

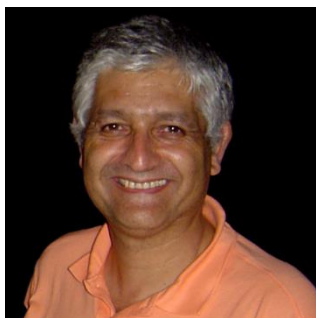
Modelos Lineales **Generalizados Mixtos** Aplicaciones en InfoStat

Julio A. Di Rienzo

Raúl Macchiavelli

Fernando Casanoves

Julio, 2017



Julio A. Di Rienzo es Profesor Asociado de Estadística y Biometría de la Facultad de Ciencias Agropecuarias de la Universidad Nacional de Córdoba, Argentina. Director del grupo de desarrollo de InfoStat y responsable de la implementación de la interfaz con R que se presenta en esta obra (dirienzo@agro.uncor.edu).



Raúl E. Macchiavelli es Catedrático de Biometría en el Facultad de Ciencias Agrícolas, Universidad de Puerto Rico - Mayagüez (raul.macchiavelli@upr.edu)



Fernando Casanoves es el Jefe de la Unidad de Bioestadística del Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). Anteriormente trabajó en la Facultad de Ciencias Agropecuarias de la Universidad Nacional de Córdoba, Argentina, donde participó del desarrollo de InfoStat (casanoves@catie.ac.cr).

Modelos lineales generalizados mixtos: aplicaciones en InfoStat / Julio Alejandro Di Rienzo; Raúl E. Macchiavelli; Fernando Casanoves. - 1a edición especial - Córdoba: Julio Alejandro Di Rienzo, 2017.

Libro digital, PDF - (InfoStat Software Estadístico / Di Rienzo, Julio Alejandro; 2)

Archivo Digital: descarga
ISBN 978-987-42-4985-2

1. Análisis Estadístico. 2. Bioestadísticas. 3. Software. I. Macchiavelli, Raúl E. II. Casanoves, Fernando III. Título CDD 519.5



Contenido

Requerimientos.....	1
Introducción a los modelos lineales generalizados mixtos.....	1
Estimación por máxima verosimilitud completa	3
Invocación del procedimiento de modelos lineales generalizados mixtos	3
Distribución Binomial: Comparación de tratamientos en papaya.....	4
Distribución Binomial: Comparación de curvas de progreso de una enfermedad.....	12
Regresión logística simple: Prediciendo éxito en inseminación artificial	24
Poisson: Ensayo clínico de progabide	28
Poisson con sobredispersión: Comparación de herbicidas.....	47
Regresión Poisson con efectos aleatorios: Análisis de hábitat de escorpiones.....	58
Análisis de recuentos en un diseño de parcelas divididas	66
Análisis de la incidencia de una enfermedad en un diseño alfa latice	74
Estimación de los componentes de varianza.....	82
Gamma: Disfunción eréctil.....	87
Referencias.....	93
Funciones de enlace en modelos lineales generalizados	99
Logit.....	99
Probit.....	99
Complementary log-log	99
Log.....	100
Sqrt.....	100
Identity.....	100
$1/\mu^2$	100
Inverse.....	101

Requerimientos

InfoStat implementa una interfaz amigable de la plataforma **R** para la estimación de modelos lineales generalizados y generalizados mixtos a través de las funciones **glm** as librería *stats* y **glmer** de la librería *lme4* (Bates, Maechler, and Matrix 2011). InfoStat se comunica con R mediante una tecnología de comunicación propia (desarrollada por el Ing. Mauricio Di Rienzo en 2016). Para que InfoStat pueda tener acceso a R, debe estar instalado en su sistema una versión actualizada de R. Para realizar en forma correcta el proceso de instalación consulte la ayuda en línea en el menú *Ayuda* de InfoStat, submenú *¿Cómo instalar R?* y siga las instrucciones brindadas sin omitir ningún paso.

Introducción a los modelos lineales generalizados mixtos

Como en los modelos lineales, tratamos de “explicar” el valor esperado de una observación mediante variables conocidas (regresoras, factores de tratamiento, etc.) dado uno o más efectos aleatorios.

La componente lineal se define como en los modelos lineales tradicionales:

$$\eta_i = x_i' \beta$$

A diferencia de los modelos lineales, la relación entre el *valor esperado* de Y y el *predictor lineal* requiere de una *función de enlace* (monótona diferenciable) $g(\cdot)$, que describe cómo el valor esperado de Y (μ_i) está relacionado al predictor lineal:

$$g(\mu_i) = x_i' \beta$$

Cuando tenemos modelos mixtos, el predictor lineal se relaciona, mediante la función de enlace, con el valor esperado condicional. Es decir, ahora se considera la distribución condicional de la respuesta dado el (los) efecto(s) aleatorio(s). Para el caso de la distribución normal y enlace identidad (modelo lineal mixto), esto no causa ningún problema de interpretación en el valor esperado, ya que al pasar de la esperanza condicional a la esperanza marginal los parámetros asociados al predictor lineal son los mismos. En el caso de otros enlaces, la no linealidad de la función del enlace hace que la relación entre la esperanza marginal y los parámetros asociados con las regresoras no se pueda especificar analíticamente (salvo algunas excepciones).

Bajo distribución normal de la variable dependiente y enlace identidad, el modelo lineal mixto es:

$$\begin{aligned} \mathbf{y} | \mathbf{u} & \text{ i.d. } N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R}) \\ \mathbf{u} & \text{ i.d. } N(\mathbf{0}, \mathbf{G}) \end{aligned}$$

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{X}\boldsymbol{\beta} \\ E(\mathbf{y} | \mathbf{u}) &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \end{aligned}$$

La relación entre las varianzas condicional y marginal es tal que, aunque condicionalmente las observaciones sean independientes (matriz \mathbf{R} diagonal), marginalmente estarán correlacionadas debido al efecto aleatorio:

$$\begin{aligned} V(\mathbf{y}) = \mathbf{V} &= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \\ V(\mathbf{y} | \mathbf{u}) &= \mathbf{R} \end{aligned}$$

Para una función de enlace g y una distribución condicional f en la familia exponencial, el modelo lineal generalizado mixto es:

$$\begin{aligned} \mathbf{y} | \mathbf{u} & \text{ i.d. } f(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ \mathbf{u} & \sim N(\mathbf{0}, \mathbf{G}) \\ g(E(\mathbf{y} | \mathbf{u})) &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \end{aligned}$$

La esperanza y varianza marginales para la observación ij serán:

$$\begin{aligned} E(y_{ij}) &= E_{\mathbf{u}}(E(y_{ij} | \mathbf{u})) \\ \text{var}(y_{ij}) &= E_{\mathbf{u}}(\text{var}(y_{ij} | \mathbf{u})) + \text{var}_{\mathbf{u}}\{E(y_{ij} | \mathbf{u})\} \end{aligned}$$

Mediante una expresión análoga para la covarianza, podemos observar nuevamente que aunque condicionalmente las observaciones sean independientes, marginalmente estarán correlacionadas debido al efecto aleatorio.

