MODELOS LINEALES GENERALIZADOS MIXTOS CON DISTRIBUCIÓN BINOMIAL NEGATIVA

Por

Jairo Arturo Ayala Godoy

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

2011

Aprobada por:

Edgar Acuña, Ph.D. Miembro, Comité Graduado	Fecha
Edgardo Lorenzo, Ph.D. Miembro, Comité Graduado	Fecha
Raúl Macchiavelli, Ph.D. Presidente, Comité Graduado	Fecha
Wilma Santiago, M.Arch. Representante, Estudios Graduados	Fecha
Omar Colón, Ph.D. Director del Departamento	Fecha

Abstract of Thesis presented to in partial fulfillment of the
requirements for the degree of Master of Science

GENERALIZED LINEAR MIXED MODELS WITH A NEGATIVE BINOMIAL DISTRIBUTION

By

Jairo Arturo Ayala Godoy

2011

Adviser: Raúl E. Macchiavelli.
Department: Mathematical Sciences.

The generalized linear mixed model (MLGM) is a widely used model with random effects. It is a good alternative to traditional linear mixed models if the Normal distribution assumption is not satisfied.

In this work we study some properties of generalized linear mixed models when the conditional distribution of observations is Negative Binomial and the random effects distribution is normal. We compare these properties with those of generalized linear mixed models with conditional Poisson distribution.

The Negative Binomial distribution has been widely used to model counts, and it is the standard alternative for overdispersed Poisson counts. For repeated measurements and other correlated data, GLMMs using negative binomial distribution can be very useful to model counts, accounting for possible correlations and for overdispersion.

We study some properties of this model, such as the induced marginal distribution, its moments, and the relationship between the conditional distributions defining the model and the induced marginal distribution. Many of these properties are studied using simulations in R and SAS, since they are analytically intractable.

Finally, we apply these models to a real problem based on the findings in a study of counts of seeds collected in traps in the dry forest of Guanica (Puerto Rico) between 2006 and 2008 under different treatments.

Resumen de tesis, presentado en cumplimiento parcial de los
requerimientos para el grado de Maestría en Ciencias

MODELOS LINEALES GENERALIZADOS MIXTOS CON DISTRIBUCIÓN BINOMIAL NEGATIVA

Por

Jairo Arturo Ayala Godoy

2011

Consejero: Raúl E. Macchiavelli.
Departamento: Ciencias Matemáticas.

El modelo lineal generalizado mixto (MLGM) es un modelo muy utilizado con efectos aleatorios. Es una buena alternativa para los modelos lineales mixtos tradicionales cuando no se satisface el supuesto de distribución Normal.

En este trabajo se estudian algunas propiedades de los modelos lineales generalizados mixtos cuando la distribución condicional de las observaciones es Binomial Negativa y la distribución de efectos aleatorios es Normal. Comparamos estas propiedades con las de los modelos lineales generalizados mixtos con distribución condicional Poisson.

La distribución Binomial Negativa ha sido ampliamente utilizada para modelos de recuentos, y es la alternativa estándar para recuentos Poisson con sobredispersión, mediciones repetidas y datos correlacionados. Los MLGMs utilizando la distribución condicional Binomial Negativa pueden ser muy útiles para modelos de recuentos, posibles correlaciones y sobredispersión.

Se estudian algunas de las propiedades de este modelo, tales como la distribución marginal inducida, sus momentos, y la relación entre las distribuciones condicionales que definen el modelo y la distribución marginal inducida. Muchas de estas propiedades se estudian por medio de simulaciones en R y SAS, ya que son analíticamente intratables.

Finalmente, aplicamos estos modelos a un problema real basado en los resultados encontrados en un estudio sobre recuentos de semillas recolectadas en trampas en el bosque seco de Guánica (Puerto Rico) entre 2006 y 2008 bajo distintos tratamientos.

A mis padres y hermanas por su amor incondicional, apoyo y dedicación.

Agradecimientos

A Dios por darme sabiduría, fuerza, perseverancia y ser mi guía siempre.

Al Dr. Raúl Macchiavelli por sus grandes consejos, paciencia y dedicación durante la realización de la tesis.

A los miembros del comité por su tiempo y sus valiosos consejos.

A todos los profesores del departamento de Matemáticas, en especial a la profesora Nilsa Toro y al Dr. Luis F. Cáceres por ayudarme a crecer profesional y personalmente.

A todos mis amigos con los cuales compartí bonitas experiencias de vida, por su apoyo y amistad brindada.

Índice general

Índice de tablas

Índice de figuras

Capítulo 1

Introducción

Los modelos son una herramienta muy utilizada para el análisis de datos que presentan una relación causa-efecto. El punto de partida en un modelo es un conjunto de datos que explica el comportamiento de una variable que queremos analizar a partir de otra u otras.

Las observaciones pueden ser recuentos. Los recuentos surgen cuando se contabilizan observaciones en un intervalo de amplitud determinada; por tanto, pueden ser considerados como realizaciones de una variable que sólo toma valores enteros no negativos. Los modelos de regresión clásicos, como el modelo de regresión lineal, presentan deficiencias a la hora de ser utilizados para analizar este tipo de datos, ya que no cumplen con todos los supuestos necesarios; por ejemplo, tener errores con varianza constante. Nelder y Weddeburn (1972) postularon los modelos lineales generalizados (MLG), extendiendo la teoría de modelos lineales, e incorporando de esta manera la posibilidad de modelar variables continuas o categóricas con distribuciones no necesariamente homocedásticas ni normales.

Un modelo lineal generalizado relaciona la distribución aleatoria de la variable dependiente en el estudio con variables regresoras a través de una función $\boldsymbol{g}$ (monótona diferenciable) llamada función de enlace. Ejemplos de estas funciones son: logaritmo, logit, probit, recíproco, entre otras.

En un MLG se asume que hay n observaciones de la variable respuesta $\boldsymbol{y}$ (independientes) y están generadas por una función de distribución de la familia exponencial. Esto implica que la varianza depende de la media a través de una función de su varianza.

$$\text{var}(y) = \phi \text{var}(\mu) \tag{1.1}$$

El parámetro de dispersión ϕ es conocido o puede estimarse. La media μ de la distribución depende de las variables independientes $\boldsymbol{x}$ a través de la fórmula:

$$g(\mathrm{E}[y]) = g(\mu) = \eta = x\beta \tag{1.2}$$

donde $x\beta$ es el predictor lineal y g es la función de enlace.

Ahora bien, los modelos lineales generalizados mixtos (MLGM) surgen cuando, además de tener variables regresoras o variables explicatorias conocidas, tenemos en el predictor lineal términos aleatorios. En un MLGM el vector de variable respuesta $\boldsymbol{y}$, se supone que tiene elementos condicionalmente independientes y cada efecto aleatorio tiene una función de densidad en la familia exponencial. En este caso,

$$g(\mathrm{E}[y \mid v]) = x\beta + zv \tag{1.3}$$

donde β es el efecto fijo y v es efecto aleatorio. La varianza condicional está dada por:

$$\mathrm{var}(y \mid v) = \phi \mathrm{var}(\mu) \tag{1.4}$$

donde $\mathrm{var}(\mu)$ relaciona la media condicional y la varianza, el parámetro ϕ es 1 para el modelo Binomial y Poisson. Por último, v tiene una distribución conocida $f(v)$, tal como la función de distribución Normal $N(0, \sigma)$, que se utiliza en muchos casos.

Por otra parte, la regresión Poisson es un método usado para modelar datos de recuentos. Sin embargo, la distribución Poisson cumple la igualdad entre la media y la varianza, esta propiedad es llamada equidispersión, una propiedad que muy pocas veces se encuentra en datos reales. Los datos que tienen mayor varianza que la media se denominan datos con sobredispersión. Hilbe (2008) menciona que en los entornos donde es de interés describir la forma en que los recuentos varían en función de variables explicatorias, a menudo hay muchos factores que no son o no pueden ser medidos. La omisión de tales factores en el modelo pueden causar sobredispersión con respecto a la distribución Poisson. La regresión Binomial Negativa es un método estándar que se utiliza para modelar datos Poisson con sobredispersión.

Cuando la Binomial Negativa se utiliza para modelar datos de recuentos Poisson con sobredispersión, la distribución puede ser pensada como una extensión del modelo de Poisson o como la derivación de un modelo mixto Poisson-Gamma.

La Binomial Negativa tradicionalmente se deriva de un modelo mixto Poisson-Gamma. Sin embargo, también puede ser considerada como un miembro de la familia exponencial, pero sólo si su parámetro auxiliar se introduce en la distribución como una constante. Esto permite aplicar al modelo Binomial Negativo las diferentes pruebas de bondad de ajuste y análisis residuales que se han desarrollado para MLG.

La derivación del modelo a partir de la función de distribución de probabilidad (FDP) no introduce el parámetro auxiliar como una constante, sin embargo, con la versión del modelo mixto Poisson-Gamma sí lo hace, solo hay que utilizar como función de enlace el logaritmo. el cual produce un MLG Binomial Negativo, llamado NB-2. La estimación de los parámetros es idéntica al usado por el modelo de enlace canónico, este modelo es llamado NB-C, sin embargo, los errores estándar de las estimaciones de los modelos difieren.

Existen varias formas para hallar los errores estándar del estimador basados en la matriz de información. El algoritmo comúnmente utilizado en MLG es el de Fisher, procedimiento que se basa en la matriz de información esperada, de ahí la diferencia en los errores estándar entre las dos versiones de la Binomial Negativa. El algoritmo del MLG Binomial Negativo puede ser modificado para permitir encontrar los errores estándar en base a la información observada. Cuando hacemos esto, el modelo NB-2 produce errores estándar idénticos al modelo NB-C. Esta forma de la Binomial Negativa fue llamada: *log-negative binomial* por Hilbe (1993a) y fue la base de un macro muy importante existente de la Binomial Negativa en SAS (Hilbe, 1994a).

Independientemente de la forma en que se estime la Binomial Negativa, es utilizada para modelar la sobredispersión Poisson. Las ventajas del MLG se basan en su capacidad de ajustar modelos y encontrar residuales que están disponibles en la mayoría de los software estadísticos de MLG.

Nosotros analizamos con detalle los dos métodos de estimación de la Binomial Negativa. Realizamos la derivación completa de ambos métodos, y la discusión de cómo los algoritmos puede ser modificados para hacer frente a datos de recuentos que no pueden ser modelados utilizando simplemente el método estándar Poisson. Por otra parte, veremos que, así como los modelos Poisson pueden tener sobredispersión, los modelos Binomiales Negativos también la pueden tener.

Extensiones de la distribución Poisson y de la Binomial Negativa se hacen en función del tipo de problema de fondo que se está abordando. Esto incluye modelos que permitan modelar datos longi-

tudinales o agrupados, efectos fijos, aleatorios y mixtos. Los modelos también pueden aplicarse en situaciones en las que los datos se pueden dividir en dos o más subgrupos de distribución. De hecho, tanto los modelos Poisson como los modelos Binomiales Negativos han sido extendidos para modelar una gran cantidad de situaciones con recuentos. Trataremos de dar una visión general de cada uno de las principales extensiones mencionadas.

Por lo general, a las extensiones al modelo Poisson le suceden extensiones análogas a la Binomial Negativa. Por ejemplo, se han formulado modelos con parámetros aleatorios y modelos de recuentos de intercepto aleatorio para tratar ciertos tipos de datos correlacionados. Las primeras implementaciones se basan en la distribución Poisson. La mayoria de la literatura discute los modelos de recuentos Poisson con parámetros aleatorios. La Binomial Negativa sólo ha aparecido en los últimos años, principalmente como resultado de la obra de Greene (2006). El software LIMPED sólo está disponible para modelos Binomiales Negativos de coeficientes aleatorios. Entre los dos modelos generales de recuentos, la Binomial Negativa tiene mayor generalidad. Este hecho, se discutirá con más detalle más adelante. La distribución Poisson se puede considerar como una distribución Binomial Negativa con el parámetro de heterogeneidad igual a cero.

Es importante observar que la Binomial Negativa se puede obtener y presentar con diferentes parametrizaciones. Algunos autores emplean una función de varianza que refleja claramente el modelo mixto Poisson-Gamma. La varianza del modelo Poisson se define como μ y en la Gamma como μ^2/α, la varianza de la Binomial Negativa es $\mu + \mu^2/\alpha$. Por tanto la mezcla del modelo Poisson-Gamma es clara. Esta parametrización es la misma que derivan Greemwood y Yule (1920). Una relación inversa entre α y μ también se utilizó para definir la varianza de la Binomial Negativa en McCullagh y Nelder (1989), pero muy pocos autores continuaron con esta representación.

Nelder y Lee (1992) desarrollaron su sistema KK, utilizado y definido por la macro Binomial Negativa en el software Genstat. En este sistema está la relación directa entre α y μ^2 que resulta la función de varianza de la Binomial Negativa como: $\mu + \alpha\mu^2$. McCullagh y Nelder (1994) han continuado con la relación directa en escritos posteriores, basándose en su obra de (1989) y la mayoría de autores han continuado usando la relación definida originalmente. Recientemente Faraway (2006) la trabajó de la misma manera.

La parametrización directa de la función de varianza en la Binomial Negativa fue realizada por Bres-

low y Lawless (1984, 1987). En la década de los noventa, la relación directa fue la más utilizada en la implementación de software de la Binomial Negativa, Hilbe (1993b y 1994a) la desarrolló en Stata y XploRe, Greene (2006) en LIMDEP, y Johnston (1997) en SAS. La parametrización directa también se especifica en Hilbe (1994a), Long (1997), Cameron y Trivedi (1998) y en la mayoría de artículos y libros sobre el tema. Recientemente Long y Freese (2003 y 2006), Hardin y Hilbe (2001) y otros autores han empleado la relación directa existente en la función de varianza. Es raro ahora encontrar mayores aplicaciones actuales que utilizan la parametrización inversa.

La razón para preferir la relación directa es porque se deriva de la Binomial Negativa en el modelado de recuentos con sobredispersión Poisson. Considerando de esta manera que α está directamente relacionado con la cantidad de sobredispersión en los datos. Si los datos no tienen sobredispersión, entonces $\alpha = 0$. El aumento de α indica crecimiento en la sobredispersión. Los valores en la práctica suelen oscilar entre 0 y alrededor de 4.

Dos libros han sido publicados recientemente (Hoffmann, 2004, y Faraway, 2006), donde afirman que la Binomial Negativa no es un modelo lineal generalizado. Sin embargo, que sea un MLG depende de si es miembro de la familia exponencial de un solo parámetro en la distribución. Esto ocurre si suponemos que el parámetro de sobredispersión, α es conocido y auxiliar, resultado que se ha llamado LIMQL (máxima cuasi-verosimilitud con información limitada) por Greene (2003), entonces la Binomial Negativa será un MLG. Por otra parte, si α es un parámetro auxiliar pero se necesita estimar, el modelo puede ser estimado como FIMQL (máxima cuasi-verosimilitud con información completa), pero este no es un MLG.

