	0.8319537	0.0475634	-0.6153980
		50	0.9407331	0.9242011	0.9408785	0.9359829	0.9542580	0.9705090	0.0281171	-1.8278170
	0.009	0	0.2287807	0.2161814	0.2278824	0.2287800	0.2404426	0.2597171	0.0180638	0.2760615
		15	0.4951691	0.4730952	0.4941792	0.4939594	0.5155776	0.5497883	0.0345466	-0.0466340
		30	0.7641808	0.7337334	0.7634114	0.7598906	0.7906391	0.8310816	0.0474546	-0.6128014
		50	0.9403896	0.9233926	0.9408823	0.9356249	0.9538583	0.9701822	0.0282386	-1.7739678
	0.01	0	0.2295659	0.2181756	0.2286766	0.2295650	0.2407453	0.2597295	0.0173666	0.1828735
		15	0.4957670	0.4731477	0.4947741	0.4945538	0.5164854	0.5520669	0.0348778	-0.0891038
		30	0.7642033	0.7341581	0.7632663	0.7598356	0.7901104	0.8328149	0.0480429	-0.6334416
		50	0.9402497	0.9240420	0.9404726	0.9353634	0.9535596	0.9706289	0.0286556	-1.8301945
	0.02	0	0.2286374	0.2163033	0.2279970	0.2286374	0.2407973	0.2599829	0.0183117	0.1064150
		15	0.4949510	0.4732915	0.4942200	0.4937513	0.5155966	0.5504093	0.0345806	-0.1009940
		30	0.7640605	0.7343082	0.7638439	0.7598192	0.7890979	0.8313383	0.0469637	-0.6023990
		50	0.9404103	0.9243635	0.9407114	0.9357262	0.9535545	0.9700606	0.0277071	-1.7128645

Tabla 4.13: Parámetros de las distribuciones marginales y condicionales del modelo de regresión beta no lineal mixto con efecto aleatorio en β_1 para diferentes valores de σ_{b_1} y $\phi = 15$.

ϕ	σ_{b_1}	Tiempo	Media condicional $b_1 = 0$	Q_1	Mediana	Media marginal	Q_3	P_{95}	Desviación Estándar	Asimetría
10	0.004	0	0.2271313	0.2113999	0.2267756	0.2271313	0.2424952	0.266563	0.0230131	0.0787584
		15	0.4944534	0.4721755	0.4940394	0.4927316	0.5151089	0.5482737	0.0353797	-0.1906325
		30	0.7649143	0.7364106	0.7627066	0.7601377	0.7891235	0.8317659	0.0471429	-0.6057216
		50	0.9408060	0.9252861	0.9406888	0.9361287	0.9534997	0.9712778	0.0279321	-1.6969673
	0.009	0	0.2256776	0.2106792	0.2256055	0.2256776	0.2398353	0.2627086	0.0219571	0.0536016
		15	0.4943387	0.4726374	0.4934556	0.4926165	0.5144540	0.5488945	0.0356523	-0.1884553
		30	0.7661623	0.7367403	0.7648165	0.7613071	0.7898264	0.8330970	0.0478851	-0.6969134
		50	0.9417444	0.9255954	0.9419746	0.9369697	0.9544926	0.9717105	0.0285088	-2.0576634
	0.01	0	0.2263295	0.2104990	0.2262395	0.2263295	0.2414111	0.2642970	0.0222071	0.0905554
		15	0.4945518	0.4715445	0.4930722	0.4928016	0.5153644	0.5502487	0.0355075	-0.1281317
		30	0.7658451	0.7367459	0.7627581	0.7609322	0.7903006	0.8339222	0.0477116	-0.6771369
		50	0.9414757	0.9255102	0.9407571	0.9366350	0.9543855	0.9718369	0.0286166	-2.1748595
	0.02	0	0.2273257	0.2119146	0.2273527	0.2273257	0.2432321	0.2651347	0.0229191	-0.0027436
		15	0.4961262	0.4733857	0.4954471	0.4944158	0.5163059	0.5500084	0.0345775	-0.1276926
		30	0.7671417	0.7374429	0.7651172	0.7624673	0.7904578	0.8335181	0.0460093	-0.5440913
		50	0.9419008	0.9263759	0.9414228	0.9374154	0.9545450	0.9717623	0.0269330	-1.6393933

Tabla 4.14: Parámetros de las distribuciones marginales y condicionales del modelo de regresión normal no lineal mixto con efecto aleatorio en β_1 para diferentes valores de σ_{b_1} y $\phi = 10$.

ϕ	σ_{b_1}	Tiempo	Media condicional $b_1 = 0$	Q_1	Mediana	Media marginal	Q_3	P_{95}	Desviación Estándar	Asimetría
15	0.004	0	0.2266282	0.2131175	0.2261873	0.2266282	0.2394155	0.2586021	0.0189529	0.1682449
		15	0.4947350	0.4730616	0.4943013	0.4931540	0.5144214	0.5479435	0.0345711	-0.2126583
		30	0.7657114	0.7369925	0.7641217	0.7609944	0.7902460	0.8316271	0.0474217	-0.7267170
		50	0.9415031	0.9259374	0.9413653	0.9366803	0.9545018	0.9706967	0.0282083	-2.0081085
	0.009	0	0.2277960	0.2144532	0.2269475	0.2277960	0.2398375	0.2605792	0.0189662	0.2149030
		15	0.4951679	0.4733067	0.4942137	0.4935632	0.5150644	0.5481781	0.0341960	-0.1871170
		30	0.7651520	0.7356646	0.7642930	0.7604100	0.7901723	0.8308994	0.0474123	-0.7275262
		50	0.9409691	0.9244408	0.9413303	0.9360743	0.9541550	0.9701989	0.0285426	-1.9797473
	0.01	0	0.2286864	0.2164190	0.2275027	0.2286864	0.2408701	0.2593486	0.0180933	0.2373778
		15	0.4957453	0.4737202	0.4952660	0.4941247	0.5158980	0.5493706	0.0346159	-0.2021940
		30	0.7650246	0.7357741	0.7643024	0.7601685	0.7897044	0.8323614	0.0481042	-0.7380380
		50	0.9407433	0.9246972	0.9409719	0.9356856	0.9537186	0.9706446	0.0290062	-1.9898452
	0.02	0	0.2277964	0.2152173	0.2268760	0.2277964	0.2411049	0.2608271	0.0193644	0.0745253
		15	0.4950315	0.4738170	0.4945273	0.4934270	0.5141185	0.5486568	0.0343899	-0.1743052
		30	0.7650226	0.7364520	0.7643204	0.7602815	0.7884443	0.8312687	0.0469499	-0.6849281
		50	0.9409575	0.9253986	0.9408635	0.9361094	0.9536056	0.9704104	0.0279875	-1.9052496

Tabla 4.15: Parámetros de las distribuciones marginales y condicionales del modelo de regresión normal no lineal mixto con efecto aleatorio en β_1 para diferentes valores de σ_{b_1} y $\phi = 15$.

A continuación se muestran los resultados de los promedios de los parámetros estimados tanto para el modelo beta no lineal mixto como para el modelo normal no lineal mixto.