Estimación por máxima verosimilitud completa

Conceptualmente pensamos que conocemos algo acerca de los coeficientes individuales a través de la variabilidad de los U . Si la variabilidad es pequeña, usamos el estimador promedio de la población, mientras que si es grande le daremos más peso a los datos propios.

La verosimilitud se puede escribir a partir de la distribución marginal de Y :

$$L(\beta, G; \underline{y}) = \prod_{i=1}^N \int \prod_{j=1}^{t_i} f(y_{ij} | U_i, \beta) f(U_i; G) dU_i$$

Para maximizar esta verosimilitud (o su logaritmo) es necesario realizar una aproximación a la integral en cada paso de optimización. Los dos métodos más comunes para esto son el de la *cuadratura adaptativa de Gauss-Hermite* y el de *Laplace*. El primero consiste en aproximar la integral con el área (o volumen) de un cierto número de rectángulos (o prismas) estratégicamente ubicados. Si se usa un solo punto de cuadratura el método es de Laplace. Este será computacionalmente más rápido por cada iteración pero requerirá más iteraciones.

Invocación del procedimiento de modelos lineales generalizados mixtos

En el menú *Estadísticas* seleccionar el submenú *Modelos lineales generalizados mixtos*, allí encontrará tres opciones. La primera, con el rótulo *Estimación*, invoca la ventana de diálogo que permite especificar la estructura del modelo. La segunda, con el rotulo *Exploración de modelos estimados* contiene herramientas de diagnóstico y comparación de modelo, la tercera opción, rotulada *Tutorial* es un enlace a este Manual. Para ilustrar la aplicación de estos modelos comenzaremos con un ejemplo binomial.

Distribución Binomial: Comparación de tratamientos en papaya

Comenzaremos indicando cómo ajustar un modelo de efectos fijos, utilizando el archivo [*PapayaBinomial.IDB2*](#) (Robles et al. 2007) del conjunto de datos de prueba de InfoStat. Los datos corresponden a un experimento en el que se comparan cuatro tratamientos, consistentes en el manejo de la cobertura del suelo para el control de áfidos, que son vectores del *ring spot* virus. Los tratamientos fueron:

- Suelo desmalezado (T) – control negativo
- Suelo con malezas (M) – control positivo
- Suelo cubierto con plástico negro (PN)
- Suelo cubierto con plástico plateado (PP)

Cada tratamiento fue aplicado a 5 parcelas en un diseño completamente aleatorizado. En cada parcela había 20 plantas de papaya y se evaluó el número de plantas enfermas cada dos semanas, totalizando 8 instancias de monitoreo. El número total de plantas observadas fue 400.

Para esta ejemplificación seleccionaremos solamente los datos del cuarto monitoreo (Días=85), donde ya se puede evaluar para este cultivo el efecto de los tratamientos. Para ello seleccionaremos la columna *Días* y con la opción *Buscar* en el menú *Datos* desactivaremos todos los casos con *Días*<>85 (ver manejo de datos en Manual del Usuario). A continuación activaremos el menú *Estadísticas*, submenú *Modelos lineales generalizados mixtos*, opción *Estimación*. En la ventana de selección de variables, los factores de clasificación, covariables y variables dependientes pueden ser especificados como en un análisis de la varianza para efectos fijos. Para los datos en el archivo *PapayaBinomial.IDB2* especificar *Enfermas* como variable respuesta y como criterio de clasificación: *Trat*. Incluiremos además a la variable *Plantas* (que contienen una constante=20) como una covariable cuya aplicación explicaremos más adelante (Figura 1).

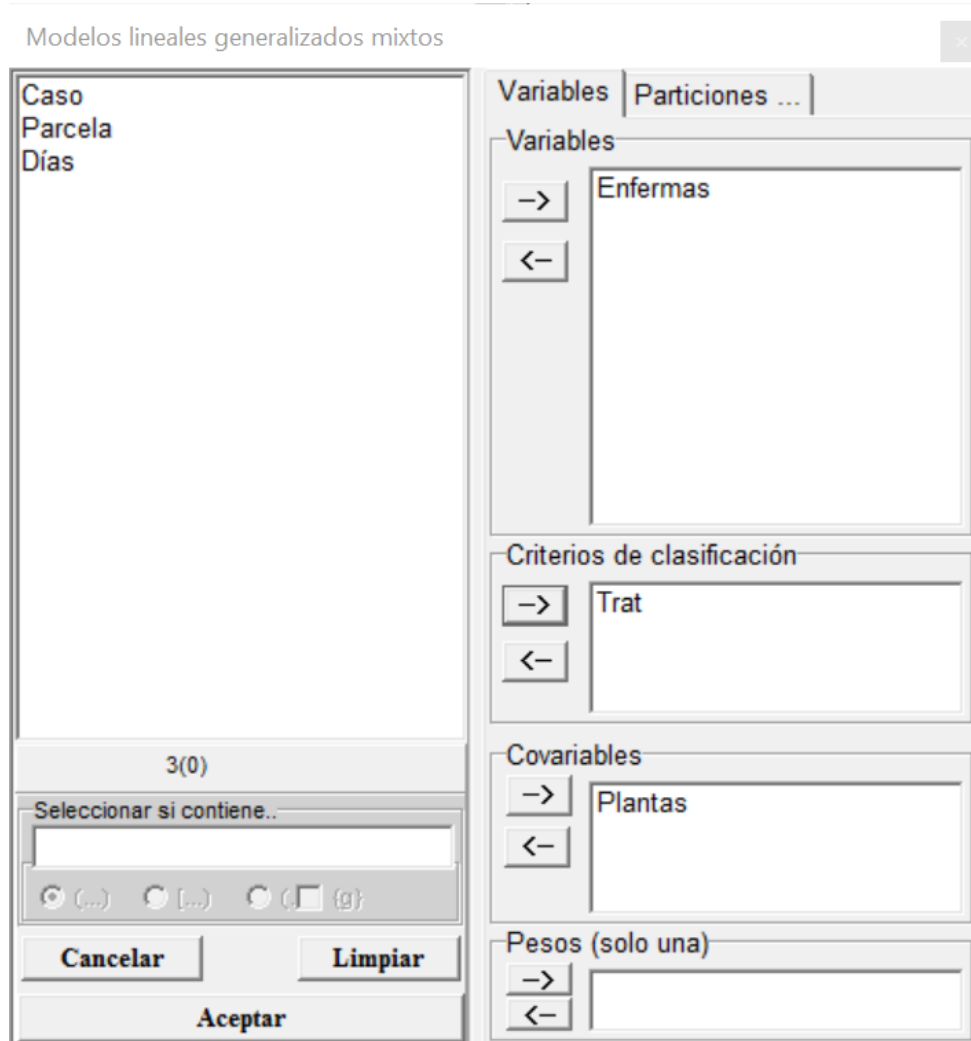


Figura 1: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo PapayaBinomial.IDB2.