En esta tesis, la Binomial Negativa se estima como MLG y como un modelo de cuasi-verosimilitud. El MLG, puede ser de amplio uso durante el proceso de modelado. Sin embargo, para obtener un valor de α, es decir, para obtener un α conocido, es necesario estimarlo. El método tradicional para estimar α es por medio del algoritmo de máxima verosimilitud. Vamos a estar utilizando los dos métodos, cuando los datos a modelar sean Binomiales Negativos.

En el capítulo 2 definimos la función de distribución de probabilidad Binomial Negativa (FDP) y los procesos para derivarla. Además, la derivación del modelo mixto Poisson-Gamma, parametrización que se utiliza en algoritmos de máxima verosimilitud. En este capítulo se hace evidente que la Binomial Negativa es un miembro de la familia exponencial usada en los modelos lineales generalizados.

El capítulo 3 está dedicado al modelo lineal generalizado mixto Binomial Negativo. Este modelo surge como una alternativa para modelar la Binomial Negativa con sobredispersión, donde estudiaremos su distribución marginal inducida por este modelo, investigaremos las propiedades mediante software estadísticos y métodos numéricos, luego compararemos la mejoría existente de estos modelos estudiados con respecto a los vistos en los capítulos anteriores. Además, se pueden aplicar para datos longitudinales y otras situaciones con recuentos correlacionados.

En el capítulo 4 se trabajarán tres simulaciones diferentes encontrando algunas propiedades a el MLGM Binomial Negativo.

El capítulo 5 está dedicado a las aplicaciones de este modelo a un caso real, donde se analizan recuentos de semillas recolectadas en trampas en el bosque seco de Guánica (Puerto Rico) entre 2006 y 2008 bajo distintos tratamientos.

Capítulo 2

Regresión Binomial Negativa

Durante todo el texto analizaremos la naturaleza y la utilidad de diversas formas de la regresión Binomial Negativa que son de interés para el modelado de datos de recuentos. Además, examinamos algunos modelos que están relacionados con la familia Binomial Negativa.
En Hilbe (2008), el modelo Poisson (el más simple para modelar recuentos) también puede ser mejorado para adaptar los datos que violen los supuestos. De hecho, muchos de los problemas de distribución tienen la misma naturaleza tanto para la distribución Poisson como para la distribución Binomial Negativa. Por lo tanto, encontraremos enfoques similares para el manejo de estos datos. Estos incluyen modelos Binomiales Negativos especificando su correlación, por ejemplo, si tiene un parámetro de heterogeneidad como el modelo NB-2, pero también se puede utilizar para modelos con sobredispersión. Todos los modelos permiten que la varianza sea igual o mayor a la media.
Con respecto a la Binomial Negativa, la asignación de variación adicional se refiere a:

1. Una adaptación en la varianza de la función NB-2.

2. Una modificación a la distribución de probabilidad NB-2, lo que resulta en una función de log-verosimilitud modificada. La función de varianza de NB-2, se expresa como $\mu + \alpha\mu^2$.

El modelo con la función de enlace canónica NB-C mantiene la verosimilitud y la función de varianza del modelo NB-2. Se puede argumentar, que el modelo NB-2 es una alteración de la forma canónica, que se deriva directamente de la función de probabilidad Binomial Negativa.

Podemos incluir el modelo Geométrico como una variedad de la Binomial Negativa, esto es cierto cuando el parámetro auxiliar α es igual a uno. Se podría argumentar que la Poisson también debe

incluirse como una variedad de Binomial Negativa desde un modelo NB-2 con α igual a cero, pero no la tendremos en cuenta ya que el parámetro auxiliar sólo puede aproximarse a cero, pero nunca podrá llegar a ser cero. Aunque en la práctica, un modelo Binomial Negativo con un valor de α cercano a cero es estadísticamente indistinguible de un modelo de Poisson.

Este capítulo se dedicará principalmente a la derivación del modelo Binomial Negativo y los dos métodos más importantes de su estimación.

2.1. Derivación de la Binomial Negativa

El modelo estándar de regresión Binomial Negativa, después de Cameron y Trivedi (1998), normalmente se conoce como NB-2, y surge de un modelo mixto Poisson-Gamma. Sin embargo, también puede ser considerado individualmente como un miembro de la familia exponencial, sólo si su parámetro auxiliar α se introduce en la distribución como una constante y por lo tanto es posible formular un MLG con esta distribución. Realizaremos la derivación de la Binomial Negativa en primera instancia como un modelo mixto Poisson-Gamma.

2.1.1. Derivación de la Binomial Negativa como modelo mixto Poisson-Gamma

Podemos pensar en la distribución de probabilidad Poisson condicionada a un efecto aleatorio Gamma, donde la función de distribución de probabilidad estaría dada por:

$$f(y \mid v) = \frac{e^{-\lambda v}(\lambda v)^y}{y!} \tag{2.1}$$

donde v tiene distribución Gamma con media 1 y varianza θ. Por tanto, la distribución marginal de $\boldsymbol{y}$ está dada por:

$$f(y) = \int_0^\infty \frac{e^{-\lambda v}(\lambda v)^y}{y!} \left[\frac{v^{\frac{1}{\theta}-1} e^{-\frac{v}{\theta}}}{\Gamma\left(\frac{1}{\theta}\right)\theta^{\frac{1}{\theta}}} \right] dv \tag{2.2}$$

La mezcla Gamma permite explicar la sobredispersión o recuentos correlacionados Poisson. Podemos

reescribir la función marginal así:

$$
\begin{aligned}
f(y) &= \frac{\lambda^y}{\Gamma\left(\frac{1}{\theta}\right) y! \frac{1}{\theta}^{\frac{1}{\theta}}} \int_0^\infty e^{-v(\lambda+\frac{1}{\theta})} v^{(y+\frac{1}{\theta})-1} dv \\
&= \frac{\lambda^y}{\Gamma\left(\frac{1}{\theta}\right) y! \frac{1}{\theta}^{\frac{1}{\theta}}} \left[\frac{\Gamma(y+1/\theta)}{(\lambda+1/\theta)^{y+\frac{1}{\theta}}}\right] \\
&= \frac{\Gamma(y+1/\theta)}{\Gamma(1/\theta)\Gamma(y+1)} \left(\frac{\lambda}{\lambda+1/\theta}\right)^y \left(\frac{1}{1+\lambda\theta}\right)^{1/\theta}
\end{aligned} \tag{2.3}
$$

Luego, cambiando $\mu = \lambda$ y $\alpha = \theta$ obtenemos la FDP de la Binomial Negativa.

$$
f(y \mid \mu, \alpha) = \frac{\Gamma(y+1/\alpha)}{\Gamma(y+1)\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha\mu}\right)^{1/\alpha} \left(\frac{\alpha\mu}{1+\alpha\mu}\right)^y \tag{2.4}
$$

2.1.2. Derivación de la Binomial Negativa como miembro de la familia exponencial

Existen dos formas de escribir la función de probabilidad de la Binomial Negativa. Ambas formas pueden ser consideradas como miembros de la familia exponencial, siempre que consideremos el parámetro auxiliar como conocido, por lo que pueden ser modeladas en el ámbito de los modelos lineales generalizados. Una de esas formas es la canónica, que se deriva directamente de la FDP, la otra es una conversión de la canónica, donde la función de enlace es el logaritmo. Esta última es conocida como NB-2 ó modelo estándar de la Binomial Negativa. La NB-2 es de gran importancia, ya que nos permite comparar las estimaciones puntuales con el modelo Poisson.

Podemos describir la FDP de la Binomial Negativa como la probabilidad de observar ***y*** fracasos ante ***r*** éxitos en una serie de ensayos Bernoulli, donde ***r*** es un entero positivo. Sin embargo, no hay ninguna razón de peso matemática para limitar este parámetro a entero. A pesar de que la Binomial Negativa puede ser parametrizada de manera diferente, siempre es posible convertir los términos para producir la forma final que usaremos a continuación.

Tenemos que la FDP de la Binomial Negativa como recuentos se expresa así:

$$
f(y \mid r, p) = \binom{y+r-1}{r-1} p^r (1-p)^y \tag{2.5}
$$

Al expresar la FDP de la Binomial Negativa como una familia exponencial, resulta:

$$f(y \mid r, p) = \exp\left\{ y \ln(1-p) + r(\ln(p)) + \ln\binom{y+r-1}{r-1} \right\} \tag{2.6}$$

donde podemos observar el parámetro natural, el enlace canónico y el parámetro de escala.

$$\theta = \ln(1-p) \Rightarrow p = 1 - e^{\theta} \tag{2.7}$$

$$b(\theta) = -r\ln(p) \;\Rightarrow -r\ln\left(1 - e^{\theta}\right) \tag{2.8}$$

$$\alpha(\phi) = 1 \tag{2.9}$$

Hallando la primera y segunda derivada, con respecto a θ encontramos la media y la varianza de la Binomial Negativa respectivamente.

Media de la Binomial Negativa

$$b'(\theta) = \frac{\partial b}{\partial p}\frac{\partial p}{\partial \theta} = -\frac{r}{p}\{-(1-p)\} = \frac{r(1-p)}{p}$$

por lo tanto,

$$\mathrm{E}[y] = \frac{r(1-p)}{p} \tag{2.10}$$

Varianza de la Binomial Negativa

$$b''(\theta) = \frac{\partial^2 b}{\partial p^2}\left(\frac{\partial p}{\partial \theta}\right)^2 + \frac{\partial b}{\partial p}\frac{\partial^2 p}{\partial \theta^2} = \frac{r}{p^2}(1-p)^2 - \frac{r}{p}(1-p) = \frac{r(1-p)}{p^2}$$

Entonces,

$$\mathrm{var}(y) = \frac{r(1-p)}{p^2} \tag{2.11}$$

Parametrizando p y r en términos de μ y α, donde $\alpha = 1/r$, se obtiene:

$$\begin{aligned} \mu &= \frac{(1-p)}{\alpha p} \\ \alpha\mu &= \frac{(1-p)}{p} \\ p &= \frac{1}{1+\alpha\mu} \end{aligned} \tag{2.12}$$

Teniendo en cuenta los valores μ, α y reemplazando la combinatoria por funciones Gamma, podemos re-parametrizar la FDP Binomial Negativa, así:

$$f(y\,|\,\mu,\alpha) = \frac{\Gamma(y+1/\alpha)}{\Gamma(y+1)\Gamma(1/\alpha)}\left(\frac{1}{1+\alpha\mu}\right)^{1/\alpha}\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y} \tag{2.13}$$

donde,

$$\mathrm{E}[y] = \mu \quad \text{y} \quad \mathrm{var}(y) = \mu + \alpha\mu^2 \tag{2.14}$$

Función de verosimilitud de la Binomial Negativa

Tomando la ecuación (2.13) podemos obtener la log-verosimilitud de la Binomial Negativa para una observación y_i:

$$\ell(\mu_i\,|\,y_i,\alpha) = y_i\ln(\alpha\mu_i) - (y_i+1/\alpha)\ln(1+\alpha\mu_i) + \ln\Gamma(y_i+1/\alpha) - \ln\Gamma(y_i+1) - \ln\Gamma(1/\alpha) \tag{2.15}$$

Deviance

La función de la *Deviance* en el MLG se deriva de la función de log-verosimilitud del modelo saturado y la función de log-verosimilitud. La función del modelo saturado se consigue sustituyendo el valor de y_i en cada valor de μ_i.

$$D = 2\sum_{i=1}^{n}\Big\{\ell(y_i\,|\,y_i) - \ell(\mu_i\,|\,y_i)\Big\} \tag{2.16}$$

sustituyendo la función de log-verosimilitud especifica en cada ecuación, obtenemos,

$$\begin{aligned} D = 2\sum_{i=1}^{n}\Big\{ &\{y_i\ln(\alpha y_i) - (y_i+1/\alpha)\ln(1+\alpha y_i) + \ln\Gamma(y_i+1/\alpha) - \ln\Gamma(y_i+1) - \ln\Gamma(1/\alpha)\} \\ &- \{y_i\ln(\alpha\mu_i) - (y_i+1/\alpha)\ln(1+\alpha\mu_i) + \ln\Gamma(y_i+1/\alpha) - \ln\Gamma(y_i+1) - \ln\Gamma(1/\alpha)\}\Big\} \end{aligned} \tag{2.17}$$

Simplificando la ecuación (2.17) tenemos que la Deviance es:

$$D = 2\sum_{i=1}^{n}\left\{y_i\ln(y_i/\mu_i) - (y_i+1/\alpha)\ln\left(\frac{1+\alpha/y_i}{1+\alpha/\mu_i}\right)\right\} \tag{2.18}$$

2.2. Modelo lineal generalizado Binomial negativo (NB-2)

La distribución Binomial Negativa tiene como función de densidad de probabilidad:

$$f(y_i \,|\, \mu_i, \alpha) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha\mu_i}\right)^{1/\alpha} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i}, \qquad y_i = \{0, 1, 2, \dots\} \tag{2.19}$$

el predictor lineal y la función de enlace son:

$$\ln(\mu_i) = x_i'\underline{\beta} \tag{2.20}$$

Esta distribución tiene como media:

$$\mathrm{E}[y_i] = \mu_i = \exp\left(x_i'\underline{\beta}\right) \tag{2.21}$$

y varianza:

$$\mathrm{var}(y_i) = \mu_i + \alpha\mu_i^2 = \mu_i\,(1 + \alpha\mu) > \mathrm{E}[y_i] \tag{2.22}$$

La varianza de la distribución Binomial Negativa es superior a la media. Es decir, hay sobredispersión, resultado de la heterogeneidad no observada. Se tienen $y_1, y_2, \dots, y_n$ observaciones independientes.

El parámetro α se denomina parámetro de dispersión. Cuando $\alpha \to 0$, $\mathrm{var}(y) \to \mu$ y la distribución Binomial Negativa converge a la distribución Poisson (Cameron y Trivedi, 1998).

Podemos dar ejemplos de predictores lineales, tales como:

1. Regresión lineal, donde $\ln(\mu_i) = \beta_0 + \beta_1 x_i$
2. Efecto de tratamiento, donde $\ln(\mu_i) = \mu + \tau_i$

Capítulo 3

Modelo lineal generalizado mixto

El modelo lineal generalizado mixto (MLGM) es muy utilizado para modelos de efectos aleatorios en el contexto de medidas repetidas y en otros casos donde las observaciones estén correlacionadas. Este modelo surge como una alternativa para modelar datos con sobredispersión, en donde suponemos que los efectos aleatorios son independientes y tienen una distribución Normal. Cuando tenemos dicho supuesto, el vector media y la matriz de varianza-covarianza de $\boldsymbol{y}$ se pueden derivar fácilmente. Durante este capítulo cuando se utilice la función Binomial Negativa, se usará la parametrización NB-2.