Efecto aleatorio en β_1 , $\phi = 10$ y variando σ_{b_1}

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.22188	0.119979	-1.23875	0.120083
β_1	0.08009	0.005477	0.08110	0.007071
<i>Logphi</i>	-2.33987	0.140890	-	-
<i>Logse</i>	-	-	-2.10260	0.070710
Logsb₁	-4.68517	0.470159	-4.59183	0.504489
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	10.36172	1.468790	-	-
σ_e	-	-	0.123730	0.008717
σ_{b_1}	0.008319	0.004358	0.008217	0.006324

Tabla 4.16: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 10$ y $\sigma_{b_1} = 0.004$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.23099	0.120291	-1.249130	0.120776
β_1	0.08060	0.005830	0.081508	0.006928
<i>Logphi</i>	-2.33683	0.140982	-	-
<i>Logse</i>	-	-	-2.100476	0.070731
Logsb₁	-4.69176	0.458852	-4.582873	0.524125
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	10.334780	1.467501	-	-
σ_e	-	-	0.123776	0.008717
σ_{b_1}	0.008376	0.004690	0.008223	0.007071

Tabla 4.17: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 10$ y $\sigma_{b_1} = 0.009$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.224053	0.120145	-1.239990	0.120249
β_1	0.080173	0.005744	0.081247	0.006928
$Logphi$	-2.338622	0.140946	-	-
$Logse$	-	-	-2.102919	0.070760
$Logsb_1$	-4.694562	0.457934	-4.589383	0.621663
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	10.416688	1.4787072	-	-
σ_e	-	-	0.123368	0.0087177
σ_{b_1}	0.008308	0.005830	0.008302	0.006403

Tabla 4.18: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 10$ y $\sigma_{b_1} = 0.01$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.219712	0.120374	-1.232789	0.120623
β_1	0.080352	0.005830	0.081288	0.006928
$Logphi$	-2.332815	0.140985	-	-
$Logse$	-	-	-2.098636	0.070753
$Logsb_1$	-4.674721	0.451404	-4.599340	0.514483
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	10.312854	1.461967	-	-
σ_e	-	-	0.123870	0.008717
σ_{b_1}	0.008135	0.004472	0.008128	0.006

Tabla 4.19: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 10$ y $\sigma_{b_1} = 0.02$.

Efecto aleatorio en β_1 , $\phi = 15$ y variando σ_{b_1}

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.224850	0.101059	-1.236657	0.099754
β_1	0.080220	0.005147	0.080956	0.005831
<i>Logphi</i>	-2.733704	0.140804	-	-
<i>Logse</i>	-	-	-2.284579	0.070717
Logsb₁	-4.721534	0.421655	-4.673795	0.473801
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	15.451601	2.193548	-	-
σ_e	-	-	0.102839	0.007281
σ_{b_1}	0.008573	0.003860	0.008412	0.004795

Tabla 4.20: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 15$ y $\sigma_{b_1} = 0.004$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.216976	0.100822	-1.226167	0.099461
β_1	0.079872	0.005128	0.080472	0.005856
<i>Logphi</i>	-2.736685	0.140757	-	-
<i>Logse</i>	-	-	-2.284060	0.070704
Logsb₁	-4.718626	0.413913	-4.662487	0.461804
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	15.530430	2.205750	-	-
σ_e	-	-	0.102676	0.007253
σ_{b_1}	0.008598	0.003987	0.008441	0.004785

Tabla 4.21: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 15$ y $\sigma_{b_1} = 0.009$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.213875	0.100528	-1.221383	0.099039
β_1	0.079794	0.005128	0.080222	0.005848
$Logphi$	-2.740954	0.140789	-	-
$Logse$	-	-	-2.285688	0.070726
$Logsb_1$	-4.718531	0.422322	-4.665304	0.462068
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	15.573223	2.210456	-	-
σ_e	-	-	0.102516	0.007239
σ_{b_1}	0.008644	0.003715	0.008530	0.004571

Tabla 4.22: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 15$ y $\sigma_{b_1} = 0.01$.

Estimación de los parámetros				
	DISTRIBUCIÓN BETA		DISTRIBUCIÓN NORMAL	
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
β_0	-1.217576	0.100891	-1.225875	0.099448
β_1	0.079895	0.005099	0.080405	0.005831
$Logphi$	-2.733985	0.140794	-	-
$Logse$	-	-	-2.283011	0.070753
$Logsb_1$	-4.726741	0.419062	-4.687035	0.472066
Estimaciones adicionales				
Parámetros	Estimación	Error Estándar	Estimación	Error Estándar
ϕ	15.4667	2.195611	-	-
σ_e	-	-	0.102726	0.007280
σ_{b_1}	0.008516	0.003987	0.008481	0.004582

Tabla 4.23: Promedio de los parámetros estimados de los modelos de regresión beta y normal no lineal mixto con efecto aleatorio en β_1 , $\phi = 15$ y $\sigma_{b_1} = 0.02$.

Conclusiones

En las Tablas 4.12 y 4.13 correspondientes al ajuste con la distribución beta se observa lo siguiente:

- La media marginal es igual a la media condicional ($b_1 = 0$) en el tiempo 0. Esto se debe a

que en el modelo marginal se adicionó el efecto aleatorio a la pendiente, lo que indica que el intercepto no se altera.

- La media condicional cuando el efecto aleatorio es cero es mayor a la mediana marginal excepto cuando la variable tiempo toma el valor de 50.
- La asimetría disminuye a medida que aumenta el valor de la variable tiempo.
- Las medias condicionales y marginales se encuentran entre el percentil 25 y el percentil 95, lo cual permite ver que la curva típica (media condicional $b_1 = 0$) y las curvas marginales para cada sujeto están acotadas por las curvas de estos percentiles.

En las Tablas 4.14 y 4.15 correspondientes al ajuste con la distribución normal se observa lo siguiente:

- La media marginal es igual a la media condicional ($b_1 = 0$) en el tiempo 0.
- La media condicional cuando el efecto aleatorio es cero es mayor a la mediana marginal excepto, cuando la variable tiempo toma el valor de 50.
- La asimetría disminuye a medida que aumenta el valor de la variable tiempo. Debido a que la distribución normal es simétrica, se considera que este ajuste no parece ser adecuado para los datos que se encuentran entre 0 y 1.
- Las medias condicionales y marginales se encuentran entre el percentil 25 y el percentil 95.

En las Tablas correspondientes a las estimaciones de los parámetros (Tabla 4.16 a tabla 4.23) se observa que los errores estándar de los parámetros estimados β_0 y β_1 son menores en el ajuste con la distribución beta que con la distribución normal.

Respecto a la convergencia, se puede decir que existe diferencia significativa entre las proporciones de convergencia en ambos modelos. A continuación, se muestra la prueba de McNemar sólo para los datos de la simulación correspondiente a la Tabla 4.16 y no hacer muy extensos los resultados. Las demás pruebas muestran resultados similares.

		BETA		Total
		Converge	No Converge	
NORMAL	Converge	648	83	731
	No Converge	184	85	269
	Total	832	168	1000

Tabla 4.24: Prueba de McNemar para la convergencia de los modelos ajustados con efecto aleatorio de pendiente

La hipótesis que se plantea es:

- H_0 : La proporción de convergencias es igual en ambos modelos.
- H_1 : La proporción de convergencias es diferente en ambos modelos.

El estadístico se calcula como:

$$\chi^2_{calculado} = \frac{(184 - 83)^2}{(184 + 83)} = 38.2059 \quad (4.11)$$

con una distribución $\chi^2_{1,0.95} = 3.8414$ y un valor $p = 6.36 \times 10^{-10}$. Por lo tanto se rechaza la hipótesis nula con un nivel de significancia del 0.05 y se concluye que sí hay diferencia significativa en la proporción de convergencias de los dos modelos.

Por último concluimos que en el modelo que asume distribución beta se obtuvo un mayor número de convergencias y un tiempo menor de ejecución para alcanzar dichas convergencias.

Aplicaciones: Estudio de Severidad de Enfermedades en Cultivos de Guineo en Puerto Rico

En este capítulo se aplicarán los resultados del modelo propuesto a un conjunto de datos reales. Los datos provienen de un estudio del desarrollo de la enfermedad Sigatoka negra y se obtuvieron en un experimento de campo en la Subestación Experimental Agrícola de Isabela, Colegio de Ciencias Agrícolas, Recinto Universitario de Mayagüez, Universidad de Puerto Rico. El objetivo principal es ajustar modelos de regresión no lineal mixtos bajo la distribución beta y la distribución normal, y ver cuál de los dos ajusta mejor a los datos. Se expondrá brevemente el origen de la Sigatoka negra y el índice de severidad, la descripción de los datos y las variables de interés, y finalmente se mostrarán los resultados obtenidos.