Una vez que se acepta la selección realizada se mostrará la ventana principal de la interfaz para modelos lineales generalizados mixtos. Esta ventana contiene tres solapas (Figura 2).

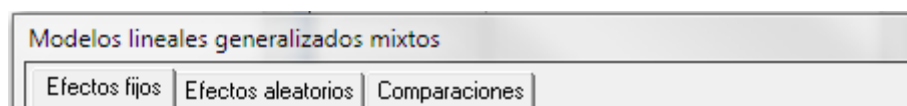


Figura 2: Solapas con las opciones para especificación de un modelo lineal generalizado mixto.

La primera permite especificar los efectos fijos del modelo y seleccionar opciones para la presentación de resultados y la generación de predicciones y residuos (deviance) del modelo. A la derecha de la ventana aparecerá una lista conteniendo las variables de clasificación y las covariables declaradas en la ventana de selección de variables. Para incluir un factor (variable de clasificación) o una covariable

a la parte fija del modelo, basta hacer doble clic sobre el nombre del factor o covariable que se quiere incluir. Esta acción agregará una línea en la lista de efectos fijos. Doble clics adicionales sobre un factor o una covariable agregarán términos en líneas sucesivas, implícitamente separados por un signo “+” (modelo aditivo). Seleccionando con el ratón los factores principales y accionando el botón “*” se introduce un término que especifica la interacción entre los factores.

Para el conjunto de datos en el archivo [PapayaBinomial.IDB2](#), incluiremos en el modelo de efectos fijos el factor *Trat* (Figura 4). Debido a que el método de estimación utilizado es máxima verosimilitud y que en los modelos lineales generalizados y generalizados mixtos los datos pueden provenir de distintas distribuciones, el usuario debe especificar la distribución de las observaciones a través de lo que para R es el parámetro familia (**family**) las siguientes opciones:

Binary
Binomial
Poisson
Binomial Negativa
Normal
Gamma
Inverse gaussian
Quasi Binomial
Quasi Poisson

Como indicamos, la variable de respuesta es el número de plantas enfermas en cada parcela. Cada parcela tiene 20 plantas, por lo que el número de plantas enfermas puede variar entre 0 y 20. Si asumimos que la infección ocurre independientemente de planta a planta y que la probabilidad de infección se mantiene constante durante todo el experimento, entonces la variable de respuesta sigue una distribución binomial con parámetros: π desconocido y $n=20$. Luego para completar la especificación del modelo que queremos estimar elegiremos *Binomial* como *familia* y al hacerlo automáticamente aparecerá seleccionada la función de enlace más comúnmente usada para esta distribución (logit). Luego seleccionaremos la variable *Plantas* en el campo desplegable a la derecha de la función de enlace ()

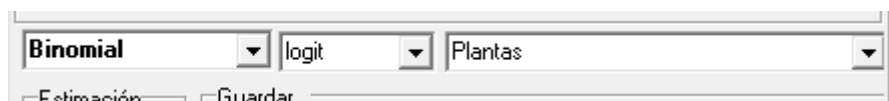


Figura 3: Campos para la selección de la distribución, la función de enlace y le covariable que indica el número de experimentos para el análisis de los datos del archivo *PapayaBinomial.IDB2*.

Si el número de ensayos Bernoulli independientes es el mismo para todas las unidades experimentales, se puede digitar este número directamente en el campo en vez de identificar la variable que lo contiene.

Como se explicó en la introducción a los modelos lineales generalizados, lo que se estima en estos modelos son los parámetros de un predictor lineal para una función $g(\cdot)$ de la esperanza de la variable de respuesta. Esta función $g(\cdot)$, conocida como función de enlace también debe especificarse. InfoStat elige una función de enlace por defecto dependiendo la familia seleccionada. La función de enlace seleccionada en este ejemplo es la *logit* y la covariable indicando el número de ensayos es *Planta* (Figura 4).

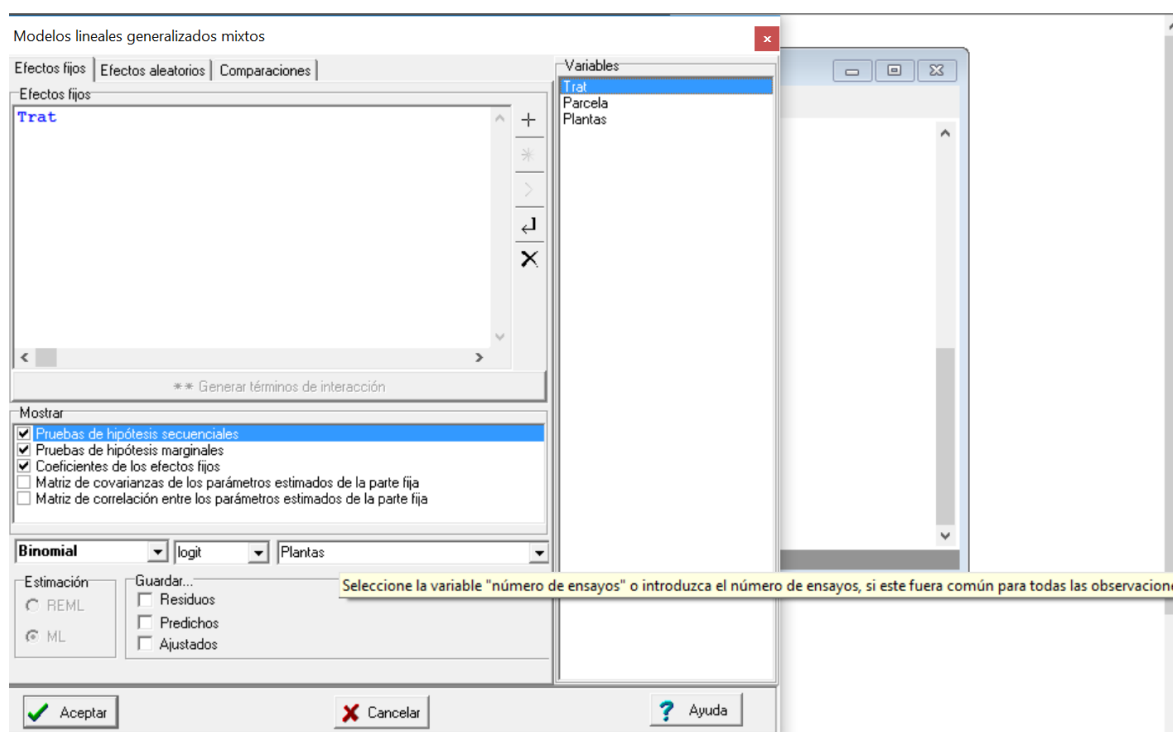


Figura 4: Ventana con la solapa *Efectos fijos* desplegada para los datos del archivo *PapayaBinomial.IDB2* con el menú contextual indicando donde colocar la variable número de ensayos.