Empezamos este capítulo realizando ejemplos sencillos en donde podremos observar y comparar los resultados obtenidos por medio de simulaciones. Trabajaremos de forma aritmética hasta donde se nos permita, puesto que existen algunos procesos que son intratables analíticamente.
En primer lugar vamos a analizar un modelo lineal generalizado mixto Poisson.

3.1. Modelo lineal generalizado mixto Poisson

3.1.1. Modelo lineal generalizado mixto Poisson con un efecto aleatorio normal

Consideremos la distribución condicional Poisson de la siguiente manera:

$$y_{ij} \,|\, v_i \sim Po(\lambda_{ij}) \tag{3.1}$$

donde los efectos aleatorios v_i son independientes, con distribución:

$$v_i \sim N\left(0, \sigma_v^2\right) \tag{3.2}$$

La función de enlace canónica es logarítmica. Supongamos un modelo en el que:

$$\ln(\lambda_{ij}) = \beta_0 + v_i + \beta_1 t_j \tag{3.3}$$

Media de la distribución marginal

Utilizando la definición de media condicional tenemos:

$$\mathrm{E}[y_{ij}] = \mathrm{E}\left[\mathrm{E}(y_{ij} \mid v_i)\right] = \mathrm{E}[\lambda_{ij}] = \mathrm{E}\left[\exp(\beta_0 + v_i + \beta_1 t_j)\right] = \exp(\beta_0 + \beta_1 t_j) \cdot \mathrm{E}\left[\exp(v_i)\right] \tag{3.4}$$

por medio de la función generadora de momentos de la distribución Normal obtenemos:

$$\mathrm{E}\left[\exp(v_i)\right] = \exp\left(\frac{\sigma_v^2}{2}\right) \tag{3.5}$$

Entonces, retomando la ecuación (3.4), la media de la distribución marginal es igual a:

$$\mathrm{E}[y_{ij}] = \mu_{ij} = \exp\left(\beta_0 + \beta_1 t_j + \frac{\sigma_v^2}{2}\right) \tag{3.6}$$

Varianza de la distribución marginal

Utilizando la definición de varianza condicional, tenemos:

$$\begin{aligned}
\mathrm{var}(y_{ij}) &= \mathrm{E}[\mathrm{var}(y_{ij} \mid v_i)] + \mathrm{var}[\mathrm{E}(y_{ij} \mid v_i)] \\
&= \mathrm{E}[\lambda_{ij}] + \mathrm{var}(\lambda_{ij}) = \mu_{ij} + \mathrm{var}\left(\exp\left(\beta_0 + v_i + \beta_1 t_j\right)\right) \\
&= \mu_{ij} + \mathrm{E}\left[\left(\exp\left(\beta_0 + v_i + \beta_1 t_j\right)\right)^2\right] - \left(\mathrm{E}\left[\exp(\beta_0 + \beta_1 t_j + v_i)\right]\right)^2 \\
&= \mu_{ij} + \exp(2(\beta_0 + \beta_1 t_j)) \cdot \mathrm{E}\left[\exp(2v_i)\right] - \mu_{ij}^2
\end{aligned} \tag{3.7}$$

Ahora, con ayuda de la función generadora de momentos de la Normal, tenemos que:

$$\mathrm{E}\left[\exp(2v_i)\right] = \exp\left(2\sigma_v^2\right) \tag{3.8}$$

Luego, retomando la ecuación (3.7) obtenemos que la varianza la distribución marginal es:

$$\mathrm{var}(y_{ij}) = \mu_{ij} + \mu_{ij}^2\left(\exp\left(\sigma_v^2\right) - 1\right) \tag{3.9}$$

Note que

$$\exp\left(\sigma_v^2\right) - 1 > 0$$

por lo tanto

$$\text{var}(y_{ij}) > \mu_{ij}$$

Es decir, el modelo tiene sobredispersión.

Realizamos una simulación en el Software estadístico SAS (v.9.1.3) de la distribución marginal, donde hallamos la media y la varianza. Esta simulación se realizó con diez mil observaciones, tomando los parámetros iniciales de $\beta_0 = 2$ y $\sigma_v^2 = 0.5$, los resultados se muestran a continuación.

Localización		Variabilidad	
Media	9.488168	Desviación	8.23081
Mediana	7	Varianza	67.74616
Moda	4	Rango	180
		Rango intercuantil	8

Tabla 3.1: Medidas estadísticas básicas

Comparando los resultados anteriores de la simulación con los valores teóricos esperados, observamos:

	Simulación	Valor teórico esperado
Media	9.488168	9.487735
Varianza	67.74616	67.88376

Tabla 3.2: Comparaciones

La media y la varianza tanto de forma teórica como por medio de la simulación son muy parecidas. Por obvias razones estos valores nunca serán idénticos, ya que estamos trabajando con una simulación.

Este mismo proceso lo podemos generalizar para un vector de efectos aleatorios, como se trabajará a continuación.

3.1.2. Modelo lineal generalizado mixto Poisson con un vector de efectos aleatorios Normales

En esta sección, presentaremos un modelo lineal generalizado Poisson condicionado a un vector de efectos aleatorios con distribución normal. El cual lo podemos presentar así:

$$y_{ij} \mid \underline{v_i} \sim Po(\lambda_{ij}) \tag{3.10}$$

El vector de efectos aleatorios independientes es:

$$\underline{v_i} \sim N\left(\underline{0}, \Sigma\right) \tag{3.11}$$

La función de enlace canónico y el predictor lineal son:

$$\ln(\lambda_{ij}) = x_j'\underline{\beta} + z_j'\underline{v_i} \tag{3.12}$$

Media de la distribución marginal

Aplicando la definición de esperanza condicional tenemos:

$$\mathrm{E}[y_{ij}] = \mathrm{E}\left[\mathrm{E}(y_{ij} \mid \underline{v_i})\right] = \mathrm{E}\left[\exp(x_j'\underline{\beta} + z_j'\underline{v_i})\right] = \exp(x_j'\underline{\beta}) \cdot \mathrm{E}\left[\exp(z_j'\underline{v_i})\right] \tag{3.13}$$

Note que,

$$z_j'\underline{v_i} \sim N\left(\underline{0}, z_j'\,\Sigma\, z_j\right) \tag{3.14}$$

Luego por medio de la función generadora de momentos de la distribución Normal obtenemos que:

$$\mathrm{E}\left[\exp(z_j'\underline{v_i})\right] = \exp\left(\frac{z_j'\,\Sigma\, z_j}{2}\right) \tag{3.15}$$

Por lo tanto, retomando la ecuación (3.13) la media de la distribución marginal es:

$$\mathrm{E}(y_{ij}) = \exp\left(x_j'\underline{\beta} + \frac{z_j'\,\Sigma\, z_j}{2}\right) \tag{3.16}$$

Varianza de la distribución marginal

Por medio de la definición de varianza condicional tenemos:

$$\begin{aligned}\text{var}(y_{ij}) &= \text{E}\left[\text{var}\left(y_{ij} \mid \underline{v_i}\right)\right] + \text{var}\left[\text{E}\left(y_{ij} \mid \underline{v_i}\right)\right] \\ &= \text{E}\left[\exp\left(x'_j\underline{\beta} + z'_j\underline{v_i}\right)\right] + \text{var}\left[\exp\left(x'_j\underline{\beta} + z'_j\underline{v_i}\right)\right] \\ &= \mu_{ij} + \text{E}\left[\left(\exp\left(x'_j\underline{\beta} + z'_j\underline{v_i}\right)\right)^2\right] - \left(\text{E}\left[\exp\left(x'_j\underline{\beta} + z'_j\underline{v_i}\right)\right]\right)^2 \\ &= \mu_{ij} + \left(\exp(2x'_j\underline{\beta})\right) \cdot \text{E}\left[\exp\left(2z'_j\underline{v_i}\right)\right] - {\mu_{ij}}^2 \end{aligned} \tag{3.17}$$

Utilizando la función generadora de momentos de la distribución Normal obtenemos:

$$\text{E}\left[\exp\left(2z'_j\underline{v_i}\right)\right] = \exp\left(2z'_j\Sigma z_j\right) \tag{3.18}$$

Por lo tanto, de la ecuación (3.17) la varianza de la función marginal es:

$$\text{var}(y_{ij}) = \mu_{ij} + {\mu_{ij}}^2\left(\exp\left(z'_j\Sigma z_j\right) - 1\right) \tag{3.19}$$

Observamos que la media y la varianza del modelo con un vector de efectos aleatorios tienen la misma estructura que el modelo con un solo efecto aleatorio, simplemente se generaliza.

Analicemos ahora el modelo de nuestro estudio, el modelo lineal generalizado mixto Binomial Negativo.

3.2. Modelo lineal generalizado mixto Binomial Negativo

En primera instancia, para una mejor comprensión, hagamos lo mismo que hicimos en el MLGM Poisson, es decir, vamos a tomar una distribución Binomial Negativa con un solo efecto aleatorio y una media condicional constante.

3.2.1. Modelo Binomial Negativo con un efecto aleatorio Normal

Consideremos,

$$y_{ij} \mid v_i \sim bn(\alpha, \xi_{ij}) \tag{3.20}$$

Los factores aleatorios son independientes de la forma:

$$v_i \sim N\left(0, \sigma_v^2\right) \tag{3.21}$$

donde la función de enlace canónico y el predictor lineal son:

$$\ln(\xi_{ij}) = \beta_0 + v_i + \beta_1 t_j \tag{3.22}$$

Media de la distribución marginal

Aplicando la definición de media condicional,

$$\mathrm{E}[y_{ij}] = \mathrm{E}\left[\mathrm{E}\left(y_{ij} \mid v_i\right)\right] = \mathrm{E}[\xi_{ij}] = \mathrm{E}\left[\exp\left(\beta_0 + v_i + \beta_1 t_j\right)\right] = \exp\left(\beta_0 + \beta_1 t_j\right) \cdot \mathrm{E}\left[\exp\left(v_i\right)\right] \tag{3.23}$$

con ayuda de la función generadora de momentos de la distribución Normal tenemos:

$$\mathrm{E}\left[\exp\left(v_i\right)\right] = \exp\left(\frac{\sigma_v^2}{2}\right) \tag{3.24}$$

Retomando la ecuación (3.23) la media de la distribución marginal es:

$$\mathrm{E}[y_{ij}] = \mu_{ij} = \exp\left(\beta_0 + \beta_1 t_j + \frac{\sigma_v^2}{2}\right) \tag{3.25}$$

Observamos que las ecuaciones (3.6) y (3.25) son iguales, es decir, la medias marginales son iguales en el MLGM Binomial Negativo y en el MLGM Poisson.

Ahora podemos comparar la esperanza marginal con la esperanza condicional para el efecto aleatorio promedio. Tomaremos como notación a:

$$\xi_{ij_0} = \mathrm{E}\left[y_{ij} \mid v_i = 0\right] \tag{3.26}$$

De la ecuación (3.25) tenemos que:

$$\begin{aligned} \mathrm{E}[y_{ij}] &= \exp\left(\beta_0 + \beta_1 t_j\right) \cdot \exp\left(\frac{\sigma_v^2}{2}\right) \\ &= \mathrm{E}\left[y_{ij} \mid v_i = 0\right] \cdot \exp\left(\frac{\sigma_v^2}{2}\right) \\ &= \xi_{ij_0} \cdot \exp\left(\frac{\sigma_v^2}{2}\right) \end{aligned} \tag{3.27}$$

Sabemos que:

$$\exp\left(\frac{\sigma_v^2}{2}\right) > 1 \tag{3.28}$$

Por lo tanto,

$$\mathrm{E}[y_{ij}] > \xi_{ij_0} \tag{3.29}$$

Note que esperanza condicional para el efecto aleatorio promedio es igual a la mediana del modelo marginal, es decir: $P_{50} = \xi_{ij_0}$. Entonces de la ecuación (3.29) obtenemos que la media es mayor que la mediana en el modelo marginal, es decir, la distribución marginal es asimétrica a la derecha o tiene asimetría positiva. Esta conclusión también se consigue en el MLGM Poisson.

Varianza de la distribución marginal

Aplicando la definición de varianza condicional tenemos:

$$\begin{aligned}
\mathrm{var}(y_{ij}) &= \mathrm{E}\left[\mathrm{var}\left(y_{ij} \mid v_i\right)\right] + \mathrm{var}\left[\mathrm{E}\left(y_{ij} \mid v_i\right)\right] \\
&= \mathrm{E}\left[\xi_{ij} + \alpha {\xi_{ij}}^2\right] + \mathrm{var}\left(\xi_{ij}\right) \\
&= \mathrm{E}\left[\xi_{ij}\right] + \mathrm{E}\left[\alpha {\xi_{ij}}^2\right] + \mathrm{E}\left[{\xi_{ij}}^2\right] - \left(\mathrm{E}\left[\xi_{ij}\right]\right)^2 \\
&= \mu_{ij} + \alpha \cdot \mathrm{E}\left[\left(\exp\left(\beta_0 + \beta_1 t_j + v_i\right)\right)^2\right] + \mathrm{E}\left[\left(\exp\left(\beta_0 + \beta_1 t_j + v_i\right)\right)^2\right] - {\mu_{ij}}^2 \\
&= \mu_{ij} + \exp\left(2(\beta_0 + \beta_1 t_j)\right) \cdot \mathrm{E}\left[\exp\left(2v_i\right)\right]\left(\alpha + 1\right) - {\mu_{ij}}^2
\end{aligned} \tag{3.30}$$

Sabemos que:

$$\mathrm{E}\left[\exp\left(2v_i\right)\right] = \exp\left(2\sigma_v^2\right) \tag{3.31}$$

Por lo tanto, reemplazando en la ecuación (3.30) la varianza de la distribución marginal es:

$$\mathrm{var}(y_{ij}) = \mu_{ij} + \alpha {\mu_{ij}}^2 \left[\left(\exp\left(\sigma_v^2\right)\right)\left(1 + \frac{1}{\alpha}\right) - \frac{1}{\alpha}\right] \tag{3.32}$$

Comparemos la varianza marginal con la varianza condicional para el efecto aleatorio promedio. Recordemos que la varianza condicional para el efecto aleatorio promedio es:

$$\mathrm{var}\left(y_{ij} \mid v_i = 0\right) = \xi_{ij_0} + \alpha \xi_{ij_0}^2 \tag{3.33}$$

Sustituyendo en la ecuación (3.32) la ecuación (3.27), obtenemos:

$$\begin{aligned}\text{var}(y_{ij}) &= \xi_{ij_0}\exp\left(\frac{\sigma_v^2}{2}\right) + \alpha\left(\xi_{ij_0}\exp\left(\frac{\sigma_v^2}{2}\right)\right)^2\left[\left(\exp\left(\sigma_v^2\right)\right)\left(1+\frac{1}{\alpha}\right)-\frac{1}{\alpha}\right] \\ &= \xi_{ij_0}\exp\left(\frac{\sigma_v^2}{2}\right) + \alpha\xi_{ij_0}^2\left(\exp\left(\sigma_v^2\right)\right)\left[\left(\exp\left(\sigma_v^2\right)\right)\left(1+\frac{1}{\alpha}\right)-\frac{1}{\alpha}\right] \end{aligned} \tag{3.34}$$

Analicemos la ecuación anterior por partes. Primero podemos probar que:

$$\exp\left(\sigma_v^2\right)\left(1+\frac{1}{\alpha}\right)-\frac{1}{\alpha} > 1$$

Sabemos que $\alpha > 0$ y $\exp\left(\sigma_v^2\right) > 1$, por tanto:

$$\begin{aligned} \exp\left(\sigma_v^2\right) &> 1 \\ \frac{1}{\alpha}\cdot\left(\exp\left(\sigma_v^2\right)\right) &> \frac{1}{\alpha} \\ \frac{1}{\alpha}\cdot\left(\exp\left(\sigma_v^2\right)\right) + \exp\left(\sigma_v^2\right) &> 1+\frac{1}{\alpha} \\ \frac{1}{\alpha}\cdot\left(\exp\left(\sigma_v^2\right)\right) + \exp\left(\sigma_v^2\right) - \frac{1}{\alpha} &> 1 \\ \exp\left(\sigma_v^2\right)\left(1+\frac{1}{\alpha}\right)-\frac{1}{\alpha} &> 1 \end{aligned} \tag{3.35}$$

Entonces, verificamos el resultado.