5.1. Enfermedad de la Sigatoka Negra

La enfermedad de Sigatoka negra en plantas de guineo, causada por el ascomiceto *Mycosphaerella fijiensis* Morelet, se caracteriza por el deterioro del área foliar de la planta, disminuyendo así su capacidad fotosintética, rendimiento del cultivo y calidad del fruto como se observa en la Figura 5.1

(Marengo, 2010). Cuando un patógeno se incrementa significativamente afectando una población de plantas en un periodo corto de tiempo se le conoce como *epidemia*. El estudio de una epidemia y de los factores que la influyen (relación causa-efecto) se le conoce como *epidemiología*. Para crear estrategias de manejo de la enfermedad se realiza el análisis detallado del progreso de la enfermedad en función del tiempo y de los factores que la influyen (Marengo, 2010). La descripción y valorización de la epidemia se realiza mediante evaluaciones de incidencia y severidad de la enfermedad del cultivo, lo cual permite expresar mediante gráficas y modelos epidemiológicos cómo se comporta la epidemia (Agrios, 1997).



Figura 5.1: Planta con síntomas característicos del ataque de Sigatoka negra (Marengo, 2010).

5.2. Descripción del Área Experimental y los Datos

Con la finalidad de examinar el desarrollo de la Sigatoka negra, se realizó un experimento de campo en la Sub-estación Experimental Agrícola de Isabela del Departamento de Ciencias Agrícolas, Recinto Universitario de Mayagüez, Universidad de Puerto Rico. Esta estación experimental se encuentra ubicada en el noroeste de Puerto Rico a una altura de 15 metros sobre el nivel del mar y con temperatura promedio de 24°C. Se realizó una siembra de guineo (*Musa acuminata*, AAA cv. “Grand Naine”) por su importancia económica en Puerto Rico y su sensibilidad a la Sigatoka Negra (Marengo, 2010). En cada parcela se sembraron 40 plantas de guineo, distribuidas en 10 filas de 10 plantas cada una.

Se seleccionaron 6 plantas para estudiar el desarrollo de la enfermedad y el resto de las plantas fue usado como borde de cada parcela. La distancia de siembra fue de 2.4 metros entre filas y 2.1 metros de planta a planta. Durante el desarrollo del experimento no se usó algún tipo de control químico para reducir el daño ocasionado por la enfermedad.

Para medir el desarrollo de la enfermedad, se evaluaron 24 plantas semanalmente durante 48 semanas consecutivas entre los años 2006 y 2007 hasta la cosecha para un total de 1,152 observaciones. De esta manera, tenemos 24 curvas de progreso, cada una con 48 observaciones. Es importante señalar que las observaciones tomadas sobre la misma planta están correlacionadas. Las tres variables respuesta de interés fueron la hoja más joven enferma, la hoja más joven manchada y el índice de severidad (IS). Sin embargo, sólo estudiaremos el IS en este trabajo. En la Tabla 5.1 se muestra una parte del conjunto de datos usado para el análisis estadístico usando modelos no lineales mixtos.

Bloque	Fecha	Planta	Número de hojas	Grado	Número de hojas por grado
1	11/13/2006	1	7	0	3
1	11/13/2006	1	7	1	1
1	11/13/2006	1	7	2	1
1	11/13/2006	1	7	3	0
1	11/13/2006	1	7	4	1
1	11/13/2006	1	7	5	1
1	11/13/2006	1	7	6	0
1	11/13/2006	2	6	0	3
1	11/13/2006	2	6	1	1
1	11/13/2006	2	6	2	2
⋮	⋮	⋮	⋮	⋮	⋮

Tabla 5.1: Visualización de la base de datos para el análisis estadístico de la Sigatoka negra.

Índice de Severidad

De acuerdo a la escala de Stover (Gauhl, 1994) (ver Figura 5.2), la severidad se clasifica en uno de los siete grados de daño de la hoja:

- Grado 0: Ningún sintoma.

- Grado 1: Menos del 1 % de la hoja está afectada.
- Grado 2: Del 1 % al 5 % del área foliar está afectada.
- Grado 3: Del 6 % al 15 % del área foliar está afectada.
- Grado 4: Del 16 % al 33 % del área foliar está afectada.
- Grado 5: Del 34 % al 50 % del área foliar está afectada.
- Grado 6: Del 51 % al 100 % del área foliar está afectada.

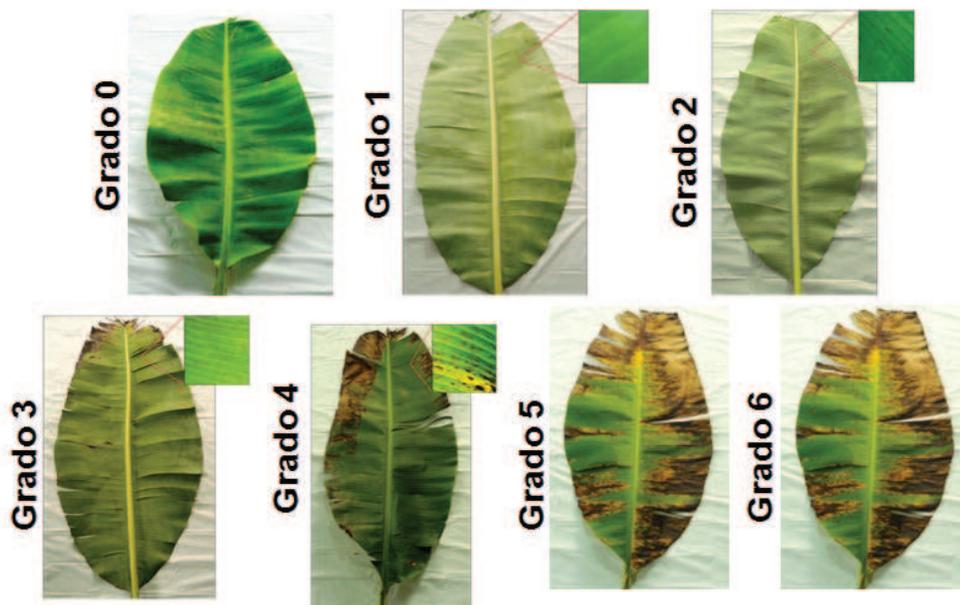


Figura 5.2: Grados de severidad de la enfermedad Sigatoka negra de acuerdo a la escala de Stover (Marengo, 2010).

El *Índice de severidad* (IS_{ij}) en el tiempo j es el promedio de los grados de severidad de las diferentes hojas en una planta o parcela representadas en una escala de 0 y 1.

Marengo (2010) realizó un estudio de la curva de progreso de la enfermedad ajustando modelos no lineales con curvas Gompertz y logística, concluyendo que el modelo logístico es el que mejor ajusta. Además realizaron correlaciones de Spearman para establecer la relación entre las condiciones climatológicas y la severidad de la enfermedad.

Como objetivo principal de este capítulo se generalizará el estudio de [Marengo \(2010\)](#), ajustando un modelo no lineal mixto para el índice de severidad y así establecer la curva de progreso de la severidad de la Sigatoka negra.

5.3. Métodos de Análisis

Los datos fueron analizados mediante modelos no lineales mixtos incorporando efectos aleatorios de planta. Los efectos aleatorios permiten modelar la correlación entre las observaciones de la misma planta. En la Figura 5.3 se puede ver la relación de los datos observados del índice de severidad y la variable independiente semana a través de cada planta, notándose una posible tendencia logística. Se ajustaron modelos para la media mediante la curva logística asumiendo distribución beta o normal para la variable respuesta y_{tj} . La variable y_{tj} es el índice de severidad para la planta t en el tiempo j . Se ajustaron modelos con uno y dos parámetros para la media. Se adicionó efecto aleatorio de intercepto y/o pendiente. La mejor combinación entre el modelo para la media y el efecto aleatorio será escogida mediante el criterio de AIC y gráfico de residuales.