Si aceptamos esta especificación, en la ventana de resultados de InfoStat se obtendrá la salida que se muestra a continuación. La primera parte contiene la especificación de la forma en que se invocó la estimación del modelo en la sintaxis de R, e indica el nombre del objeto R que contiene al modelo y su estimación. En este caso `modelo.g01_Enfermas_ML`. Esta especificación es sólo de interés para aquellos que están acostumbrados a ver las sentencias en R.

A continuación se indica la variable de respuesta (*Enfermas*) y un cuadro que indica la familia, el enlace y si el modelo convergió. El parámetro escala, se discutirá más adelante, pero si la familia fuera la normal y la función de enlace identidad, este parámetro corresponde al cuadrado medio residual del análisis de la varianza.

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.001_Enfermas_ML<-glm(cbind(Enfermas
,as.numeric(as.character(Plantas))-Enfermas)~1+Trat
,family=myFamily
,na.action=na.omit
,data=R.data00)
```

Resultados para el modelo: *mlgm.modelo.001_Enfermas_ML*

Variable dependiente: *Enfermas*

General

Familia	Enlace	Convergencia	Escala
binomial	logit	Alcanzada	1.00

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
160	2642.71	2655.01	-1317.35	2332.98

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			159	2425.80	
Trat	3	92.81	156	2332.98	<0.0001

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.05	0.07	0.71	0.4795
TratPN	-0.52	0.10	-5.16	<0.0001
TratPP	-0.66	0.10	-6.48	<0.0001
TratT	0.16	0.10	1.60	0.1089

La tabla *Medidas de ajuste del modelo* contiene estadísticos útiles para comparar el ajuste de distintos modelos. *AIC* hace referencia al criterio de Akaike, *BIC* al Criterio Bayesiano de Información, *logLik* al logaritmo de la verosimilitud.

Para modelos lineales generalizados como la Poisson, Binomial, Bernoulli (Binary) y Binomial negativa, el cociente entre la *deviance relativa* (en adelante *deviance*) y los grados de libertad residuales debería estar cerca de 1 si el ajuste es razonable. Un valor grande de este cociente significa que la varianza es mayor que la esperada bajo la distribución correspondiente (sobredispersión). Se suele usar un valor de hasta 2.5 para el cociente como indicador de que no hay una sobredispersión importante. Si la familia fuera la `normal` y la función de enlace `identidad`, la *deviance* corresponde a la suma de cuadrado residual del análisis de la varianza y el cociente es el cuadrado medio residual.

La tercera tabla de esta salida presenta la tabla de pruebas de hipótesis secuenciales para los efectos fijos. Esta tabla es equivalente a una tabla de análisis de la varianza de un modelo lineal clásico. En la línea NULL, que identifica a un modelo que contiene sólo la constante, la *deviance* es 79.60, mientras que la *deviance* para el modelo que incluye a la constante y el efecto de los tratamientos, es 31.15. La reducción en la *deviance* fue de 48.44 unidades, correspondiente al efecto de los tratamientos. El p-valor (<0.0001) surge de comparar