También tenemos que:

$$\xi_{ij_0}\exp\left(\frac{\sigma_v^2}{2}\right) > \xi_{ij_0} \tag{3.36}$$

De las ecuaciones (3.35) y (3.36) podemos concluir que:

$$\alpha\xi_{ij_0}^2\left(\exp\left(\sigma_v^2\right)\right)\left[\left(\exp\left(\sigma_v^2\right)\right)\left(1+\frac{1}{\alpha}\right)-\frac{1}{\alpha}\right] > \alpha\xi_{ij_0}^2 \tag{3.37}$$

Finalmente, sumando las ecuaciones (3.36) y (3.37) podemos concluir que:

$$\text{var}\left(y_{ij}\right) > \text{var}\left(y_{ij}\,|\,v_i = 0\right) \tag{3.38}$$

Podemos concluir que la varianza marginal es mayor que la varianza condicional para un efecto aleatorio promedio. Este es un resultado muy conocido para variables aleatorias.

Construimos una simulación con el Software estadístico SAS (v.9.1.3), con diez mil observaciones y los parámetros iniciales como $\beta_0 = 1$, $\sigma_v = 0.5$ y $\alpha = 0.5$. Obtenemos:

Localización		Variabilidad	
Media	9.483094	Desviación	11.92346
Mediana	2	Varianza	142.16884
Moda	2	Rango	411
		Rango intercuantil	9

Tabla 3.3: Medidas estadísticas básicas

Comparando los resultados anteriores de la simulación con los valores teóricos esperados, observamos:

	Simulación	Valor teórico esperado
Media	9.483094	9.487735
Varianza	142.16684	142.09034

Tabla 3.4: Comparaciones

En comparación con los modelos sin efectos aleatorios, note que la presencia de efectos aleatorios induce un aumento en la media y la varianza marginal en ambas distribuciones. Por lo tanto es más apropiado el uso del modelo mixto Binomial Negativo para recuentos con mayor dispersión que el modelo mixto Poisson.

3.2.2. Modelo Binomial Negativo con dos efectos aleatorios normales

En muchas situaciones, se podría suponer que cada sujeto responde de manera diferente a través del tiempo, y por lo tanto, cada uno debe tener un intercepto y una pendiente diferente. Un modelo que describe esta situación para y_{ij}, medida a través del tiempo t_j está dado por:

$$y_{ij} \,|\, \underline{v_i} \sim bn\left(\alpha, \xi_{ij}\right) \tag{3.39}$$

El vector de efectos aleatorios independientes es una Normal bivariada.

$$\underline{v_i} = (s_i, p_i) \sim N\,(0, D) \tag{3.40}$$

donde,

$$D = \begin{pmatrix} \sigma_s^2 & \rho\sigma_s\sigma_p \\ \rho\sigma_s\sigma_p & \sigma_p^2 \end{pmatrix} \tag{3.41}$$

La función de enlace canónica y el predictor lineal son:

$$\ln(\xi_{ij}) = (\beta_0 + s_i) + (\beta_1 + p_i)\, t_j \tag{3.42}$$

Media de la distribución marginal

Aplicando la definición de media condicional tenemos:

$$\begin{aligned} \mathrm{E}\left(y_{ij}\right) = \mathrm{E}\Big[\mathrm{E}\left(y_{ij} \mid \underline{v_i}\right)\Big] &= \mathrm{E}\Big[\exp\left((\beta_0 + s_i) + (\beta_1 + p_i)\, t_j\right)\Big] \\ &= \exp\left(\beta_0 + \beta_1 t_j\right) \cdot \mathrm{E}\Big[\exp\left(\exp(s_i)\right)\Big] \cdot \mathrm{E}\Big[\exp\left(p_i t_j\right)\Big] \end{aligned} \tag{3.43}$$

Note que,

$$s_i \sim N\left(0, \sigma_s^2\right) \tag{3.44}$$

y

$$p_i t_j \sim N\left(0, t_j' D\, t_j\right) \tag{3.45}$$

Definamos a:

$$x_j' = [1 \;\; t_j] \tag{3.46}$$

Retomando la ecuación (3.43) encontramos que:

$$\mu_{ij} = \mathrm{E}\left(y_{ij}\right) = \exp\left(x_j' \underline{\beta} + \frac{1}{2} x_j' D\, x_j\right) \tag{3.47}$$

Realizamos una simulación con el Software estadístico SAS (v.9.1.3), donde suponemos que existen 10 sujetos, los parámetros iniciales los tomamos como: $\beta_0 = 1$, $\beta_1 = 0.5$, $\alpha = 0.5$, $\sigma_s^2 = 0.5$, $\sigma_p^2 = 0.2$ y $\rho = 0.5$.

En la Figura (3.1), podemos observar las rectas producidas entre $\ln(\mu)$ vs t, $\quad t = \{0, 2, 4, 6, \ldots 20\}$, donde las líneas delgadas azules respresentan el $\ln(\mu)$ de cada sujeto específico, la línea roja el promedio de estas líneas y la línea verde representa el logaritmo del promedio.

El modelo presentado en la ecuación (3.39) es un caso especial del modelo lineal generalizado con un vector de efectos aleatorios Normales, donde asumimos que el vector de efectos aleatorios independientes es una Normal multivariada. Este lo analizaremos con detalle en la siguiente sección.

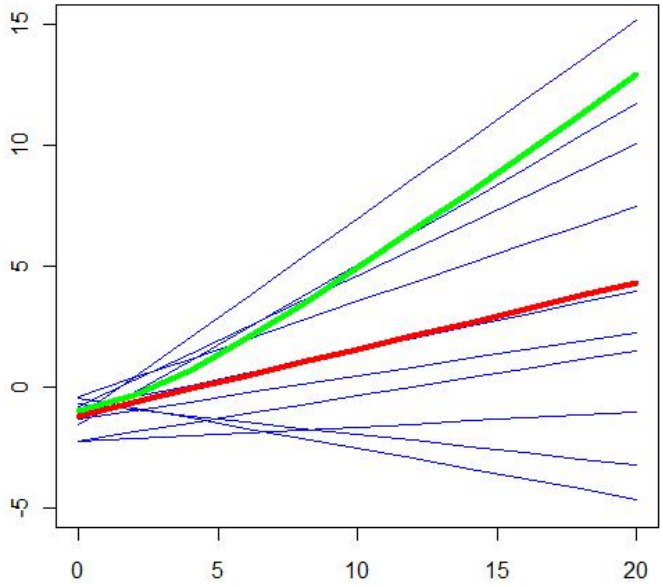

Figura 3.1: $\ln(\mu)$ vs t

3.2.3. Modelo Binomial Negativo con un vector general de efectos aleatorios Normales

Supongamos que tenemos una distribución condicional Binomial Negativa y un vector de efectos aleatorios normales, de la siguiente manera,

$$Y_{ij} \mid \underline{v_i} \sim bn\left(\alpha, \xi_{ij}\right) \tag{3.48}$$

El vector de efectos aleatorios independientes es Normal Multivariado.

$$\underline{v_i} \sim N\left(0, \Sigma\right) \tag{3.49}$$

La función de enlace canónico y el predictor líneal son:

$$\ln(\xi_{ij}) = x_j'\underline{\beta} + z_j'\underline{v_i} \tag{3.50}$$

donde, $\underline{\beta}$ es el vector de efectos fijos y $\underline{v_i}$ es el vector de efectos aleatorios.

Media de la distribución marginal

Encontremos de forma teórica su media.

$$\mathrm{E}\big(y_{ij}\big) = \mathrm{E}\Big[\mathrm{E}\big(y_{ij} \mid \underline{v_i}\big)\Big] = \mathrm{E}\left[\exp\left(x_j'\underline{\beta} + z_j'\underline{v_i}\right)\right] = \exp\left(x_j'\underline{\beta}\right) \cdot \mathrm{E}\Big[\exp\left(z_j'\underline{v_i}\right)\Big] \tag{3.51}$$

Note que

$$z_j'\underline{v_i} \sim N\left(0, z_j'\,\Sigma\, z_j\right) \tag{3.52}$$

por lo tanto,

$$\mathrm{E}\Big[\exp\left(z_j'\underline{v_i}\right)\Big] = \exp\left(\frac{z_j'\,\Sigma\, z_j}{2}\right) \tag{3.53}$$

Reemplazando en la ecuación (3.51) tenemos que la media marginal es igual a:

$$\mathrm{E}\big(y_{ij}\big) = \mu_{ij} = \exp\left(x_j'\underline{\beta} + \frac{z_j'\,\Sigma\, z_j}{2}\right) \tag{3.54}$$

Varianza de la distribución marginal

Aplicando la definición de varianza condicional tenemos:

$$\begin{aligned}
\mathrm{var}(y_{ij}) &= \mathrm{E}\left[\mathrm{var}\big(y_{ij} \mid \underline{v_i}\big)\right] + \mathrm{var}\left[\mathrm{E}\big(y_{ij} \mid \underline{v_i}\big)\right] \\
&= \mathrm{E}\left[\xi_{ij} + \alpha{\xi_{ij}}^2\right] + \mathrm{var}\big(\xi_{ij}\big) \\
&= \mathrm{E}\big[\xi_{ij}\big] + \mathrm{E}\left[\alpha{\xi_{ij}}^2\right] + \mathrm{E}\left[{\xi_{ij}}^2\right] - \big(\mathrm{E}\big[\xi_{ij}\big]\big)^2 \\
&= \mu_{ij} + \alpha \cdot \mathrm{E}\left[\left(\exp\left(x_j'\underline{\beta} + z_j'\underline{v_i}\right)\right)^2\right] + \mathrm{E}\left[\left(\exp\left(x_j'\underline{\beta} + z_j'\underline{v_i}\right)\right)^2\right] - {\mu_{ij}}^2 \\
&= \mu_{ij} + \exp\left(2x_j'\underline{\beta}\right) \cdot \mathrm{E}\left[\exp\left(2z_j'\underline{v_i}\right)\right]\left(\alpha + 1\right) - {\mu_{ij}}^2
\end{aligned} \tag{3.55}$$

Note que,

$$\mathrm{E}\left[\exp\left(2z_j'\underline{v_i}\right)\right] = \exp\left(2z_j'\,\Sigma\, z_j\right) \tag{3.56}$$

Por lo tanto, sustituyendo en la ecuación (3.55) la varianza de la distribución marginal es:

$$\mathrm{var}(y_{ij}) = \mu_{ij} + \alpha{\mu_{ij}}^2\left[\exp\left(z_j'\,\Sigma\, z_j\right)\left(1 + \frac{1}{\alpha}\right) - \frac{1}{\alpha}\right] \tag{3.57}$$

3.2.4. Inferencias del MLGM Binomial Negativo

Para trabajar con un MLGM es posible utilizar dos alternativas: el enfoque Bayesiano y máxima verosimilitud. En el enfoque Bayesiano, es necesario especificar la función de densidad a priori de β y Σ. Dado esto, podemos encontrar la función posterior.

Por otra parte, la densidad conjunta para MLGM se puede escribir a partir del producto de las distribuciones conocidas $y_{ij} \mid \underline{v_i}$ y $\underline{v_i}$. La función de densidad se puede obtener como:

$$L\left(y_{ij}\,,\,\underline{v_i}\right) = L\left(y_{ij} \mid \underline{v_i}\right) \cdot L\left(\underline{v_i}\right) \tag{3.58}$$

Como asumimos que $\underline{v_i}$ tiene una distribución Normal multivariada $N(0, \Sigma)$, entonces:

$$L\left(y_{ij}\,,\,\underline{v_i}\right) \propto L\left(y_{ij} \mid \underline{v_i}\right) \mid \Sigma \mid^{-1/2} \cdot \exp\left(-\frac{1}{2}\underline{v_i}'\,\Sigma^{-1}\,\underline{v_i}\right) \tag{3.59}$$

Con base en la ecuación (3.59), la verosimilitud marginal para $\boldsymbol{y}$ se obtiene de la integración de los efectos aleatorios $\underline{v_i}$.

En el MLGM se puede encontrar la verosimilitud marginal mediante la integración de los efectos aleatorios, donde se tiene i sujetos independientes cada uno con n_i observaciones.
La verosimilitud del i-ésimo sujeto específico viene dada por:

$$L_i(y_i) = f_i(y_{ij}) = \int \prod_{j=1}^{n_i} f(y_{ij} \mid \underline{v_i}) f(\underline{v_i}) d\underline{v_i} \tag{3.60}$$

Por lo tanto, la función de verosimilitud es:

$$L(y_{ij}) = \prod_{i=1}^{N} L_i = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f(y_{ij} \mid \underline{v_i}) f(\underline{v_i}) d\underline{v_i} \tag{3.61}$$

donde y_i es un vector n_i-dimensional que contiene la información del i-ésimo elemento. Observemos que la ecuación (3.61) tiene la forma de la ecuación (3.59). Esta integral es a través de la distribución q-dimensional de v_i, y es imposible evaluar analíticamente para la mayoria de los MLGM. Por tanto, es necesario utilizar aproximaciones numéricas con el fin de calcular la verosimilitud.