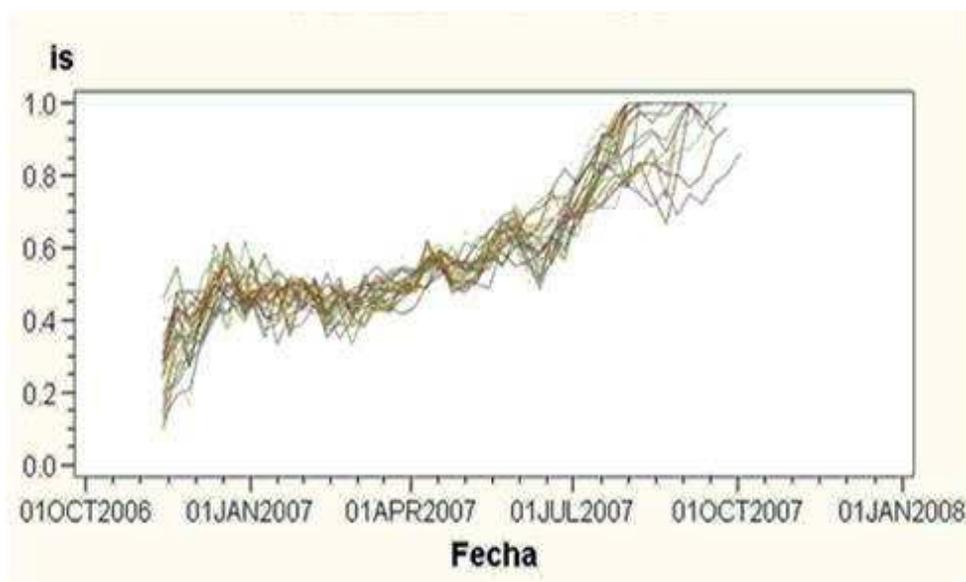


Figura 5.3: Datos observados del IS vs. fecha en la enfermedad de Sigatoka negra.

A continuación se formulan los modelos no lineales mixtos para el índice de severidad asumiendo distribución beta o normal.

Modelo No Lineal Mixto con Efecto Aleatorio de Intercepto

El modelo de regresión no lineal mixto beta con efecto aleatorio de intercepto se define como:

$$y_{tj}|b_{t0} \sim \text{beta}(\mu_{tj}, \phi), \quad (5.1)$$

y el modelo de regresión no lineal mixto normal como:

$$y_{tj}|b_{t0} \sim \mathbf{N}(\mu_{tj}, \sigma_{\varepsilon}^2). \quad (5.2)$$

Así, el índice de severidad para los dos modelos tiene media condicional μ_{tj} :

$$\mu_{tj} = E(IS_{tj}) = \frac{\exp(\beta_0 + b_{t0} + \beta_1 * \text{semana}_{tj})}{1 + \exp(\beta_0 + b_{t0} + \beta_1 * \text{semana}_{tj})} \quad (5.3)$$

y distribución para el efecto aleatorio b_{t0} :

$$b_{t0} \sim \mathbf{N}(0, \sigma_{b_0}^2). \quad (5.4)$$

Modelo No Lineal Mixto con Efecto Aleatorio de Pendiente

El modelo de regresión no lineal mixto beta con efecto aleatorio de pendiente se define como:

$$y_{tj}|b_{t1} \sim \text{beta}(\mu_{tj}, \phi), \quad (5.5)$$

y el modelo de regresión no lineal mixto normal como:

$$y_{ij}|b_{t1} \sim \mathbf{N}(\mu_{tj}, \sigma_{\varepsilon}^2). \quad (5.6)$$

Así, el índice de severidad para los dos modelos tiene media condicional μ_{tj} :

$$\mu_{tj} = E(IS_{tj}) = \frac{\exp(\beta_0 + (\beta_1 + b_{t1}) * semana_{tj})}{1 + \exp(\beta_0 + (\beta_1 + b_{t1}) * semana_{tj})} \quad (5.7)$$

y distribución para el efecto aleatorio b_{t1} :

$$b_{t1} \sim \mathbf{N}(0, \sigma_{b_1}^2). \quad (5.8)$$

Modelo No Lineal Mixto con Efecto Aleatorio de Intercepto y Pendiente

El modelo de regresión no lineal mixto beta con efecto aleatorio de intercepto y pendiente se define como:

$$y_{ij}|b_{t0}, b_{t1} \sim \text{beta}(\mu_{tj}, \phi), \quad (5.9)$$

y el modelo de regresión no lineal mixto normal como:

$$y_{ij}|b_{t0}, b_{t1} \sim \mathbf{N}(\mu_{tj}, \sigma_{\varepsilon}^2), \quad (5.10)$$

Así, el índice de severidad para los dos modelos tiene media condicional μ_{tj} :

$$\mu_{tj} = E(IS_{tj}) = \frac{\exp((\beta_0 + b_{t0}) + (\beta_1 + b_{t1}) * semana_{tj})}{1 + \exp((\beta_0 + b_{t0}) + (\beta_1 + b_{t1}) * semana_{tj})}, \quad (5.11)$$

donde b_{t0} y b_{t1} son los efectos aleatorios de planta en el intercepto y pendiente, respectivamente. La distribución de los efectos aleatorios se define como:

$$\mathbf{b}_t = [b_{t0}, b_{t1}] \sim \mathbf{N}([0, 0], \mathbf{G}) \quad (5.12)$$

donde la matriz \mathbf{G} puede ser escrita como:

$$\mathbf{G} = \begin{pmatrix} \sigma_{b_0}^2 & \rho\sigma_{b_0}\sigma_{b_1} \\ \rho\sigma_{b_0}\sigma_{b_1} & \sigma_{b_1}^2 \end{pmatrix} \quad (5.13)$$

La correlación ρ entre los efectos aleatorios de intercepto y pendiente fue parametrizada usando la transformación de Fisher como:

$$z = \frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right) \quad (5.14)$$

de donde se obtiene

$$\rho = \frac{\exp(2 * z) - 1}{\exp(2 * z) + 1} \quad (5.15)$$

y $\sigma_{b_0}, \sigma_{b_1}$ son las desviaciones estándar de los efectos aleatorios.

El análisis fue realizado en PROC NLMIXED de SAS (SAS Institute, 2003). Las estimaciones fueron obtenidas mediante máxima verosimilitud usando la técnica quasi-Newton para la optimización y la cuadratura Gaussiana para la integración. Los errores estándar de los parámetros del modelo se estimaron mediante la matriz Hessiana. El parámetro de dispersión de la beta y todas las varianzas fueron parametrizadas en la escala *log* para evitar problemas de acotamiento; sus errores estándar fueron calculados usando el método delta.

5.4. Resultados

Modelo No Lineal Mixto con Efecto Aleatorio de Intercepto

Los criterios de ajuste para los dos modelos ajustados se muestran en la Tabla 5.2.

Criterios de Ajuste		
Estadístico	Beta	Normal
-2 Log Likelihood	-1510	-1876
AIC	-1502	-1868
BIC	-1497	-1863

Tabla 5.2: Criterios de ajuste para el modelo no lineal mixto con efecto aleatorio de intercepto.