la reducción de la *deviance* con una chi-cuadrado con 3 grados de libertad. De acuerdo a estos resultados, existe un efecto significativo de tratamientos. Para establecer patrones de diferencias podemos aplicar una prueba de comparaciones múltiples. InfoStat implementa dos: Una es equivalente a una prueba LSD de Fisher, la otra es la implementación, para estos modelos de la prueba DGC (Di Rienzo, Guzmán, y Casanoves 2002). Estas pruebas están disponibles en la solapa *Comparaciones* – sub-solapa *Medias*. En esta ventana tildaremos el factor o factores de clasificación cuyas medias queremos comparar: en este caso `Trat`, y si seleccionamos la prueba DGC (Figura 5) se obtiene la salida correspondiente a las comparaciones de medias que se encuentra a continuación de la Figura 5.

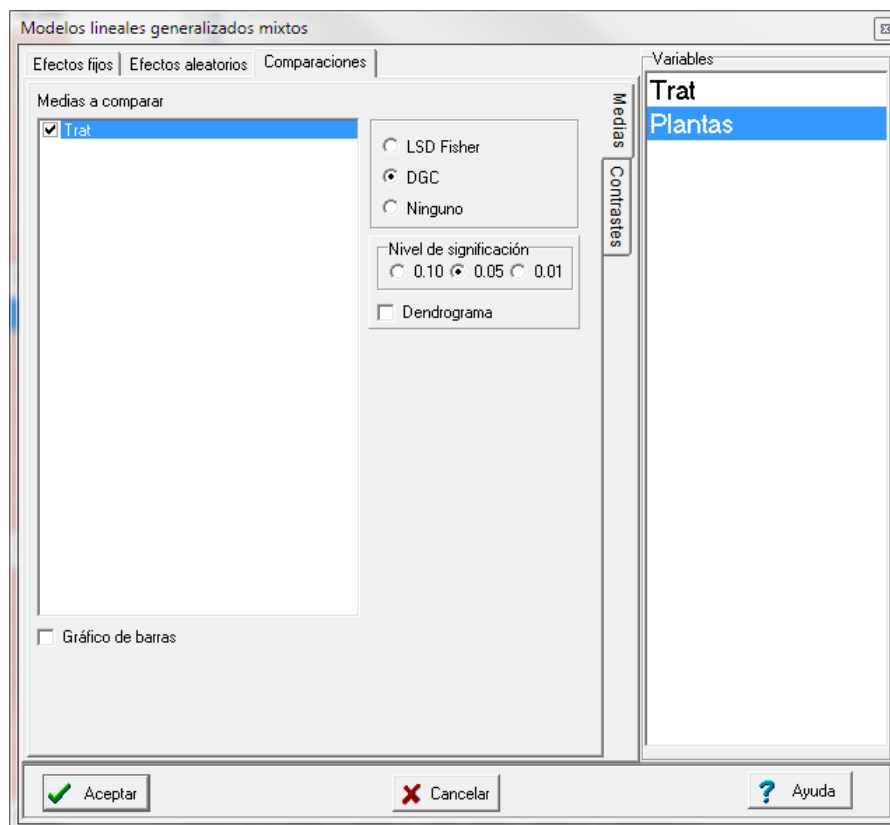


Figura 5: Ventana desplegada con la solapa Comparaciones para los datos del archivo PapayaBinomial.IDB2.

Enfermas - Medias ajustadas y errores estándares para Trat

Inversa de la función de enlace con efecto aleatorio=0

DGC (Alfa=0.05)

Trat	PredLin	E.E.	Media	E.E.	
T	-0.36	0.20	0.41	0.05	A
M	-0.71	0.21	0.33	0.05	A
PN	-2.20	0.33	0.10	0.03	B
PP	-2.44	0.37	0.08	0.03	B

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

En el cuadro anterior hay una columna denominada `PredLin`, que corresponde a los valores predichos para los tratamientos en la escala de la función de enlace. La columna `E.E.` a la derecha de la primera contiene sus errores estándares. Las pruebas que se presentan a través del código de letras a la derecha de la tabla se realizan sobre `PredLin`, sin embargo, la escala de la función de enlace suele no ser útil a los fines de la interpretación. Por ello, se incluyen las columnas `Media` y `E.E.` que están en la escala original. En este caso la escala original es una escala de proporciones (Enfermas/Plantas). Así, la proporción de plantas enfermas en la cuarta fecha de monitoreo fue para M del 33% y para T del 41%, no siendo la diferencia entre ellas estadísticamente significativa

(ambas comparten la letra A). Los tratamientos PN y PP, tiene respectivamente proporciones de plantas enfermas del 10% y 8% respectivamente, no siendo la diferencia, estadísticamente significativa (comparten la letra B). Por otra parte los controles M y T, difieren de los tratado PN y PP de manera significativa (no comparten letras). Sería interesante estimar la diferencia entre controles (M y T) y tratados (PN y PP) mediante un contraste. Para ello, utilizando la solapa Comparaciones, sub-solapa Contrastes. Como existe sólo un factor fijo (Trat) éste aparece preseleccionado y sus niveles (M, PN, PP y T) listado. Si seleccionamos con el ratón los tratamientos PN y PP y accionamos el botón [+], los nombres de estos tratamientos aparecerán en **rojo**. Si hacemos lo propio con los tratamientos M y T y utilizamos el botón [-], estos se marcarán en **azul**. Apretando el botón marcado con una flecha en la Figura 6, se obtienen los coeficientes que aparece al pie de la figura ($1/2 \ -1/2 \ -1/2 \ 1/2$). Estos coeficientes muestran que lo que compararemos son los promedios de los predictores lineales de PN y PP vs el promedio de los predictores de M y T. A continuación de la Figura 6 se presenta la parte correspondiente a la estimación del contraste que aparece en la salida de InfoStat.