Las aproximaciones numéricas de la integral de la verosimilitud marginal en la ecuación (3.61) se pueden realizar de diversas formas: La primera se basa en la aproximación de la integral propia, tal como la cuadratura de Gauss-Hermite (Molenberghs y Verbeke, 2005). La segunda forma es por medio de la aproximación del integrando, como la aproximación de Laplace. Por último, una aproximación de los datos, como la cuasi-verosimilitud marginal y penalizada (Molenberghs y Verbeke, 2005).

Aproximación de la integral: la cuadratura de Gauss-Hermite

En muchas aplicaciones, es conveniente considerar la integral de la forma:

$$\int f(x)\phi(x)dx \tag{3.62}$$

donde $\phi(x)$ es una función no negativa en el intervalo de integración y se denomina función de peso. La integral puede ser interpretada como un promedio ponderado de $f(x)$.

La integral aproximada mediante la cuadratura de Gauss-Hermite es de la forma:

$$\int f(x)\exp\left(-x^2\right)dx \tag{3.63}$$

donde la función de peso $\phi(x) = \exp\left(-x^2\right)$. La cuadratura de Gauss-Hermite es comúnmente utilizada debido a la relación directa que tiene con la densidad de Gauss (Lange, 1988). En muchas aplicaciones estadísticas, la densidad Gausiana es un factor explícito del integrando. Utilizando una transformación líneal este factor tiene la forma $\exp\left(-x^2\right)$. Cuando la densidad Gausiana no es un factor del integrando, el integrando original debe multiplicarse y dividirse por $\exp\left(-x^2\right)$, con el fin de ponerlo de la forma de la cuadratura de Gauss-Hermite.

Queremos aplicar la cuadratura de Gauss-Hermite para aproximar integrales de la forma:

$$\int g(x)dx \tag{3.64}$$

Cuando $g(x) > 0$, ésta es una función suave y unimodal. La cuadratura de Gauss-Hermite puede ser reescrita en términos de densidad Gausiana como:

$$\int f(x)\phi(x\,|\,\mu,\sigma)dx \approx \sum_{q=1}^{Q}\omega_q f\left(z_q\right) \tag{3.65}$$

Donde $\phi(x\,|\,\mu,\sigma)$ es una densidad Gausiana arbitraria. Las abscisas son $z_q = \mu + \sqrt{2}\sigma x_q$, y los pesos son modificados de ω_q a ω_q/π, tenemos $q = 1, 2, \ldots, Q$ puntos de cuadratura. Esta aproximación se conoce como la cuadratura clásica.

Torres-Saavedra (2006) implementó la función "phermite" en *R* para el cálculo de las abscisas y los pesos de la cuadratura de Gauss-Hermite clásica. Estos valores son simétricos alrededor de cero y se pueden encontrar en libros de análisis numérico. A veces es necesario tener muchos puntos para lograr

una buena aproximación a la integral, ya que las abscisas pueden quedar en una región inadecuada. La cuadratura adaptativa nos proporciona abscisas en una región apropiada donde $\tilde{\mu}$ será la forma del integrando $g(x)$ y $\tilde{\sigma} = \frac{1}{\sqrt{\tilde{q}}}$,

$$\tilde{q} = -\frac{\partial^2}{\partial x^2} \log g(x)\Big|_{x=\tilde{\mu}} \tag{3.66}$$

Utilizando $\phi(x\,|\,\tilde{\mu}, \tilde{\sigma})$ la aproximación es:

$$\int g(x) \approx \sqrt{2}\tilde{\sigma} \sum_{q=1}^{Q} \omega_q^* \, g\left(\tilde{\mu} + \sqrt{2}\tilde{\sigma} x_q\right) \tag{3.67}$$

donde,

$$\omega_q^* = \omega_q \exp\left(x_q^2\right) \tag{3.68}$$

Cuando en la ecuación (3.67) se aplica con una sola abscisa, el resultado de la aproximación es la integral de Laplace.

$$\int g(x) \approx \sqrt{2\pi}\tilde{\sigma}\, g(\tilde{\mu}) \tag{3.69}$$

Por lo tanto, la cuadratura de m-orden de Gauss-Hermite es una alternativa de la aproximación de Laplace de m-orden. Algunos resultados de las simulaciones sugieren un número grande de abscisas para obtener una alta precisión en la cuadratura clásica (100 puntos o más), mientras que la cuadratura adaptativa proporciona una buena precisión con 20 puntos o menos (Demidenko, 2004).

La aproximación de la integral para calcular la verosimilitud es muy sencilla, y por lo tanto, la maximización numérica de la verosimilitud puede ser evaluada con precisión. Esta aproximación funciona relativamente bien en situaciones simples, con un efecto aleatorio. Sin embargo, no para estructuras más complicadas. Además, podría ser computacionalmente muy pesada para dos o más efectos aleatorios. En algunos casos Cadenas de Markov Monte Carlo (MCMC) son una buena alternativa para estimar los parámetros en estos modelos. Algunos de los mejores algoritmos conocidos son Monte Carlo EM (MCEM), Monte Carlo Newton Raphson (MCNR), y simulación de máxima verosimilitud

.

Existen paquetes de software estadísticos para la aproximación de Gauss-Hermite, como SAS PROC NLMIXED y varias funciones de *R*.

Aproximación del integrando: Aproximación de Laplace

El objetivo cuando se aproximan los integrandos es obtener una integral tratable de tal manera que se puedan obtener expresiones de forma cerrada, para que la maximización de la aproximación de la verosimilitud sea posible. Se han propuesto varios métodos, pero en el fondo todos se reducen a las aproximaciones de Laplace. El método de Laplace se ha diseñado para aproximar integrales de la forma:

$$I = \int e^{-Q(v)} dv \tag{3.70}$$

donde la función $Q(v)$ es conocida, unimodal y acotada. Sea $\tilde{v}$ el valor de v para que Q se minimice. Entonces, la expansión de Taylor de segundo orden de $Q(v)$ alrededor de $\tilde{v}$ es de la forma:

$$Q(v) \approx Q(\tilde{v}) + \frac{1}{2}(v - \tilde{v})' Q''(\tilde{v})(v - \tilde{v}) \tag{3.71}$$

donde $Q''(\tilde{v})$ es el Hessiano de Q evaluado en $\tilde{v}$. Reemplazando $Q(v)$ en la ecuación (3.70) tenemos:

$$I \approx (2\pi)^{q/2} |-Q''(\tilde{v})|^{-1/2} e^{-Q(\tilde{v})} \tag{3.72}$$

Cuando Q es bimodal, es necesario utilizar una mejor aproximación de Laplace. En este método se utilizan tantas estimaciones de v como sea necesario de acuerdo a las diferentes formas de la función Q.

Claramente se observa que la integral en la ecuación (3.61) es proporcional a cada integral I, para funciones $Q(v)$ de la forma:

$$Q(v) = \phi^{-1} \sum_{j=1}^{n_i} \left[y_{ij} \left(x_j'\beta + z_j' v \right) - \psi \left(x_j'\beta + z_j' v \right) \right] - \frac{1}{2} v' \Sigma^{-1} v \tag{3.73}$$

de tal manera que el método de Laplace se puede aplicar.

Aproximación de los Datos: Cuasi-verosimilitud penalizada y marginal

Cuasi-verosimilitud penalizada (PQL): Para MLG, la máxima cuasi-verosimilitud es un buen método, por su capacidad para generar estimadores eficientes sin suponer distribuciones. La cuasi-verosimilitud no especifica una distribución, sólo la media y la varianza. Para MLGM, la función a maximizar es modificada por un término de la penalidad debido a los efectos aleatorios, y por lo tanto,

se le llama cuasi-verosimilitud penalizada (PQL). En la literatura, se pueden encontrar diferentes versiones para MLGM. Nosotros trabajaremos con la presentada por Breslow (1984).

En un contexto de MLG, y en base a la ecuación (3.58), para el MLGM la forma de la cuasi-verosimilitud viene dada por:

$$QL\left(y_{ij}\,,\,\underline{v_i}\right) = QL\left(y_{ij}\,|\,\underline{v_i}\right)\cdot L\left(\underline{v_i}\right) \tag{3.74}$$

Luego, la función integrada de cuasi-verosimilitud (PQL) es:

$$\text{PQL} \;= (2\pi)^{-1/2}\,|\Sigma|^{-1/2}\int \exp\left[-\frac{1}{2\phi}\sum_{i=1}^{N} d_i\left(y_i,\tilde{\mu}_i\right) - \frac{1}{2}\underline{v_i}'\,\Sigma^{-1}\,\underline{v_i}\right] d\underline{v_i} \tag{3.75}$$

donde

$$d_i\left(y_i,\tilde{\mu}_i\right) = -2\int_{y_i}^{\tilde{\mu}_i}\frac{y_i - t}{a_i xi(t)}dt \tag{3.76}$$

denota la deviance ponderada. Como $\boldsymbol{y}$ es Binomial Negativa, esta expresión contiene integrales que deben ser resueltas numéricamente con métodos tales como la aproximación de Laplace.

La ecuación (3.75) tiene la forma $C\,|\Sigma|^{-1/2}\int \exp(Q(v)db$, con C una constante, y Q en función de v. Por lo tanto, la aproximación de Laplace se puede aplicar a la aproximación de esta integral. Utilizando el resultado de la ecuación (3.72), el log-cuasiverosimilitud viene dado por:

$$\log\left(\text{pql}\right) \approx -\frac{1}{2}\log|\,\Sigma\,| - \frac{1}{2}\log\left|\,Q''\left(\underline{\tilde{v}_i}\right)\right| - Q\left(\underline{\tilde{v}_i}\right) \tag{3.77}$$

donde $\underline{\tilde{v}_i}$ es la solución de:

$$Q'\left(\underline{v_i}\right) = -\sum_{i=1}^{N}\frac{\left(y_i - \tilde{\mu}_i\right)z_i}{\phi a_i xi\left(\tilde{\mu}_i\right)g'\left(\tilde{\mu}_i\right)} + \Sigma^{-1}\,\underline{v_i} = 0 \tag{3.78}$$

que minimiza $Q\left(\underline{v_i}\right)$. Además, la aproximación de la segunda derivada de Q viene dada por:

$$Q''\left(\underline{v_i}\right) = \sum_{i=1}^{N}\frac{z_i z_i'}{\phi a_i xi\left(\tilde{\mu}_i\right)\left[g'\left(\tilde{\mu}_i\right)\right]^2} + \Sigma^{-1} \;+ R = Z'W\,Z + \Sigma^{-1} \tag{3.79}$$

Por lo tanto, reemplazando en la ecuación (3.77) el log-cuasiverosimilitud es:

$$\log\left(\text{pql}\right) \approx -\frac{1}{2}\log\left|\,\Sigma\,\right| - \frac{1}{2}\log\left|\,I + Z'W\,Z\,\Sigma\,\right| - \frac{1}{2\phi}\sum_{i} d_i\left(y_i,\widehat{\tilde{\mu}_i}\right) - \frac{1}{2}\underline{\tilde{v}_i}'\,\Sigma^{-1}\,\underline{\tilde{v}_i} \tag{3.80}$$

donde $\underline{\tilde{v}_i}$ se elige para maximizar los dos últimos términos, I es la matriz identidad, y W es la matriz diagonal $N \times N$ con términos en la diagonal de la forma: $\omega_i = \left\{\phi a_i xi\left(\widehat{\mu_i}\right)\left[g'\left(\tilde{\mu}_i\right)\right]^2\right\}^{-1}$. Esta expresión da lugar a que:

$$\sum_{i=1}^{N} \frac{\left(y_i - \tilde{\mu}_i\right) x_i}{\phi a_i xi\left(\tilde{\mu}_i\right) g'\left(\tilde{\mu}_i\right)} = 0 \tag{3.81}$$

y

$$\sum_{i=1}^{N} \frac{\left(y_i - \tilde{\mu}_i\right) z_i}{\phi a_i xi\left(\tilde{\mu}_i\right) g'\left(\tilde{\mu}_i\right)} = \Sigma^{-1} \underline{v_i} \tag{3.82}$$

Las soluciones de las ecuaciones (3.81) y (3.82) se basan en el vector Y^* siendo,

$$Y_i^* = \left(x_i' \widehat{\tilde{\beta}} + z_i \widehat{\underline{v_i}}\right) + \left(y_i + \widehat{\tilde{\mu}_i}\right) g'\left(\widehat{\tilde{\mu}_i}\right) \tag{3.83}$$

Una alternativa para derivar del algoritmo **PQL** fue desarrollado por Breslow (2003) en el contexto de datos longitudinales. Este método propone la utilización de una linealización de la media condicional en función de los efectos fijos y aleatorios, con una expansión de la serie de Taylor de la media que es una función no lineal del predictor lineal. Para nuestro modelo es:

$$y_{ij} = \upsilon_{ij} + \varepsilon_{ij} = h\left(x_j' \tilde{\beta} + z_j' \underline{v_i}\right) + \varepsilon_{ij} \tag{3.84}$$

Consideremos una expansión lineal de Taylor de la ecuación anterior, donde β son los efectos fijos, v_i los efectos aleatorios y $h = g^{-1}$. Por lo tanto,

$$\begin{aligned} y_{ij} \approx\ & h\left(x_j' \widehat{\tilde{\beta}} + z_j' \widehat{\underline{v_i}}\right) \\ & + h'\left(x_j' \widehat{\tilde{\beta}} + z_j' \widehat{\underline{v_i}}\right) x_j' \left(\tilde{\beta} - \widehat{\tilde{\beta}}\right) \\ & + h'\left(x_j' \widehat{\tilde{\beta}} + z_j' \widehat{\underline{v_i}}\right) z_j' \left(\underline{v_i} - \widehat{\underline{v_i}}\right) + \varepsilon_{ij} \\ =\ & \widehat{\upsilon_{ij}} + \ \text{var}\left(\widehat{\upsilon_{ij}}\right) x_j' \left(\tilde{\beta} - \widehat{\tilde{\beta}}\right) + \ \text{var}\left(\widehat{\upsilon_{ij}}\right) z_j' \left(\underline{v_i} - \widehat{\underline{v_i}}\right) + \varepsilon_{ij} \end{aligned} \tag{3.85}$$

donde,

$$\widehat{\upsilon} = h\left(x_j' \widehat{\tilde{\beta}} + z_j' \widehat{\underline{v_i}}\right) \tag{3.86}$$

para la media condicional $E\left(y_{ij} \,|\, \underline{v_i}\right)$. Reescribiendo la ecuación (3.85) en notación de vectores, y reordenando los términos tenemos:

$$Y_i^* \equiv \widehat{V}_i^{-1}\left(Y_i - \widehat{\upsilon_i}\right) + X_i \widehat{\beta} + Z_i \widehat{\underline{v_i}} \approx X_i \beta + Z_i \underline{v_i} + \varepsilon_i^* \tag{3.87}$$

donde $\varepsilon_i^* = \widehat{V}_i^{-1}\varepsilon_i$ con media cero. La ecuación anterior se puede ver como un modelo lineal mixto para seudo-datos de Y_i, con efectos fijos β, efectos aleatorios $\underline{v_i}$ y los términos de error ε_i^*. Esto produce estimaciones para un MLGM. Teniendo en cuenta los valores iniciales para los parámetros β y Σ en la verosimilitud marginal, las estimaciones empíricas Bayesianas para $\underline{v_i}$ se calculan mediante la función de densidad posterior y los seudo-datos de Y_i^*. Entonces, la aproximación del modelo lineal mixto presentado en la ecuación (3.87) se va modificando, luego, las estimaciones de los parámetros se actualizan. Este proceso se repite hasta que se alcanza la convergencia. Las estimaciones resultantes se denominan estimaciones cuasi-verosimilitud penalizada, ya que se obtienen de la optimización de la función de cuasi-verosimilitud mediante aproximaciones de primer y de segundo orden.