De acuerdo al ajuste de este modelo (ver Anexo A) se obtiene el \widehat{IS}_{ij} asumiendo distribución beta o normal como sigue:

Asumiendo Distribución Beta

$$\widehat{IS}_{ij} = \frac{\exp(-1.20 + b_{t0} + 0.08 * semana_{tj})}{1 + \exp(-1.20 + b_{t0} + 0.08 * semana_{tj})} \quad (5.16)$$

Así, la distribución condicional de las observaciones queda determinada como:

$$y_{ij}|b_{t0} \sim beta(\widehat{IS}_{ij}, 6.36), \quad (5.17)$$

y la distribución del efecto aleatorio como:

$$b_{t0} \sim \mathbf{N}(0, 0.21^2) \quad (5.18)$$

Asumiendo Distribución Normal

$$\widehat{IS}_{tj} = \frac{\exp(-0.83 + b_{t0} + 0.05 * semana_{tj})}{1 + \exp(-0.83 + b_{t0} + 0.05 * semana_{tj})} \quad (5.19)$$

Así, la distribución condicional de las observaciones queda determinada como:

$$y_{tj}|b_{t0} \sim \mathbf{N}(\widehat{IS}_{tj}, 0.45^2) \quad (5.20)$$

y la distribución del efecto aleatorio como:

$$b_{t0} \sim \mathbf{N}(0, 0.01^2) \quad (5.21)$$

En la Figura 5.4 se muestran las curvas ajustadas y las gráficas de residuales condicionados para ambos modelos. Se puede observar que los residuales de la distribución beta, a diferencia de la distribución normal, están más dispersos y muestran un patrón no tan definido. Esto indica un mejor comportamiento de residuos en el ajuste asumiendo distribución beta para la variable respuesta.

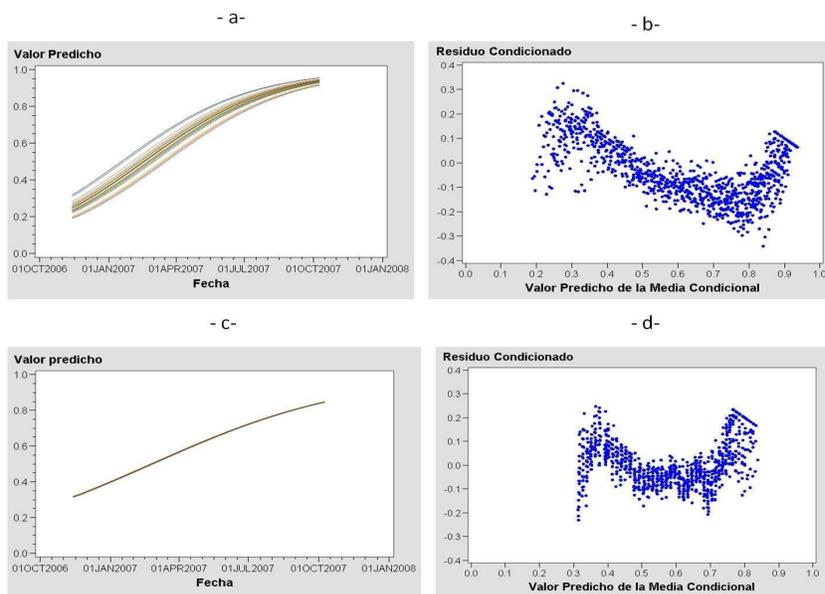


Figura 5.4: Curvas ajustadas y gráficas de residuales del modelo no lineal mixto con efecto aleatorio de intercepto asumiendo distribución beta (-a- y -b-) y distribución normal (-c- y -d-).

Modelo No Lineal Mixto con Efecto Aleatorio de Pendiente

Los criterios de ajuste para los dos modelos ajustados se muestran en la Tabla 5.3.

Criterios de Ajuste		
Estadístico	Beta	Normal
-2 Log Likelihood	-1604	-1876
AIC	-1596	-1868
BIC	-1592	-1863

Tabla 5.3: Criterios de ajuste para el modelo no lineal mixto con efecto aleatorio de pendiente.

De acuerdo al ajuste de este modelo (ver Anexo B) se obtiene el \widehat{IS}_{ij} asumiendo distribución beta o normal como sigue:

Asumiendo Distribución Beta

$$\widehat{IS}_{tj} = \frac{\exp(-1.17 + (0.08 + b_{t1}) * semana_{tj})}{1 + \exp(-1.17 + (0.08 + b_{t1}) * semana_{tj})} \quad (5.22)$$

Así, la distribución condicional de las observaciones queda determinada como:

$$y_{tj}|b_{t1} \sim \text{beta}(\widehat{IS}_{tj}, 6.94), \quad (5.23)$$

y la distribución del efecto aleatorio como:

$$b_{t0} \sim \mathbf{N}(0, 0.01^2) \quad (5.24)$$

Asumiendo Distribución Normal

$$\widehat{IS}_{tj} = \frac{\exp(-0.83 + (0.05 + b_{t1}) * semana_{tj})}{1 + \exp(-0.83 + (0.05 + b_{t1}) * semana_{tj})} \quad (5.25)$$

Así, la distribución condicional de las observaciones queda determinada como:

$$y_{tj}|b_{t1} \sim \mathbf{N}(\widehat{IS}_{tj}, 0.20^2) \quad (5.26)$$

y la distribución del efecto aleatorio como:

$$b_{t0} \sim \mathbf{N}(0, 0.0007^2) \quad (5.27)$$

En la Figura 5.5 se muestran las curvas ajustadas y las gráficas de residuales condicionados para

ambos modelos. Se puede observar un mejor comportamiento de residuos en el ajuste asumiendo distribución beta para la variable respuesta.

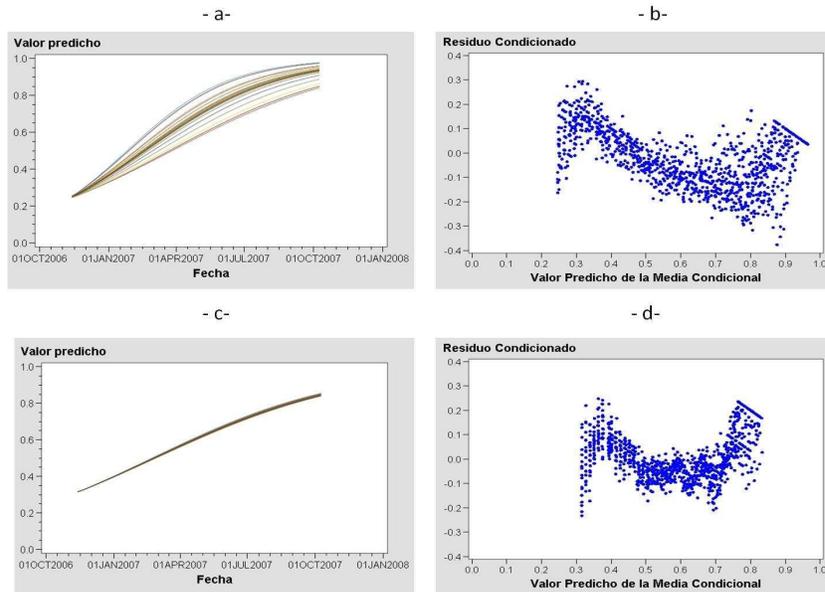


Figura 5.5: Curvas ajustadas y gráfica de residuales del modelo no lineal mixto con efecto aleatorio de pendiente asumiendo distribución beta (-a- y -b-) y distribución normal (-c- y -d-)

Modelo No Lineal Mixto con Efecto Aleatorio de Intercepto y Pendiente

Los criterios de ajuste para los dos modelos ajustados se muestran en la Tabla 5.4.

Criterios de Ajuste		
Estadístico	Beta	Normal
-2 Log Likelihood	-1604	-1968
AIC	-1592	-1956
BIC	-1585	-1949

Tabla 5.4: Criterios de ajuste para el modelo no lineal mixto con efecto aleatorio de intercepto y pendiente.