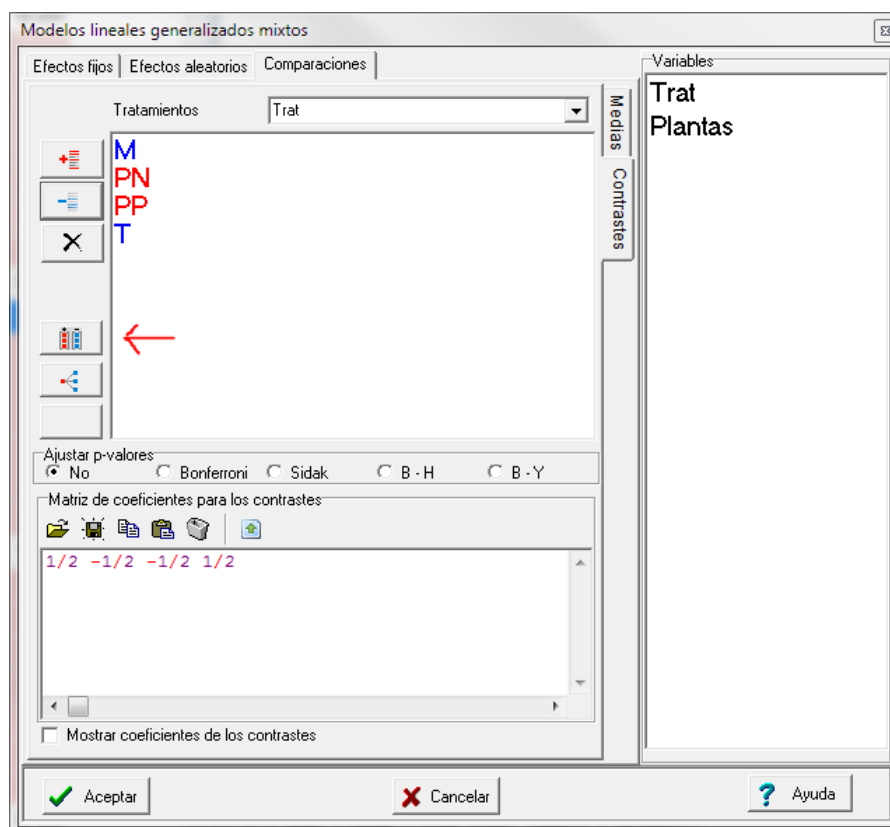


Figura 6: Especificación de contrastes para comparar tratamientos. Archivo PapayaBinolial.IDB2.

Pruebas de hipótesis para contrastes

Trat	PredLin	E.E.	Chi-Cuadrado	gl	p-valor
Ct.1	-1.78	0.29	38.15	1	<0.0001
Total			38.15	1	<0.0001

La diferencia entre los promedios de los predictores lineales estimada es -1.78, siendo esta diferencia significativa ($p < 0.0001$). Este resultado no hace más que confirmar lo que ya se observaba en la tabla de comparaciones múltiples y la construcción del contraste no sería necesaria excepto que nuestro interés fuera estimar la diferencia. Como antes, la diferencia de los predictores lineales está en la escala de la función de enlace. En algunos casos, es posible interpretar el contraste. En particular, para el caso binomial con función de enlace `logit`, usando la transformación $\exp(\text{contraste})$ se puede obtener una razón de chances (odds-ratio). Para nuestro ejemplo $\exp(-1.78) = 0.16$, indica que la chance de que una planta tratada se infecte es 0.16 veces la chance de que se infecte un no tratado. Esta forma de expresar puede ser difícil de interpretar, por lo que esta razón de chance se puede invertir ($1/0.16$) para expresar la chance de infección de una planta no tratada versus una tratada, en este caso da 6.25.

Un intervalo de confianza para la razón de chances puede obtenerse a partir de la exponenciación de los límites del intervalo de confianza para el predictor lineal. Así, utilizando una aproximación normal, los límites de un intervalo de confianza al 95% para el predictor son $-1.78 \pm 1.96 \times 0.29$ de donde los límites para el odds-ratio se calculan como:

$$[\exp(-1.78 - 1.96 \times 0.29), \exp(-1.78 + 1.96 \times 0.29)] = [0.29, 0.09]$$

Reordenando el intervalo nos queda $[0.09, 0.29]$. Si se quiere expresar con relación al no tratado, estos límites serán $[3.36, 10.47]$, es decir la chance de tener plantas infectadas en los no tratados esta, con un 95% de confianza, entre 3.36 y 10.47.

Distribución Binomial: Comparación de curvas de progreso de una enfermedad

En el ejemplo anterior nos limitamos al análisis de los datos del cuarto monitoreo (85 días). Para cada parcela se disponía entonces de un único dato. Si por el contrario quisiéramos estudiar el progreso de enfermedad en función del tiempo (Días), entonces tendríamos que reformular el modelo. Por una