Cuasi-verosimilitud marginal: Es una aproximación muy similar al método de **PQL**, la diferencia es que ésta se basa en una expansión lineal de Taylor de la media υ_{ij} alrededor de las estimaciones actuales de β en los efectos fijos, y alrededor de $\underline{v_i} = 0$ para los efectos aleatorios. Este método produce expresiones similares al **PQL**, donde υ_{ij} toma la forma $h\left(x_j'\widehat{\beta}\right)$ en lugar de $h\left(x_j'\widehat{\beta} + z_j'\widehat{\underline{vi}}\right)$. En este caso los seudo datos toman la forma:

$$Y_i^* \equiv \widehat{V}_i^{-1}\left(Y_i - \widehat{\upsilon_i}\right) + X_i\widehat{\beta} \tag{3.88}$$

satisfacen la aproximación del modelo lineal mixto,

$$Y_i^* \approx X_i\beta + Z_i\underline{v_i} + \varepsilon_i^* \tag{3.89}$$

Las estimaciones resultantes son llamadas estimaciones cuasi-verosimilitud marginales **MQL**.

Capítulo 4

Estudios de Simulación

En este capítulo trabajaremos simulaciones del MLGM Binomial Negativo, buscando encontrar propiedades importantes que no se pueden obtener de forma analítica. Consideramos 3 modelos diferentes:

1. Regresión lineal con un efecto aleatorio de intercepto.
2. Regresión lineal con efectos aleatorios de intercepto y de pendiente.
3. Efecto de tratamiento (2 niveles) y un efecto aleatorio de intercepto.

Nos concentraremos en estudiar las propiedades de las distribuciones marginales de y inducidas por el modelo.

$$f(y_j) = \int \prod_{i=1}^{N} \left(\frac{\Gamma\left(y_{ij} + 1/\alpha\right)}{\Gamma(y_{ij}+1)\Gamma(1/\alpha)} \right) \left(\frac{\left(\alpha e^{x_j\underline{\beta} + z_j\underline{v_i}}\right)^{y_{ij}}}{\left(1 + \alpha e^{x_j\underline{\beta} + z_j\underline{v_i}}\right)^{y_{ij}+1/\alpha}} \right) \left(\frac{e^{-\frac{1}{2}\underline{v_i} D \underline{v_i}}}{|D|^{1/2}(2\pi)^{1/2}} \right) dv_i \tag{4.1}$$

Encontramos para la distribución marginal, el cuartil inferior, la mediana, la media, el cuartil superior, el percentil 90, el percentil 95, la asimetría (*Skewness*) y la desviación estándar.

Para la distribución condicional la analizaremos cuando los efectos aleatorios son cero y encontraremos la media y la desviación estándar. Luego haremos algunas comparaciones entre las medias marginales, las medianas marginales y las medias condicionales.

La asímetria existente en la distribución marginal, se analizó por medio del coeficiente de asimetría (*Skewness*).

Coeficiente de Asimetría (*Skewness*): se define como el cociente entre el tercer momento de la distribución y el cubo de la desviación estándar:

$$\frac{E(y-\mu)^3}{\sigma^3} \tag{4.2}$$

Note en la fórmula que las potencias son cúbicas, por lo tanto, los signos de las desviaciones se mantendrán. Lo que nos lleva a concluir si la asimetría es positiva o negativa.

Por lo general el coeficiente de asimetría (*Skewness*) varía entre -4 y 4, aunque podemos encontrar valores extremos cuando tenemos asimetrías muy altas. La asímetria (*Skewness*) también se puede interpretar como la tendencia a que un extremo de la población sea más denso que el otro.

Las dos primeras simulaciones se realizarón con el siguiente algoritmo general:

Algoritmo

1. Creamos los datos:
    1.1 Generamos v, el efecto aleatorio para un sujeto específico i, $(v_i \sim N(0, \sigma_v))$.
        1.1.1 Creamos el vector $x_j = \{0, 2, 4, \ldots, 20\}$
        1.1.2 Creamos la media condicional μ para cada valor de j
        1.1.3 Creamos a $p = \frac{1}{1+\alpha\mu}$ y $k = \frac{1}{\alpha}$
        1.1.4 Generamos la variable y con la distribución Binomial Negativa para cada valor de j, $(y_j \sim bn(p, k))$
    1.2 Repetimos el proceso para 20 sujetos.
2. Análisis de los datos: Los datos se analizaron por máxima verosimilitud.
    2.1 Estimamos los parámetros para el conjunto de datos.
    2.2 Estimamos el valor predicho de la media condicional para cuando los efectos aleatorios sean cero.
3. Repetimos los pasos 1 y 2, 1000 veces.

4. Estimación de la distribución marginal.

    4.1 Estimamos los paramétros, los cuantiles, la mediana, la media, el percentil 90, el percentil 95, la asimetría (*Skewness*) y la desviación estándar.

Así x_j es la variable independiente fija y v_i es el efecto aleatorio de cada sujeto. Para estos modelos tomamos el vector $\underline{\beta}$ constante, α y (σ_v ó D) varían con valores dados en las tablas de los resultados de cada modelo. Este escenario simula una situación en la que hay 11 mediciones tomadas a cada uno de los 20 sujetos aleatoriamente elegidos.

Los datos fueron generados por medio del Proc IML con el programa SAS(v.9.1.3), donde el efecto aleatorio se generó con la función ***rannor*** y la variable dependiente ***y*** por medio de la función ***randgen(negbin)*** . Luego realizamos 1000 simulaciones Monte Carlo para esta sitación, estos datos simulados fueron analizados con el Proc NLMIXED con el método de máxima-verosimilitud.

Es importante resaltar que para propósitos de la estimación se utilizó una reparametrización para las desviaciones estándar de los efectos aleatorios y para el coeficiente de correlación, usando en la estimación una parametrización. Tal que el dominio sea $\mathbb{R}$, de la siguiente manera:

1. $\sigma_v = \exp(s_v)$
2. $\rho = \dfrac{\exp(2z) - 1}{\exp(2z) + 1}$

Las tablas que se muestran más adelante en los resultados de cada modelo son características de la distribución marginal, donde se le agregó una columna de la media condicional cuando los efectos aleatorios son cero.

4.1. Modelo con un efecto aleatorio de intercepto

Este es un modelo de regresión lineal donde en el predictor lineal se incluye un efecto aleatorio de intercepto. Planteado así:

$$y_{ij}|v_i \sim bn\left(\alpha, \xi_{ij}\right) \tag{4.3}$$

El factor de efectos aleatorios es:

$$v_i \sim N\left(0, \sigma_v^2\right) \tag{4.4}$$

y la función de enlace y el predictor lineal son:

$$\ln(\xi_{ij}) = \beta_0 + v_i + \beta_1 x_j \tag{4.5}$$

4.1.1. Resultados

Se consideraron diferentes escenarios variando los parámetros α y σ_v, donde los parámetros varían entre 0.2 y 3. Los demás parámetros son constantes con los valores de $\beta_0 = -1.5$ y $\beta_1 = 0.5$. Para poder hacer un resumen de los resultados y no tener tablas muy grandes, sólo mostramos las distribuciones marginales para los valores de $x = 0, 10$ y 20. En los demás valores de x observamos conclusiones similares.

α	σ_v	x	Q_1	Mediana	Media condicional ($v = 0$)	Media Marginal	Q_3	P_{90}	P_{95}	Asimetría (*Skewness*)	Desviación estándar
0.2	0.2	0	0	0	0.23	0.23	0	1	1	2.31	0.49
		10	21	31	33.23	33.92	43	57	67	1.11	17.79
		20	3193.5	4544	4915	5003.11	6315	8311	9722.5	1.16	2502.41
	0.5	0	0	0	0.22	0.25	0	1	1	2.49	0.54
		10	18	30	33.23	37.23	48	71	91	2.38	28.17
		20	2823	4498.5	4944.37	5561.31	7082	10578.5	13364	2.09	4068.97
	0.8	0	0	0	0.23	0.31	0	1	2	4.23	0.67
		10	16	30	33.54	45.12	56	97	135	3.95	50.28
		20	2363	4461.5	4987.36	6713.82	8282.5	14339.5	19903.5	5.25	7624.93
	1.2	0	0	0	0.24	0.48	1	1	2	8.16	1.19
		10	12	31	35.38	70.48	74	164	249	13.08	145.67
		20	1891	4532	5250.33	10501.6	10794	23952	38308	10.98	21571.58
	2	0	0	0	0.25	1.86	1	3	5	126.38	30.87
		10	8	31	37.83	322.06	126	433.5	911	135.89	9398.66
		20	1157.5	4486.5	5623.33	39583.63	18743	63976	132497	87.94	400633.53
	3	0	0	0	0.28	16.79	2	10	30	113.57	425.95
		10	4	30	41.89	2677.08	230	1437	4440	118.14	78241.42
		20	590	4482	6234.1	321601.48	34103.5	217254.5	644086.5	54.67	4558417.85

Tabla 4.1: Distribuciones marginales variando σ_v. Donde $\ln(\xi_{ij}) = -1.5 + v_i + 0.5x_j, \quad x = \{0, 10, 20\}$

α	σ_v	x	Q_1	Mediana	Media condicional ($v = 0$)	Media Marginal	Q_3	P_{90}	P_{95}	Asimetría (*Skewness*)	Desviación estándar
0.5	0.2	0	0	0	0.22	0.23	0	1	1	2.36	0.51
		10	15	27	33.01	33.19	45	67	83	1.6	25.59
		20	2298	4039.5	4924.34	4955	6602.5	9801	12130.5	1.64	3686.05
0.8		0	0	0	0.23	0.23	0	1	1	2.8	0.53
		10	11	25	33.13	34.13	47	76	98	2.06	32.07
		20	1734	3643	4922.2	5057.41	6823.5	11088.5	14419	2.05	4759.62
1.2		0	0	0	0.22	0.23	0	1	2	2.87	0.54
		10	7	21	33.01	33.46	46	82	109	2.5	38.43
		20	1128	3112.5	4969.46	5105.27	6940	12460	16705	2.45	5887.79
2		0	0	0	0.23	0.23	0	1	1	3.33	0.58
		10	3	15	32.78	33.6	43	92	130	3.25	50.04
		20	506.5	2249	4881.52	4941.1	6431.5	13276.5	19038	3.06	7111.74
3		0	0	0	0.23	0.23	0	1	1	3.81	0.62
		10	1	10	33.01	33.31	39	95	147	3.58	58.6
		20	1158	1418.5	4999.19	5060.26	5844	14610.5	22470	3.87	9081.89

Tabla 4.2: Distribuciones marginales variando α. Donde $\ln(\xi_{ij}) = -1.5 + v_i + 0.5x_j, \quad x = \{0, 10, 20\}$

α	σ_v	x	Q_1	Mediana	Media condicional ($v = 0$)	Media Marginal	Q_3	P_{90}	P_{95}	Asimetría (*Skewness*)	Desviación estándar
0.5	0.5	0	0	0	0.22	0.25	0	1	1	2.71	0.56
		10	14	27	33.15	37.59	49	80	105	3.25	37.13
		20	2110	4046	4966.9	5625.94	7337.5	11985	15674.5	2.64	5345.45
1	1	0	0	0	0.23	0.36	0	1	2	6.74	0.96
		10	7	21	33.40	54.91	56	133	215	10.51	118.73
		20	1041	3088.5	4980.85	7921.83	8395.5	19986	33153	7.43	15866.47
2	1	0	0	0	0.24	0.36	0	1	2	9.05	1.12
		10	2	13	34.12	54.58	49	137	236	8.08	138.04
		20	412	2009	5102.48	8003.06	7110	19400	33398.5	12.21	22428.5
2	2	0	0	0	0.26	1.47	0	2	6	77.01	14.33
		10	1	12	38.1	242.28	70	322	777.5	37.97	2105.43
		20	232	1749	5762.34	38744.72	10540	49346.5	121555.5	65.07	411252.93
3	2	0	0	0	0.25	1.51	0	2	5	23.38	10.91
		10	0	7	36.82	284.37	57	306.5	768.5	39.37	3182.17
		20	87	1048.5	5662.79	37422.03	8419	44109.5	112874	36.02	34424.64

Tabla 4.3: Distribuciones marginales variando α y σ_v. Donde $\ln(\xi_{ij}) = -1.5 + v_i + 0.5x_j, \quad x = \{0, 10, 20\}$

4.2. Modelo con efectos aleatorios de intercepto y de pendiente

Este es un modelo de regresión lineal donde en el predictor lineal se incluye un efecto aleatorio de intercepto y de pendiente. Planteado así:

$$Y_{ij} \mid \underline{v_i} \sim bn\left(\alpha, \xi_{ij}\right) \tag{4.6}$$

El vector de efectos aleatorios independientes es una Normal bivariada.

$$\underline{v_i} = (s_i, p_i) \sim N(0, D) \tag{4.7}$$

donde,

$$D = \begin{pmatrix} \sigma_s^2 & \rho\sigma_s\sigma_p \\ \rho\sigma_s\sigma_p & \sigma_p^2 \end{pmatrix} \tag{4.8}$$

La función de enlace y el predictor lineal son:

$$\ln(\xi_{ij}) = (\beta_0 + s_i) + (\beta_1 + p_i)\, x_j \tag{4.9}$$

Se consideraron diferentes escenarios variando los parámetros α, σ_s y σ_p

4.2.1. Resultados

Se consideraron diferentes escenarios variando los parámetros α, σ_s y σ_p; donde α varía entre 0.2 y 2, σ_s varía entre 0.2 y 1.2, y σ_p varía entre 0.05 y 0.15. Los demás parámetros son constantes $\beta_0 = -2$, $\beta_1 = 0.2$ y $\rho = 0.5$. Para poder hacer un resumen de los resultados y no tener tablas muy grandes, sólo mostramos las distribuciones marginales para los valores de $x = 0, 10$ y 20. En los demás valores de x observamos conclusiones similares.