De acuerdo al ajuste de este modelo (ver Anexo C) se obtiene el \widehat{IS}_{ij} asumiendo distribución beta o normal como sigue:

Asumiendo Distribución Beta

$$\widehat{IS}_{ij} = \frac{\exp((-1.1797 + b_{i0}) + (0.08 + b_{i1}) * semana_{ij})}{1 + \exp((-1.1797 + b_{i0}) + (0.08 + b_{i1}) * semana_{ij})} \quad (5.28)$$

Así, la distribución condicional de las observaciones queda determinada como:

$$y_{ij}|b_{i0}, b_{i1} \sim \text{beta}(\widehat{IS}_{ij}, 6.94), \quad (5.29)$$

y la distribución de los efectos aleatorios como:

$$[b_{i0}, b_{i1}] \sim \mathbf{N} \left([0, 0], \begin{pmatrix} 0.0001^2 & -0.0000012 \\ -0.0000012 & 0.0127^2 \end{pmatrix} \right) \quad (5.30)$$

Asumiendo Distribución Normal

$$\widehat{IS}_{ij} = \frac{\exp((-0.82 + b_{i0}) + (0.05 + b_{i1}) * semana_{ij})}{1 + \exp((-0.82 + b_{i0}) + (0.05 + b_{i1}) * semana_{ij})} \quad (5.31)$$

Así, la distribución condicional de las observaciones queda determinada como:

$$y_{ij}|b_{i0}, b_{i1} \sim \mathbf{N}(\widehat{IS}_{ij}, 0.09^2) \quad (5.32)$$

y la distribución de los efectos aleatorios como:

$$[b_{i0}, b_{i1}] \sim \mathbf{N} \left([0, 0], \begin{pmatrix} 0.0962^2 & -0.00043 \\ -0.00043 & 0.004^2 \end{pmatrix} \right) \quad (5.33)$$

En la Figura 5.6 se muestran las curvas ajustadas y las gráficas de residuales condicionados para ambos modelos. Se puede observar un mejor comportamiento de residuos en el ajuste asumiendo distribución beta para la variable respuesta.

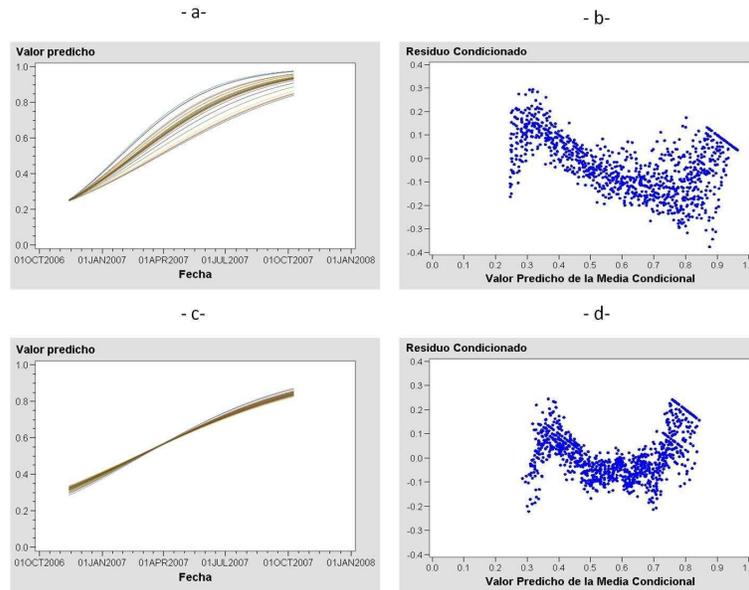


Figura 5.6: Curvas ajustadas y gráficas de residuales del modelo no lineal mixto con efecto aleatorio de intercepto y pendiente asumiendo distribución beta (-a- y -b-) y distribución normal (-c- y -d-)

5.5. Conclusiones

Mediante el criterio de AIC y gráfico de residuales se eligieron los dos modelos a los que se les adicionó simultáneamente el efecto aleatorio de intercepto y pendiente con una curva logística, como el que mejor ajusta a los datos observados. Por lo tanto, nos concentramos en las conclusiones para los modelos no lineales mixtos con efecto aleatorio de intercepto y pendiente. Se concluye lo siguiente:

- El valor estimado de la pendiente (tasa de progreso de la enfermedad) y su error estándar bajo la distribución beta son mayores que en la que la distribución normal.
- Los errores estándar para las estimaciones de las varianzas bajo distribución beta son menores que los de la distribución normal.

- El ajuste bajo la distribución beta presenta un mejor ajuste a los datos observados, lo que implica un mejor comportamiento de residuos (ver Figuras 5.3 y 5.6).
- El ajuste bajo la distribución normal no contempla varianzas heterogéneas mientras que en la beta sí (implícitamente).
- El ajuste bajo la distribución normal necesitó más puntos de cuadratura para poder alcanzar su convergencia.
- El ajuste con la distribución beta resultó ser mejor alternativa que el modelo normal.

Conclusiones Generales y Trabajos Futuros

6.1. Conclusiones Generales

La distribución beta resulta ser muy útil en el modelamiento de datos medidos como una proporción continua en el intervalo $(0,1)$. Este tipo de ajuste presenta ventajas importantes, como:

- No hay necesidad de hacer transformaciones a los datos para hacer cumplir algún supuesto.
- De acuerdo a la parametrización que se usó en este trabajo, el supuesto de homoscedasticidad se flexibiliza, ya que la varianza de la variable respuesta está en función de la media. Esto es muy importante porque permite situaciones en que las varianzas son heterogéneas, característica muy común en este tipo de datos.
- La asimetría es muy común en este tipo de datos, por lo que el ajuste asumiendo distribución beta es muy útil ya que se pueden modelar diferentes formas de la distribución.

En situaciones donde se incumplen los supuestos de independencia y varianzas constantes, los modelos no lineales mixtos resultan ser una herramienta adecuada para obtener inferencias confiables y válidas. Recordemos que este tipo de datos suelen estar correlacionados, ya que las observaciones proceden de un mismo individuo.

De los resultados obtenidos en las simulaciones podemos concluir lo siguiente que:

Un modelo no lineal mixto asumiendo distribución normal no es una buena aproximación para datos de proporciones en el intervalo continuo (0,1), algunas de las falencias que se encontraron fueron:

- la asimetría en los datos es bastante clara y por definición la distribución normal es simétrica.
- Este ajuste no contempló varianzas heterogéneas, situación muy común en este tipo de datos.

Las medias condicionales y las medias marginales en ambos modelos se encuentran entre el percentil 25 y el percentil 95, lo cual permite ver que la curva típica (media condicional para $b = 0$) y las curvas marginales para cada sujeto estarán acotadas por las curvas de estos percentiles.

Las estimaciones de los parámetros muestran que los errores estándar de los parámetros estimados β_0 y β_1 son menores en el ajuste del modelo asumiendo distribución beta que en el que se asume distribución normal.

Al intentar ajustar un modelo no lineal mixto con efecto aleatorio de intercepto y pendiente simultáneamente, no fue posible encontrar valores iniciales adecuados para los parámetros que garantizaran una convergencia satisfactoria de los modelos. Aunque se observó un mayor número de convergencias para el modelo beta no lineal mixto, no se considera una cantidad suficiente para tomar decisiones o conclusiones generales. El tiempo de computación para la estimación del modelo aumentó considerablemente respecto al tiempo que se tomó el ajuste de los modelos con intercepto o pendiente aleatoria solamente. Ya se había dicho con anterioridad que al aumentar el número de efectos aleatorios al modelo el ajuste se hacía más complejo. Además, al trabajar con datos aleatorios simulados hace un poco más difícil la estabilidad de las estimaciones.

De los resultados obtenidos en la aplicación podemos concluir que el ajuste bajo un modelo no lineal mixto asumiendo distribución beta para la variable respuesta *índice de severidad* resultó ser el mejor ajuste a los datos observados.