parte debemos agregar *Días* como una covariable y su interacción con *Trat*, para incluir regresiones específicas para cada tratamiento. Esto sería suficiente si no fuera por el hecho de que las observaciones en el tiempo sobre una misma parcela se encuentran correlacionadas. Una manera apropiada de tener en cuenta la falta de independencia entre observaciones de una misma parcela, es incorporando el efecto parcela como aleatorio (Agresti 2002, página 496). La Figura 8, muestra la especificación de la parte fija (izquierda) y de la parte aleatoria (derecha). La inclusión de un efecto aleatorio, es por defecto, sobre la constante del modelo (opción por defecto). Esto implica que estamos suponiendo que la tasa de progreso de la enfermedad no depende de la parcela lo que, para este ejemplo, es un supuesto razonable.

Para evaluar la presencia de sobredispersión se usará la deviance condicional, que aparece directamente en la salida de InfoStat si se han declarado efectos aleatorios. Hay que enfatizar que esta deviance es la correcta en estos casos ya que la distribución de la variable dependiente especificada en el modelo lineal generalizado es condicional al efecto aleatorio, por lo que la deviance para evaluar una posible sobredispersión debe ser la condicional. Si al valor de la deviance de 143.16 lo dividimos por los grados de libertad del denominador (152) nos da un cociente de 0.94, lo que indica ausencia de sobredispersión.

A continuación de la Figura 8 se presenta la tabla de pruebas de hipótesis secuenciales, en la que se observa una interacción *Trat*Días* significativa, indicando que la tasa de progreso de la enfermedad depende de los tratamientos. Por otra parte encontramos que la varianza del efecto aleatorio parcela es 0.41.

Lo que interesaría ahora establecer es si cada tratamiento tiene una curva diferente o, si como ocurrió en la ejemplificación anterior, sólo se observan diferencias significativas en el progreso de la enfermedad entre controles, por una parte, y tratados por la otra. Recalcularemos el modelo declarando ahora a *Días* como covariable (Figura 7) y luego pediremos que se muestren los coeficientes estimados de la parte fija; esta es una opción de la solapa *Efectos fijos* (Figura 8). En la solapa de Efectos aleatorios se declara a Parcela (Figura 9).

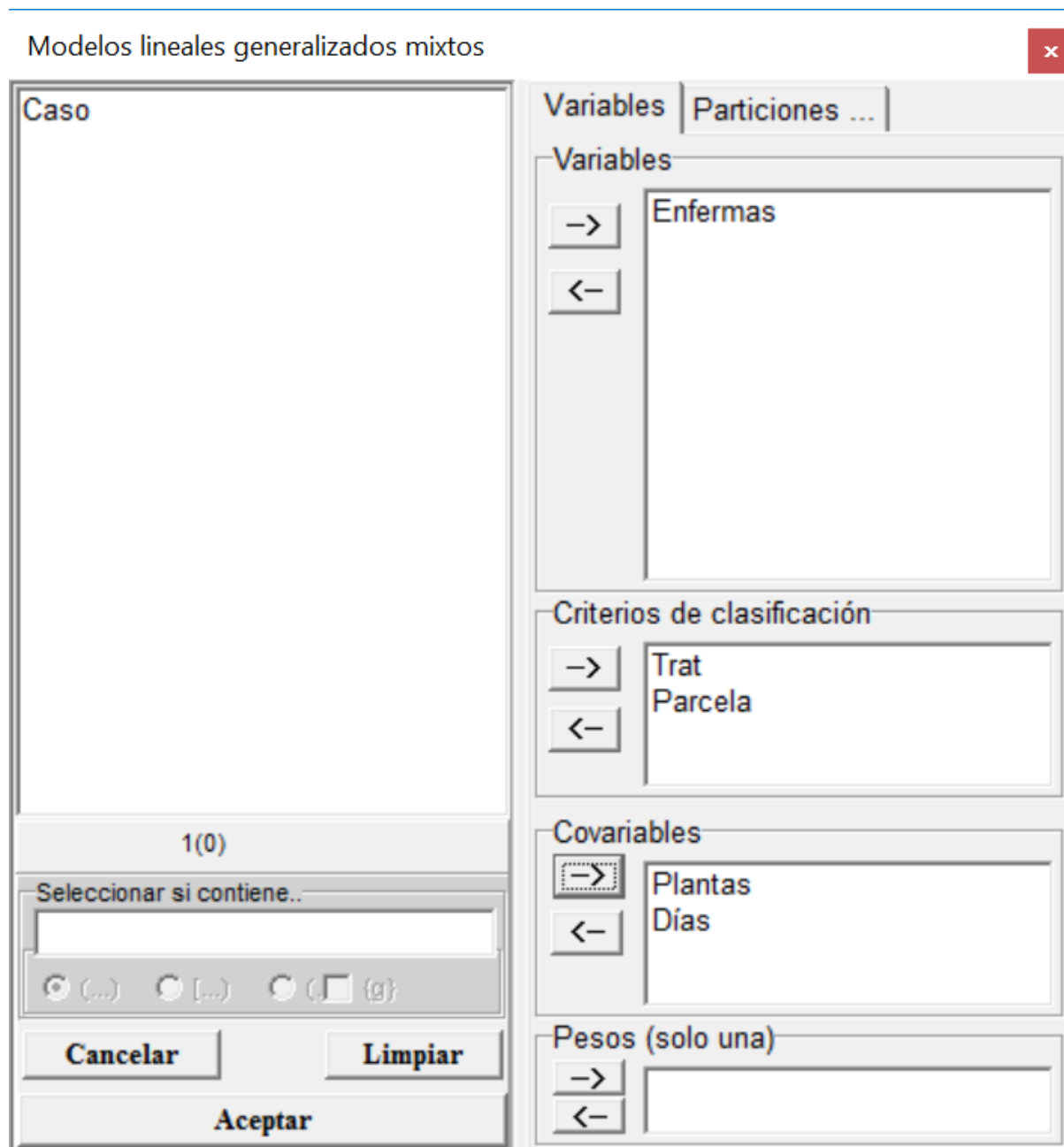


Figura 7: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables y covariables para el análisis de los datos del archivo PapayaBinomial.IDB2.

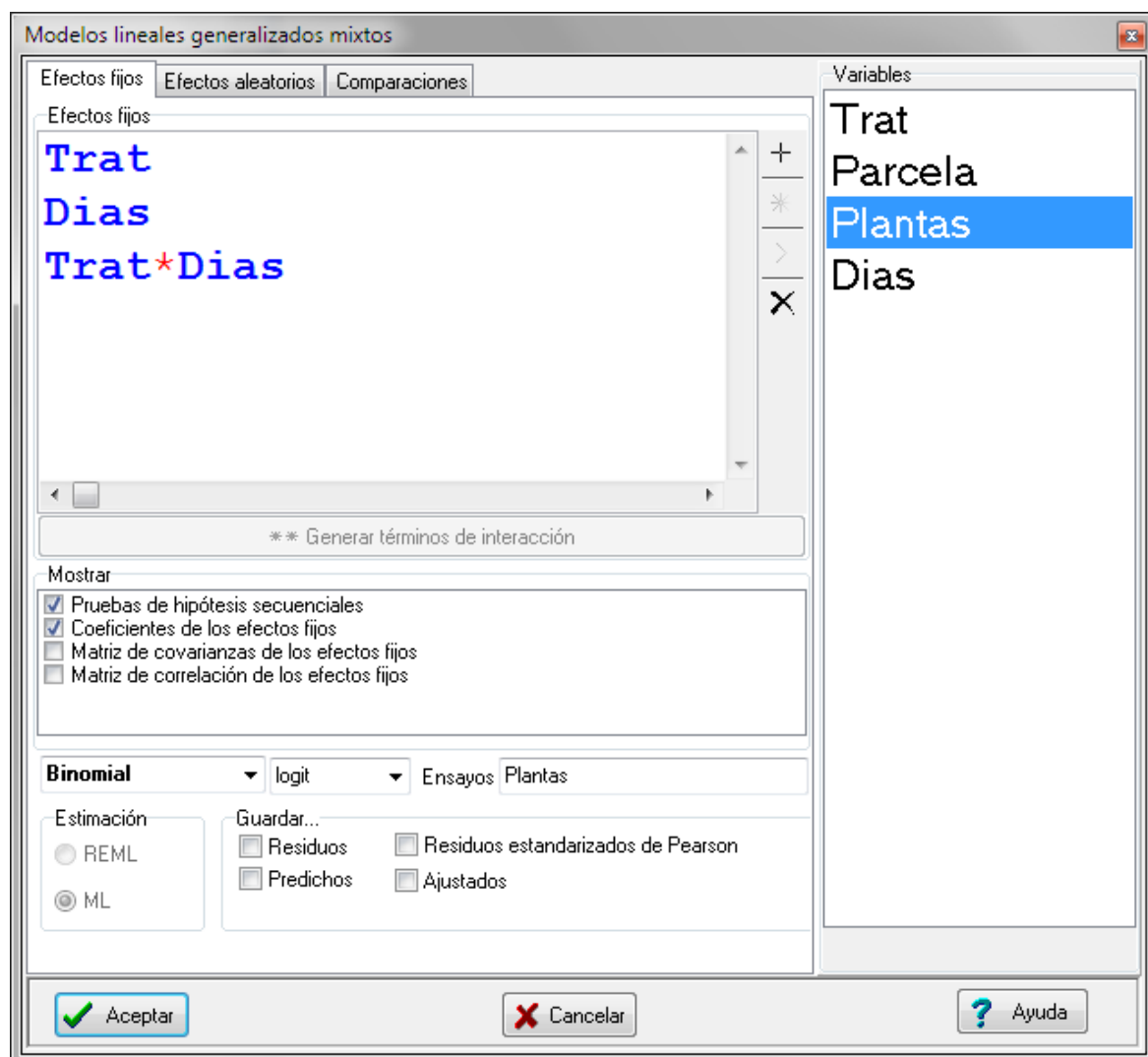


Figura 8: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de efectos de Trat, Dias y su interacción, usando una familia Binomial y enlace logit para el análisis de los datos del archivo PapayaBinomial.IDB2.

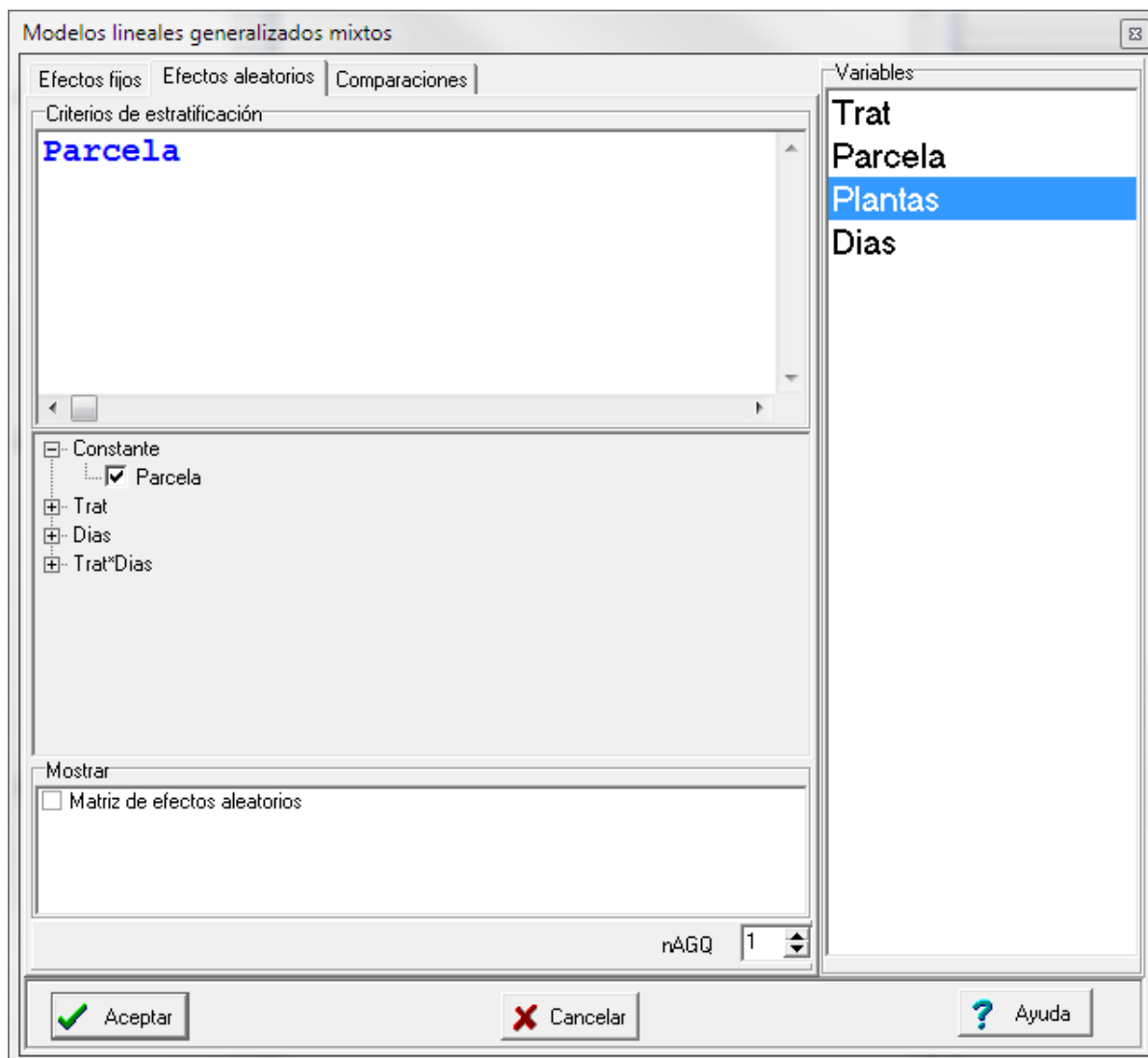


Figura 9: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y declaración de efectos de Parcela y enlace logit para el análisis de los datos del archivo PapayaBinomial.IDB2..

A continuación, se presenta la salida obtenida con las especificaciones de las figuras anteriores:

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.002_Enfermas_ML<-glmer(cbind(Enfermas
,as.numeric(as.character(Plantas)) -
Enfermas)~1+Trat+Dias+Trat:Dias+(1|Parcela)
,family=myFamily
```

```
,na.action=na.omit
,REML=F
,AGQ=1
,data=R.data02)
```