α	σ_s	σ_p	x	Q_1	Mediana	Media condicional ($v = 0$)	Media Marginal	Q_3	P_{90}	P_{95}	Asimetría (*Skewness*)	Desviación estándar
0.5	0.2	0.05	0	0	0	0.14	0.15	0	1	1	2.88	0.41
			10	0	1	1.09	11.96	4	17	39	45.92	102.25
			20	0	6	11.87	177318.7	121	1763.5	8484	99.09	14057881.6
1			0	0	0	0.14	0.15	0	1	2	3.42	0.42
			10	0	1	1.11	12.69	4	15	40.5	29.14	101.34
			20	0	5	13.21	39049.18	108.5	1655.5	8603.5	38.95	839513.73
1.5			0	0	0	0.13	0.15	0	1	1	3.98	0.42
			10	0	0	1.13	19.37	3	14	36	47.46	421.91
			20	0	4	13.59	122285.1	67	1200	7498	48.65	4960867.15
2			0	0	0	0.14	0.16	0	1	1	4	0.48
			10	0	0	1.41	208.76	6	63	221	20.93	2309.57
			20	0	2	28.61	8209111	365	29721	481986	14.57	101382047

Tabla 4.4: Distribuciones marginales variando α. Donde $\ln(\xi_{ij}) = (s_i - 2) + (p_i + 0.2)\, x_j, \quad x = \{0, 10, 20\}$

α	σ_s	σ_p	x	Q_1	Mediana	Media condicional ($v = 0$)	Media Marginal	Q_3	P_{90}	P_{95}	Asimetría (*Skewness*)	Desviación estándar
0.2	0.2	0.05	0	0	0	0.14	0.15	0	1	1	2.82	0.40
			10	0	1	1.13	12.12	4	17	41	41	101.92
			20	0	7	13.06	55088.04	139	1993	10316.5	96.94	2642402.71
	0.5		0	0	0	0.14	0.17	0	1	1	3.11	0.45
			10	0	1	1.14	15.45	5	20	46	56.03	155.95
			20	0	7	13.27	68925.67	151.5	2317.5	12331	96.5	3610695.3
	0.8		0	0	0	0.14	0.2	0	1	1	3.53	0.51
			10	0	1	1.17	15.91	5	20	48	30.92	130.32
			20	0	7	13.67	32295.84	141	2176.5	11724	46.48	684111.89
	1.2		0	0	0	0.14	0.25	0	1	1	4.58	0.65
			10	0	1	1.23	21.74	5	22	54	71.14	351.19
			20	0	7	14.19	83908.44	146	2402.5	12483	79.78	3673229.02

Tabla 4.5: Distribuciones marginales, variando los valores de σ_s. Donde $\ln(\xi_{ij}) = (s_i - 2) + (p_i + 0.2)\, x_j, \quad x = \{0, 10, 20\}$

α	σ_s	σ_p	x	Q_1	Mediana	Media condicional ($v = 0$)	Media Marginal	Q_3	P_{90}	P_{95}	Asimetría (*Skewness*)	Desviación estándar
0.2	0.2	0.05	0	0	0	0.14	0.15	0	1	1	2.82	0.40
			10	0	1	1.13	12.12	4	17	41	41	101.92
			20	0	7	13.06	55088.04	139	1993	10316.5	96.94	2642402.71
		0.1	0	0	0	0.14	0.14	0	1	1	2.74	0.38
			10	0	1	1.31	73.98	8	57	189	16.78	576.92
			20	0	7	24.31	2062435	566	23835	241287	26.35	34620960
		0.15	0	0	0	0.14	0.15	0	1	1	3.06	0.41
			10	0	1	1.44	167.52	12	99	401	18.29	1310.28
			20	0	7	48.89	6374296.8	1029	89995	1197491	15.91	71340479.7

Tabla 4.6: Distribuciones marginales, variando los valores de σ_p. Donde $\ln(\xi_{ij}) = (s_i - 2) + (p_i + 0.2)\, x_j, \quad x = \{0, 10, 20\}$

4.3. Modelo de efecto de tratamiento (2 niveles) y un efecto aleatorio de intercepto.

En este modelo vamos a analizar la potencia de la prueba para las dos medias de los tratamientos, el modelo lo planteamos así:

$$Y_{ij} \mid v_j \sim bn\left(\alpha, \xi_{ij}\right) \tag{4.10}$$

El factor de efectos aleatorios es:

$$v_j \sim N\left(0, \sigma_v^2\right) \tag{4.11}$$

La función de enlace y el predictor lineal son:

$$\ln(\xi_{ij}) = \mu + \tau_i + v_j \tag{4.12}$$

Esta simulación se realizó con el siguiente algoritmo general:

Algoritmo

1. Creamos los datos:

    1.1 Creamos el siguiente preceso para el tratamiento 1.

        1.1.1 Generamos v, el efecto aleatorio para un sujeto específico j, $(v_i \sim N(0, \sigma_v))$.

            1.1.1.1 Creamos un vector de n observaciones.

            1.1.1.2 Creamos la media condicional μ para cada observación.

            1.1.1.3 Creamos a $p = \dfrac{1}{1+\alpha\mu}$ y $k = \dfrac{1}{\alpha}$

            1.1.1.4 Generamos la variable y con la distribución Binomial Negativa para cada observación, $(y_j \sim bn(p, k))$

            1.1.1.5 Repetimos el proceso para 3 observaciones.

        1.1.2 Repetimos el proceso para 10 sujetos.

    1.2 Repetimos el proceso para el tratamiento 2.

2. Análisis de los datos: Los datos se analizaron por pseudo-verosimilitud.

2.1 Estimamos los paramétros para el conjunto de datos.

2.2 Estimamos la media condicional y la convergencia de la prueba.

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

4. Estimación de la distribución marginal.

4.1 Estimamos los paramétros, la media de convergencia, la media de la potencia de la prueba.

La potencia de la prueba F de $\tau_1 = \tau_2$ se calcula como la proporción de veces que el p-valor fue inferior a 0.05. En esta simulación los datos fueron analizados con Proc GLIMMIX y pseudo-verosimilitud por el método PQL.

4.3.1. Resultados

Se generaron varios escenarios, realizando variaciones en los parametros α, μ, σ_v y $\tau_1 - \tau_2$; donde α varía entre 0.1 y 3.5, μ varía entre 2 y 16, σ_v varía entre 0.1 y 2, y $\tau_1 - \tau_2$ varía entre 0.1 5.

En las tablas que se muestran a continuación denotamos a μ_1 y σ_1, μ_2 y σ_2 como las medias y los errores estándar de las medias de cada tratamiento respectivamente.

μ	$\tau_1 - \tau_2$	α	σ_v	Potencia de la prueba	$\frac{\mu_1}{\mu_2} = \exp(\tau_1 - \tau_2)$
2	0.1	0.1	0.1	0.1	1.1
2	0.1	0.5	0.1	0.1	1.1
2	0.5	0.5	0.1	0.7	1.7
2	0.5	0.5	0.5	0.4	1.7
2	1	0.5	0.1	1	2.7
2	1	1.5	0.1	0.8	2.7
2	5	0.5	0.1	1	148
2	5	0.5	0.5	1	148

Tabla 4.7: Diferentes escenarios para la potencia de la prueba.

μ	$\tau_1 - \tau_2$	α	σ_v	Potencia de la prueba	$\frac{\mu_1}{\mu_2} = \exp(\tau_1 - \tau_2)$
2	0.5	0.5	0.1	0.7	1.7
2	0.5	1.5	0.1	0.3	1.7
2	0.5	0.5	0.5	0.4	1.7
2	0.5	1.5	0.5	0.3	1.7
4	0.5	0.5	0.1	0.7	1.7
6	0.5	0.5	0.1	0.7	1.7
8	0.5	0.5	0.1	0.7	1.7
10	0.5	0.5	0.1	0.7	1.7
16	0.5	0.5	0.1	0.8	1.7

Tabla 4.8: Potencia de la prueba aumentando μ.

μ	$\tau_1 - \tau_2$	α	σ_v	μ_1	μ_2	σ_1	σ_2
2	1	0.5	0.1	7.25	19.65	1.11	2.8
		1.5		7.20	19.17	1.82	4.61
		2.5		7.13	19.73	2.28	5.86
		3.5		7.70	19.99	2.82	7.40
		0.5	0.5	7.33	19.73	1.64	4.25
			1	7.39	19.83	2.62	7.00
			1.5	7.99	21.94	4.03	11.18
			2	8.80	22.79	6.06	15.31

Tabla 4.9: Medias y desviaciones estándar de los tratamientos.

4.4. Discusión

1. Observamos en los dos primeros modelos, que cuando aumentamos α o las desviaciones estándar, la media condicional cuando los efectos aleatorios son cero es más grande que la mediana marginal. Esta situación la podemos justificar de la siguiente manera:

   Analicemos para el primer modelo. Note que:

   $$y_i = \beta_0 + \beta_1 t_j + v_i \tag{4.13}$$

   Si encontramos la esperanza de $\widehat{y_i}$ cuando el efecto aleatorio es cero, tenemos,

   $$\begin{aligned} E[\widehat{y_i} \mid v_i = 0] &= E\left[\exp\left(\widehat{\beta_0} + \widehat{\beta_1} t_j\right)\right] \\ &= E\left[\exp\left(\beta_0 + \varepsilon_{\beta_0} + \left(\beta_1 + \varepsilon_{\beta_1}\right) t_j\right)\right] \\ &= E\left[\exp\left(\beta_0 + \beta_1 t_j\right) \cdot \exp\left(\varepsilon_{\beta_0} + \varepsilon_{\beta_1} t_j\right)\right] \\ &= \exp\left(\beta_0 + \beta_1 t_j\right) \cdot E\left[\exp\left(\varepsilon_{\beta_0} + \varepsilon_{\beta_1} t_j\right)\right] \end{aligned} \tag{4.14}$$

   Ahora, hacemos un cambio de variable, tal que:

   $$u = \varepsilon_{\beta_0} + \varepsilon_{\beta_1} t_j \tag{4.15}$$

   Ahora, de Evens (2005) recordamos la desigualdad de Jensen, la cual dice:

Desigualdad de Jensen:

Sea u una variable aleatoria y sea f una función convexa, entonces,

$$f(E[u]) \leq E\left[f(u)\right] \tag{4.16}$$

Note que u es una variable aleatoria y que la función $f(u) = \exp(u)$ es convexa. Por lo tanto podemos aplicar la desigualdad de Jensen, es decir:

$$\exp\left(E[u]\right) \leq E\left[\exp(u)\right] \tag{4.17}$$

Recordemos que u es una variable aleatoria con esperanza asintótica igual a cero, entonces:

$$1 \leq E\left[\exp(u)\right] \tag{4.18}$$

Retomando la ecuación (4.13) tenemos tenemos que:

$$E[\widehat{y_i} \,|\, v_i = 0] \geq \exp\left(\beta_0 + \beta_1 t_j\right) \tag{4.19}$$

Recuerde que la mediana marginal de la distribución es $P_{50} = \exp\left(\beta_0 + \beta_1 t_j\right)$, por lo tanto, concluimos:

$$E[\widehat{y_i} \,|\, v_i = 0] \geq P_{50} \tag{4.20}$$

De forma similar se podría demostrar para el segundo modelo.

2. Cuando aumentamos σ_v o α en el primer modelo, las medias marginales se mantienen por debajo de cuartil superior, en cambio en el segundo modelo se encuentran por encima del percentil 95. Esta situación la podemos justificar de la siguiente manera:

   Analicemos cuando aumentan las desviaciones estándar de los efectos aleatorios. Recordemos las medias marginales para cada modelo, las cuales són:

   Media marginal del modelo con intercepto aleatorio.

$$E\left[y_{ij}\right] = \exp\left(x_j'\underline{\beta} + \frac{\sigma_v^2}{2}\right) \tag{4.21}$$

   Media marginal del modelo con intercepto y pendiente aleatoria.

$$E\left[y_{ij}\right] = \exp\left(x_j'\underline{\beta} + \frac{x_j' D\, x_j}{2}\right) \tag{4.22}$$

Observamos que las desviaciones estándar de los efectos aleatorios están directamente relacionados con la media marginal en ambos modelos, entonces cuando las desviaciones estándar aumenten, la media marginal deberá aumentar.

Pero, el segundo modelo tiene una forma cuadrática que involucra el vector x, el cual hace que la media marginal sea mucho más grande que en el primer modelo.

Ahora, consideremos el efecto de aumentar α. Recordemos las varianzas de cada modelo.

Varianza marginal del modelo con intercepto aleatorio.

$$\text{var}(y_{ij}) = \mu_{ij} + \alpha {\mu_{ij}}^2 \left[\exp\left(\sigma_v^2\right)\left(1 + \frac{1}{\alpha}\right) - \frac{1}{\alpha} \right] \tag{4.23}$$

Varianza marginal del modelo con intercepto y pendiente aleatoria.

$$\text{var}(y_{ij}) = \mu_{ij} + \alpha {\mu_{ij}}^2 \left[\exp\left(x_j' \, D \, x_j\right)\left(1 + \frac{1}{\alpha}\right) - \frac{1}{\alpha} \right] \tag{4.24}$$

Sabemos que α está directamente relacionado con la varianza de la distribución marginal, y utilizando el mismo argumento anterior podemos entender porqué la media marginal del segundo modelo se hace mucho más grande que la media marginal del primero.

3. En los dos primeros modelos, observamos que la asimetría de la distribución marginal es bastante alta a medida que aumentamos α o los errores estándar de los efectos aleatorios. Este resultado se esperaba, ya que la asimetría esta relacionada directamente con el crecimiento de las medias marginales.

4. Cuando aumentamos α o σ_v ó simultaneamente y mantenemos los demás parámetros constantes, en el tercer modelo, observamos que la potencia de la prueba disminuye, como se esperaba, ya que estos parámetros están directamente relacionados con la variabilidad de los datos.

5. A medida que aumenta α manteniendo los demás parámetros constantes, en el tercer modelo, el error estándar de la medias de los tratamientos crece. Lo mismo se observa cuando σ_v aumenta, manteniendo todos los demás parámetros constantes.

6. No parece haber un patrón claro de cambio en la potencia de la prueba cuando aumenta μ y $\tau_1 - \tau_2$ se mantiene constante (es decir, la proporción se mantiene constante).

$$\frac{\mu_1}{\mu_2} = \exp\left(\tau_1 - \tau_2\right) \tag{4.25}$$

Capítulo 5

Aplicaciones: Comparando recuentos de semillas en el bosque seco de Guánica.

Los estudios ecológicos a menudo se realizan con diseños bastante complejos, donde deben considerarse en el análisis varios niveles de aleatorización, diversos efectos aleatorios, desbalances, entre otros. En muchos de estos casos los datos no tienen distribución Normal, por lo tanto es necesario modelarlos por medio de otras distribuciones. Hemos observado como la Binomial Negativa modela de manera eficiente los datos de recuentos y aplicaremos los resultados obtenidos en esta tesis en un estudio ecológico, que describimos a continuación.