6.2. Trabajos Futuros

- En este trabajo se estudiaron las propiedades del modelo no lineal mixto asumiendo distribución beta con parámetro de precisión constante, por lo que para estudios posteriores se podría estudiar el ajuste de modelos con un parámetro de precisión variable.
- En los estudios de simulación se consideró únicamente la curva logística para la media de los modelos ajustados, sin embargo se recomienda usar otro tipo de curvas. Por ejemplo, la curva Gompertz.
- En este trabajo asumimos que la distribución de los efectos aleatorios es normal, pero se podría estudiar el ajuste de modelos asumiendo otro tipo de distribuciones para el efecto aleatorio.
- Se deberían escribir nuevos programas para el ajuste de modelos no lineales mixtos asumiendo distribución beta para la variable respuesta (por ejemplo, el paquete nlme4 de R está diseñado para el ajuste de modelos no lineales mixtos solamente con distribución normal).

Bibliografía

Agresti, A. (2002). *Categorical Data Analysis*. John Wiley and Sons, Hoboken, New Jersey.

Agrios, G. (1997). *Plant Pathology*. Academic Press, San Diego, CA.

Atkinson, A. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford University Press, New York.

Casella, G. y Berger, R. (2002). *Statistical Inference*. Duxbury Press, Belmont, CA.

Cox, C. (1996). Nonlinear Quasi-likelihood Models: Applications to Continuous Proportions. *Journal of Computational Statistics and Data Analysis*, 2:449–641.

Cribari-Neto, F. y Souza, T. (2011). Testing inference in Variable Dispersion Beta Regression. *Journal of Statistical Computation and Simulation*, iFirst:1–17.

Cribari-Neto, F. y Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(2):1–23.

Davidian, M. y Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, Boca Raton, Florida.

Draper, N. y Smith, H. (1981). *Applied Regression Analysis*. John Wiley and Sons, New York.

Espinheira, O. (2007). *Regressão Beta*. Tesis de Doctorado en Ciencias, Universidade de São Paulo, São Paulo.

- Ferrari, S. y Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, 31(7):799–815.
- Gauhl, F. (1994). *Epidemiology and Ecology of Black Sigatoka (Micosphaerella fijensis Morelet) on Plantain and Banana (Musa sp) in Costa Rica, Central America*. Ph. D Thesis, University of Göttingen, Germany.
- Godfrey, L. (1988). *Misspecification Test in Econometrics: The Lagrange Multiplier Principle and other Approaches*. Cabridge University Press, New York.
- Hermalin, B. y Wallace, N. (1972). The Determinants of Efficiency and Solvency in Savings and Loans. *The RAND Journal of Economics*, 25:361–381.
- Kieschnick, R. y McCullough, B. (2003). Regression Analysis of Variates Observed on (0,1): Percentages, Proportions and Fractions. *Journal of Statistical Modelling*, 3:193–213.
- Kvam, P. y Vidakovic, B. (2007). *Nonparametric Statistics with applications to Science and Engineering*. John Wiley and Sons, Hoboken, New Jersey.
- Lindstrom, M. y Bates, D. (1990). Nonlinear Mixed-Effects Models for Repeated Measures Data. *Biometrics*, 46:673–687.
- Macchiavelli, R. (2010). Aplicaciones de Modelos Mixtos en Ciencias Agrícolas. [disponible en: <<http://academic.uprm.edu/rmacchia/nolinealesmixtos.pdf>>]. Accesado el 14 octubre de 2011.
- Marengo, J. (2010). Epidemiología de la Sigatoka Negra (*Micosphaerella fijensis* Morelet) en una Plantilla de Guineo en Puerto Rico. Tesis de Maestría, Universidad de Puerto Rico, Recinto Universitario de Mayagüez.
- McCullagh, P. y Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London.
- Nelder, J. y Wedderburn, R. (1972). Generalized Linear Models. *Journal of Royal Statistical Society*, 135(3):370–384.

- Ospina, R. y Ferrari, S. (2007). Inflated Beta Distributions. [disponible en: <<http://www.de.ufpe.br/raydonal/research.htm>>]. Accesado el 10 de diciembre de 2011.
- Ospina, R. y Ferrari, S. (2012). A General Class of Zero - or - one Inflated Beta Regression Models. *Journal of Computational Statistical and Data Analysis*, 56(6):1609–1623.
- Owen, C. (2008). Parameter Estimation for the Beta Distribution. Tesis de Maestría, Brigham Young University.
- Paolino, P. (2001). Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables. *Journal of Political Analysis*, 9(4):325–346.
- Pinheiro, J. y Bates, D. (2000). *Mixed - Effect Models In S and S-plus*. Springer, New York.
- R-Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rousseeuw, P. y Leroy, A. (1987). *Robust Regression and Outliers Detection*. John Wiley and Sons, New York.
- Salinas, R., Pérez, N., y Avila, B. (2006). Modelos de Regresión para Variables Expresadas como una Proporción Continua. *Salud Pública de México*, 48(5):395–404.
- Santos de Oliveira, M. (2004). Um Modelo de Regressão Beta: Teoria y Aplicações. Tesis de Maestría en Estadística, Universidade de São Paulo, São Paulo.
- SAS Institute, . (2003). *SAS System for Windows v9.1.3*. Cary, NC, USA.
- Smithson, M. y Verkuilen, J. (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables. *Journal of Psychological Methods*, 11:54–71.
- Wedderburn, R. (1974). Quasi-Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method. *Biometrics*, 61:439–447.

Anexos

ANEXO A:

Salidas del Modelo No Lineal Mixto con Efecto Aleatorio de Intercepto para los Datos de Severidad

Asumiendo distribución beta

Especificaciones	
Data Set	WORK.TOTALES
Variable dependiente	is
Distribución para la variable dependiente	General (Beta)
Efectos aleatorios	b_0
Distribución para los efectos aleatorios	Normal
Sujeto	planta
Técnica de optimización	Dual Quasi-Newton
Método de integración	Cuadratura Gaussiana Adaptativa

Parámetros				
β_0	β_1	logphi	logsb0	NegLogLike
-1.22	0.08	0.7	-0.7	415.6869

Criterios de Ajuste	
-2 Log Likelihood	-1510
AIC	-1502
AICC	-1502
BIC	-1497

Parámetros Estimados								
Parámetro	Estimación	Error Estándar	GL	valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
β_0	-1.2085	0.00198	23	-609.91	<.0001	0.05	-1.2126	-1.2044
β_1	0.0820	0.00028	23	286.34	<.0001	0.05	0.0814	0.08261
$\log\phi$	-1.8507	0.00044	23	-4160.6	<.0001	0.05	-1.8516	-1.8498
\logsb_0	-1.5324	0.000991	23	-1546.0	<.0001	0.05	-1.5345	-1.5304

Estimaciones Adicionales								
Nivel	Estimación	Error Estándar	GL	Valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
ϕ	6.3644	0.002831	23	2248.11	<.0001	0.05	6.3585	6.3703
σ_{b_0}	0.216	0.000214	23	1008.85	<.0001	0.05	0.2156	0.2165

Asumiendo distribución normal

Especificaciones	
Data Set	WORK.TOTALES
Variable dependiente	is
Distribución para la variable dependiente	Normal
Efectos aleatorios	b_0
Distribución para los efectos aleatorios	Normal
Sujeto	planta
Técnica de optimización	Dual Quasi-Newton
Método de integración	Cuadratura Gaussiana adaptativa

Parámetros				
β_0	β_1	\logsb_0	\logse	NegLogLike
-1.22	0.08	-4	-0.7	-727.403

Criterios de Ajuste	
-2 Log Likelihood	-1876
AIC	-1868
AICC	-1868
BIC	-1863

Parámetros Estimados								
Parámetro	Estimación	Error Estándar	GL	valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
β_0	-0.8365	0.02737	23	-30.56	<.0001	0.05	-0.8931	-0.7798
β_1	0.0539	0.00118	23	45.43	<.0001	0.05	0.05148	0.0564
<i>Logsb0</i>	-4	64.8096	23	-0.06	0.9513	0.05	-138.07	-130.07
<i>Logse</i>	-0.7937	0.01119	23	-70.9	<.0001	0.05	-0.8169	-0.7706