5.1. Descripción del estudio

Wolfe (2009) realizó un estudio para analizar los patrones de recuentos de semillas recolectadas en trampas en un bosque seco caribeño degradado por el uso agrícola y por fuegos. Se seleccionaron cuatro sitios en el bosque estatal de Guánica, Puerto Rico (ver Figura 5.1).

1. La Hoya.
2. Ensenada.
3. Pitirre.
4. Cuevas.

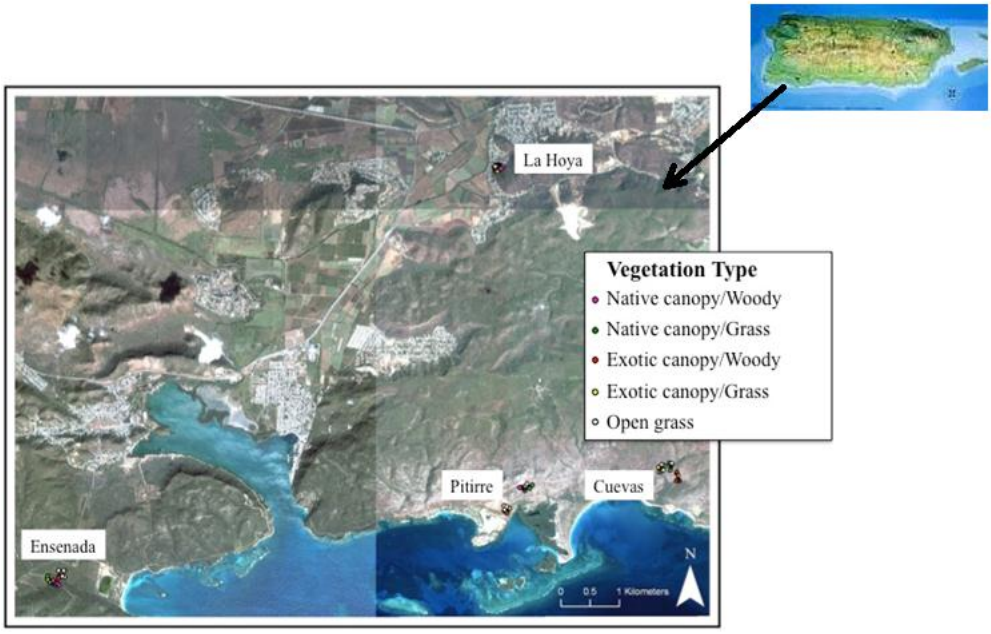

Figura 5.1: Imagen satelital del Bosque de Guánica y sus alrededores. El lugar donde están ubicadas las trampas de semillas fueron tomadas con un GPS y están marcadas con círculos de colores, (Wolfe, 2009)

.

En cada sitio se identificaron cinco tipos de vegetación en un gradiente de degradación.

1. Pastizal exótico abierto (OG) (más degradado)
2. Bosque exótico con pastos exóticos en el sotobosque (EG)
3. Bosque exótico con sotobosque leñoso (EW)
4. Bosque nativo con pastos exóticos en el sotobosque (NG)
5. Bosque nativo con sotobosque leñoso (NW) (menos degradado)

Las semillas se recolectaron mensualmente durante un año, realizando un proceso de: identificación, conteo y clasificación en tres categorías:

1. Dispersas por animales (ANIMAL)
2. Dispersas por el viento (WIND)
3. Leucaena leucocephala (LEUC)

5.2. Objetivo

Se pueden plantear varios objetivos, dependiendo de lo que se quiera analizar. En este ejemplo analizaremos los datos para responder al siguiente objetivo:

Comparar los totales anuales de recuentos de semillas de Leucaena en los cinco tipos de vegetación.

En las siguientes tablas observamos los recuentos de las semillas de Leucaena.

Recuento	**Frecuencia anual**
0	115
1	25
2	13
3	7
4	7
5	6
6-10	13
11-20	9
21-30	14
31-40	11
41-50	10
51-100	22
101-500	35
501-1000	7
Más de 1000	6

Tabla 5.1: Recuentos anuales de semillas de Leucaena

5.3. Método

Se compararon los totales anuales de recuentos de semillas de Leucaena recolectada en los cinco tipos de vegetación. Se consideraron los siguientes factores:

- Vegetación: Tiene 5 niveles y es un factor fijo.
- Sitio: Tiene 4 niveles y es un factor aleatorio.

Modelo lineal generalizado mixto Binomial Negativo

En cada vegetación y sitio hay 5 repeticiones, es decir las trampas donde se recogieron las semillas. Podemos formular un MLGM Binomial Negativo para los recuentos anuales,

$$y_{ijk} \Big| s_i, s\tau_{ij} \sim bn\Big(\alpha, \xi_{ij}\Big) \tag{5.1}$$

Los efectos aleatorios independientes son:

$$\begin{aligned} s_i &\sim N\Big(0, \sigma_s^2\Big) \\ st_j &\sim N\Big(0, \sigma_{s\tau}^2\Big) \end{aligned} \tag{5.2}$$

donde,

$$\ln\Big(\xi_{ij}\Big) = \mu + \tau_j + s_i + s\tau_{ij} \tag{5.3}$$

Para un sitio y vegetación determinados, la cantidad de semillas tiene una distribución condicional Binomial Negativa con media condicional $\exp(\mu + \tau_i + s_j + s\tau_{ij})$ y parámetro de escala $\alpha = 1.120$ con $(p < 0.0001)$.

Cuando se comparan las medias de vegetación se pueden realizar para un sitio "típico", es decir, un sitio en el que los efectos aleatorios son cero. Es decir, para un efecto promedio de sitio $s = 0$ y un efecto promedio de vegetación-sitio $s\tau = 0$, la media condicional es: $\exp(\mu + \tau_i)$.

En este modelo los datos fueron analizados con Proc GLIMMIX (SAS v. 9.1.3) y pseudo-verosimilitud por el método PQL. Luego mediante el uso de simulaciones Monte Carlo con los parámetros estimados, se encontraron las distribuciones marginales.

Comparando el análisis de los datos realizado en este capítulo con el hecho por Wolfe (2009) para los recuentos de semillas recolectadas en el Bosque seco de Guánica, tenemos:

Wolfe aplicó a los recuentos anuales un Modelo lineal generalizado (MLG) con distribución Binomial Negativa, donde se modela el tipo de vegetación y el sitio como factores fijos, luego se incluye en el modelo la interacción entre vegetación-sitio, así como un efecto de la trampa, que fue anidado dentro del sitio (Diseño de parcelas divididas). También realizó contrastes entre los tres tipos de semillas y los cinco tipos de vegetación.

Nosotros, como una contribución adicional a este trabajo, formulamos un modelo lineal generalizado mixto (MLGM) para los recuentos anuales de Leucaena. La distribución condicional de las observaciones es Binomial Negativa, y la distribución de los efectos aleatorios es Normal. En el modelo se incluye el efecto de vegetación como fijo y los efectos de sitio y la interacción de vegetación-sitio como aleatorios. Además a esto, analizamos las distribuciones marginales y las distribuciones condicionales cuando los efectos aleatorios son cero.

5.4. Resultados

Conjunto de datos	WORK COUNTS
Variable respuesta	Count
Distribución	Binomial Negativa
Función de enlace	ln

Tabla 5.2: Información del modelo

-2 Res. ln(pseudo-verosimilitud)	371.23
Chi-cuadrada generalizada	78.07
Chi-cuadrada generalizada / DF	0.82

Tabla 5.3: Estadísticas

Efecto	Num. DF	Den. DF	F	Pr > F
Veg	4	12	4.55	0.0181

Tabla 5.4: Prueba de efectos fijos Tipo III

Veg	Estimado (escala log)	Error estándar	Media	Error estándar de la media
EG	5.1405	1.2645	170.81	215.98
EW	4.678	1.2646	107.55	136.01
NG	0.08671	1.3526	1.0906	1.4751
NW	-0.3712	1.3781	0.6899	0.9508
OG	1.1985	1.3064	3.3152	4.331

Tabla 5.5: Medias condicionales para cada tipo de vegetación cuando los efectos aleatorios son cero

Veg.	Cuantil inferior	Mediana	Media	Cuantil superior	Percentil 90	Percentil 95
EG	16	107	4771.12	663	3389	8844.5
EW	10	67	2218.74	423	2136.5	5670
NG	0	1	21.82	4	22	57
NW	0	0	14.60	3	14	37
OG	0	2	75.39	13	66	178

Tabla 5.6: Distribuciones marginales

5.5. Conclusiones

1. Las distribuciones marginales son muy sesgadas, y la media marginal en cada caso está cerca de P_{90}.

2. La mediana marginal, es inferior a la media condicional calculada para cuando los efectos aleatorios sean cero, es decir, cuando $s = 0$ y $s\tau = 0$.

3. Existe diferencia significativa entre las medias condicionales de los diferentes tipos de vegetación.

Capítulo 6

Conclusiones generales y trabajos futuros

6.1. Conclusiones generales

Dados los resultados obtenidos en el transcurso de este trabajo se observan y resuelven los interrogantes planteados al inicio de la investigación. Dado que existen numerosas propiedades para las distribuciones marginales, en este trabajo nos hemos enfocado en algunas de ellas, las cuales fueron descritas anteriormente, es por esto que las conclusiones obtenidas están sujetas al enfoque que le hemos dado.

1. Cuando comparamos los MLGMs Binomial Negativo y Poisson con los modelos sin efectos aleatorios, notamos que la presencia de los efectos aleatorios inducen un aumento en la media y la varianza marginal en ambas distribuciones. Por lo tanto para recuentos con mayor dispersión es más apropiado el uso del MLGM Binomial Negativo que el MLGM Poisson.

2. Cuando tenemos efectos aleatorios con varianzas altas, observamos que las distribuciones marginales del MLGM Binomial Negativo poseen una asimetría positiva bastante alta, con medias marginales que se ubican por encima del percentil 90.

3. Cuando un MLGM Binomial Negativo posee varios efectos aleatorios, las medias condicionales se hacen más grandes que las medianas de las distribuciones marginales.

6.2. Trabajos futuros

En esta investigación observamos algunas propiedades de las distribuciones marginales, pero existen diversas caraterísticas que se pueden observar, esto se podría conseguir realizando más diversidad de escenarios, analizando los datos no en un valor central, trabajando con la media condicional para otros valores de efectos aleatorios.

Una de las propiedades que se podría estudiar a fondo son las correlaciones existentes entre diferentes tratamientos o medidas repetidas en el mismo sujeto, usando predictores Bayesianos empíricos.

Por último podemos pensar en aplicar el MLGM Binomial Negativo a diferentes tipos de datos, encontrados en cualquier área de investigación donde se observen recuentos con mucha variabilidad.

Bibliografía

Agresti, Alan (2002). *Categorical Data Analysis*, Segunda edición. John Wiley & Sons. Inc. Hoboken. New Jersey.

Booth, James; Casella, George; Friedl, Herwig; Hobert, James (2003).Negative binomial loglinear mixed models. *Statistical Modelling* 3: 179-191.

Breslow, N. E.(2003). *Whither pql?* Paper Series 192. University of Washington.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* 33(1): 38-44.

Breslow, N. E. y Clayton (1993). Approximate Inference in Generalized Linear Models. *Journal of the American Statistical Association*. 88:9-25.

Cameron, A. P. y P. K. Trivedi (1998). *Regression Analysis of Count Data*. Cambridge University Press. New York.

Demidenko, E (2004). *Mixed Models. Theory and Applications*. Primera edición. John Wiley & Sons. Eugene.

Dobson, Annette J. (2002). *An Introduction to Generalized Linear Models*, Segunda edición. Chapman & Hall/CRC. New York.

Evens, Michael J. y Rosenthal, Jeffrey (2005). *Probabilidad y Estadística*. Reverté. Barcelona.

Faraway, J. (2006). *Extending the Linear Model with R*. Boca Raton, FL. Chapman & Hall. New York.

Greene, W. H. (2003). *Econometric Analysis*. Primera edición. Macmillan. New York.

Greene, W. H. (2006). *LIMPED Econometric Modeling Guide*. Version 9. Econometric Software Inc. Plainview. New York.

Greenwood, M. y G. U. Yule (1920). An inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society A*. 83:255-279.

Hardin, J. W. y M. Hilbe (2001). *Generalized Linear Models and Extensions*. Stata Press. College Station. TX.

Hardin, J. W. y M. Hilbe (2003). *Generalized Linear Models and Extensions*. Segunda edición. Stata Press. College Station. TX.

Hilbe, Joseph M. (2008). *Negative Binomial Regression*. Cambridge University Press. New York.

Hilbe, Joseph M. (1993a). *Log negative binomial regression as a generalized linear Model*, Technical Report COS 93/945-26, Departament of Sociology. Arizona State Univestity. Arizona.

Hilbe, Joseph M. (1993b). General linear models. *Stata Technical Bulletin*. STB-11, sg16.

Hilbe, Joseph M. (1994a). General linear models. *The American statistician* 48(3):255-265.

Hoffmann, J. (2004). *Generalized Linear Models*. Allyn & Bacon. Boston.

Johnston, G. (1997). *SAS/STAT/GENMOD Procedure, SAS Institute*. Cary. NC.

Lange, K. (1988). *Numerical Analysis for Statisticians*. Primera edición, Springer. Groningen.

Lawless, J. F. (1987). Negative Binomial and Mixed Poisson Regression. *Canadian Journal of Statistics* 15,(3): 209-225.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Sage. Thousand Oaks. CA.

Long, J. S. y J. Freese (2003,2006). *Regression Models for Categorical Dependent Variables using Stata*. Segunda edición. Stata Press. College Station, TX.

McCullagh, P. y J.A. Nelder (1989). *Generalized Linear Model*. Segunda edición. Chapman & Hall. New York.

McCullagh, P. y J.A. Nelder (1994). *Generalized Linear Model*. Volumen 37 de Monographs on statistics and applied probability. Segunda edición. Chapman & Hall. New York.

Molenberghs, Geert y Verbeke, Geert (2005). *Models for Discrete Longitudinal Data*. Springer. New York.

Nelder, J. A. y R. W. Wedderburn (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, Series A, 135(3): 370-384.

Nelder, J. A. y Y. Lee (1992). Likelihood, and pseudo-likelihood: some comparations, *Journal of the Royal Statistical Society*. Series B, 54: 273-284.

Torres-Saavedra, Pedro A. (2006). *Percentile Curves in Binary Longitudinal Data*. Tesis de Maestría, Departamento de Matemáticas, Universidad de Puerto Rico. Mayagüez.

Wolfe, Brett (2009). *Post-Fire Regeneration in Subtropical Dry Forest of Puerto Rico*. Tesis de Maestría, Departamento de Cultivos y Ciencias Agroambientales, Universidad de Puerto Rico. Mayagüez.