Estimaciones Adicionales								
Nivel	Estimación	Error Estándar	GL	Valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
σ_{b0}	0.0183	0.04348	23	0.01	0.9939	0.05	-0.08961	0.09029
σ_e	0.4521	0.004577	23	44.67	<0.0001	0.05	0.195	0.2139

ANEXO B:

Salidas del Modelo No Lineal Mixto con Efecto Aleatorio de Pendiente para los Datos de Severidad

Asumiendo distribución beta

Especificaciones	
Data Set	WORK.TOTALES
Variable dependiente	is
Distribución para la variable dependiente	General (Beta)
Efectos aleatorios	b_1
Distribución para los efectos aleatorios	Normal
Sujeto	planta
Técnica de optimización	Dual Quasi-Newton
Método de integración	Cuadratura Gaussiana Adaptativa

Parámetros				
β_0	β_1	logphi	logsb1	NegLogLike
-1.22	0.08	0.7	-0.7	415.2368

Criterios de Ajuste	
-2 Log Likelihood	-1604
AIC	-1596
AICC	-1596
BIC	-1592

Parámetros Estimados								
Parámetro	Estimación	Error Estándar	GL	valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
β_0	-1.1796	0.009688	23	-121.76	<.0001	0.05	-1.1997	-1.1596
β_1	0.08057	0.003169	23	25.42	<.0001	0.05	0.07402	0.08713
\logphi	-1.9376	0.003451	23	-561.48	<.0001	0.05	-1.9447	-1.9305
$\logsb1$	-4.3612	0.0533	23	-81.82	<.0001	0.05	-4.4715	-4.251

Estimaciones Adicionales								
Nivel	Estimación	Error Estándar	GL	Valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
ϕ	6.9442	0.02396	23	289.78	<.0001	0.05	6.8925	6.9916
σ_{b1}	0.01276	0.00068	23	18.76	<.0001	0.05	0.01136	0.01417

Asumiendo distribución normal

Especificaciones	
Data Set	WORK.TOTALES
Variable dependiente	is
Distribución para la variable dependiente	Normal
Efectos aleatorios	b_1
Distribución para los efectos aleatorios	Normal
Sujeto	planta
Técnica de optimización	Dual Quasi-Newton
Método de integración	Cuadratura Gaussiana adaptativa

Parámetros				
β_0	β_1	$\logsb1$	\logse	NegLogLike
-1.22	0.08	-7	-1	-655.025

Criterios de Ajuste	
-2 Log Likelihood	-1876
AIC	-1868
AICC	-1868
BIC	-1863

Parámetros Estimados								
Parámetro	Estimación	Error Estándar	GL	valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
β_0	-0.8369	0.02741	23	-30.53	<.0001	0.05	-0.8936	-0.7802
β_1	0.0539	0.001204	23	44.82	<.0001	0.05	0.05148	0.05646
<i>Logsb1</i>	-7.2100	2.2636	23	-3.19	0.0041	0.05	-11.8927	-2.5273
<i>Logse</i>	-1.5882	0.0226	23	-70.28	<.0001	0.05	-1.6349	-1.5414

Estimaciones Adicionales								
Nivel	Estimación	Error Estándar	GL	Valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
σ_{b1}	0.000739	0.001673	23	0.44	0.6628	0.05	-0.00272	0.0042
σ_e	0.2043	0.004617	23	44.25	<0.0001	0.05	0.1947	0.2138

ANEXO C:

Salidas del Modelo No Lineal Mixto con Efecto Aleatorio de Intercepto y Pendiente para los Datos de Severidad

Asumiendo distribución beta

Especificaciones	
Data Set	WORK.TOTALES
Variable dependiente	is
Distribución para la variable dependiente	General (Beta)
Efectos aleatorios	b_0 b_1
Distribución para los efectos aleatorios	Normal
Sujeto	planta
Técnica de optimización	Dual Quasi-Newton
Método de integración	Cuadratura Gaussiana Adaptativa

Parámetros						
β_0	β_1	logphi	ztanhr	logsb0	logsb1	NegLogLike
-1.22	0.08	0.7	0.4	-8	-4	356.207618

Criterios de Ajuste	
-2 Log Likelihood	-1604
AIC	-1592
AICC	-1592
BIC	-1585

Parámetros Estimados								
Parámetro	Estimación	Error Estándar	GL	valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
β_0	-1.1797	0.06323	22	-18.66	<.0001	0.05	-1.3108	-1.0485
β_1	0.08058	0.003882	22	20.76	<.0001	0.05	0.07253	0.08863
$\log\phi$	-1.9377	0.002133	22	-908.43	<.0001	0.05	-1.9421	-1.9332
$z\tanh r$	-1.7055	0.01259	22	-135.48	<.0001	0.05	-1.7316	-1.6794
$\logsb0$	-9.1969	0.3197	22	-28.76	<.0001	0.05	-9.8599	-8.5338
$\logsb1$	-4.361	0.04354	22	-100.17	<.0001	0.05	-4.4513	-4.2707

Estimaciones Adicionales								
Nivel	Estimación	Error Estándar	GL	Valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
σ_{b0}	0.000101	0.000032	22	3.13	0.0049	0.05	0.000034	0.000169
σ_{b1}	0.01277	0.000556	22	22.97	<.0001	0.05	0.01161	0.01392
ϕ	6.9425	0.01481	22	468.83	<.0001	0.05	6.9117	6.9732
ρ	-0.9361	0.001558	22	-600.99	<.0001	0.05	-0.9393	-0.9329

Asumiendo distribución normal

Especificaciones	
Data Set	WORK.TOTALES
Variable dependiente	is
Distribución para la variable dependiente	Normal
Efectos aleatorios	$b_0 b_1 w$
Distribución para los efectos aleatorios	Normal
Sujeto	planta
Técnica de optimización	Dual Quasi-Newton
Método de integración	Cuadratura Gaussiana adaptativa

Parámetros						
β_0	β_1	$\log\phi$	$\logsb0$	$\logsb1$	\logse	NegLogLike
-1.22	0.08	0.4	-1	-4	-1.2	-196.44766

Criterios de Ajuste	
-2 Log Likelihood	-1968
AIC	-1956
AICC	-1956
BIC	-1949

Parámetros Estimados								
Parámetro	Estimación	Error Estándar	GL	valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
β_0	-0.8256	0.03232	22	-25.55	<.0001	0.05	-0.8926	-0.7586
β_1	0.05286	0.001457	22	36.29	<.0001	0.05	0.04984	0.05588
$z_{\tanh r}$	-2.1442	1.3117	22	-1.63	0.1164	0.05	-4.8645	0.5762
$Logsb0$	-2.3434	0.3955	22	-5.93	<.0001	0.05	-3.1636	-1.5232
$Logsb1$	-5.3709	0.3407	22	-15.77	<.0001	0.05	-6.0774	-4.6644
$Logse$	-2.4001	0.02278	22	-105.37	<.0001	0.05	-2.4473	-2.3528

Estimaciones Adicionales								
Nivel	Estimación	Error Estándar	GL	Valor t	Prob > t	Alpha	Límite Inferior	Límite Superior
σ_e	0.09071	0.002066	22	43.9	<.0001	0.05	0.08643	0.095
σ_{b0}	0.096	0.03797	22	2.53	0.0191	0.05	0.01726	0.1747
σ_{b1}	0.00465	0.001584	22	2.94	0.0077	0.05	0.001365	0.007935
ρ	-0.9729	0.07009	22	-13.88	<.0001	0.05	-1.1183	-0.8276