Resultados para el modelo: mlgm.modelo.002_Enfermas_ML

Variable dependiente: Enfermas

General

Familia	Enlace	nAGQ
binomial	logit	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
160	521.134	548.811	-251.567	143.157

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
Trat	18.809	3	0.0003
Dias	2066.950	1	<0.0001
Trat:Dias	47.067	3	<0.0001

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.982	0.497	-12.038	<0.0001
TratPN	-5.208	1.000	-5.209	<0.0001
TratPP	-5.732	1.034	-5.542	<0.0001
TratT	0.226	0.700	0.322	0.7472
Dias	0.066	0.004	15.222	<0.0001
TratPN:Dias	0.041	0.009	4.546	<0.0001
TratPP:Dias	0.042	0.009	4.625	<0.0001
TratT:Dias	0.001	0.006	0.166	0.8681

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Parcela	(Intercept)	0.409	0.639

Se podrá observar que TratM y TratM:Días no aparece en la lista de efectos fijos. Esto se debe a la forma en que R re-parametriza los modelos lineales para poder estimarlos. En la práctica esto

implica que TratM y TratM:Días han sido fijados en cero. Los parámetros estimados deben interpretarse a la luz de esta parametrización.

De acuerdo a la tabla anterior, las rectas estimadas para los predictores lineales son las siguientes:

Tratamiento	Recta estimada para el predictor lineal		
M	(-5.983) +	(0.066) × Días + Parcela
T	(-5.983 + 0.226)	+ (0.066 + 0.001)	× Días + Parcela
PN	(-5.983 - 5.207)	+ (0.066 + 0.041)	× Días + Parcela
PP	(-5.983 - 5.731)	+ (0.066 + 0.042)	× Días + Parcela

Es fácil observar en la tabla anterior que el parámetro estimado como TratT (0.226) es la diferencia de constantes entre el tratamiento T respecto y el tratamiento M y que esta diferencia es no significativa ($p=0.3799$). Igualmente el parámetro TratT:Días (0.001), corresponde a la diferencia en la pendiente del tratamiento T respecto del M y que esta diferencia es no significativa. Con igual lógica observamos que las constantes de los tratamientos PN y PP, difieren de la del tratamiento M y que lo mismo ocurre con las pendientes. No estamos aquí estableciendo si hay diferencias de constantes y/o pendiente entre PN y PP. Esto puede hacerse estimando algunas combinaciones lineales específicas para los parámetros de interés y probando si son distintas de cero. Recordemos que los parámetros estimados son los siguientes:

```
(Intercept)
TratPN
TratPP
TratT
Días
TratPN:Días
TratPP:Días
TratT:Días
```

La combinación lineal que permite estimar la diferencia de constantes entre PN y PP es:

$$L0 = 0 \times (\text{Intercept}) + 1 \times \text{TratPN} - 1 \times \text{TratPP} + 0 \times \text{Días} + 0 \times \text{TratPN:Días} + 0 \times \text{TratPP:Días} + 0 \times \text{TratT:Días}$$

y la combinación lineal que permite estimar la diferencia de pendientes entre PN y PP es:

$$L1 = 0 \times (\text{Intercept}) + 0 \times \text{TratPN} + 0 \times \text{TratPP} + 0 \times \text{Días} + 1 \times \text{TratPN:Días} - 1 \times \text{TratPP:Días} + 0 \times \text{TratT:Días}$$

Las hipótesis que se prueban en ambos casos es si las diferencias estimadas difieren estadísticamente de cero.

Para realizar estas estimaciones y sus contrastes de hipótesis invocaremos el módulo de *Exploración de modelos estimados*. Este módulo está disponible sólo si existen modelos estimados previamente.

Cuando se invoca este menú aparece la ventana que se muestra en la Figura 10. Se puede observar la lista de los parámetros estimados en el panel superior. Los coeficientes de las combinaciones L0 y L1 se escriben en el panel inferior, tal como aparecen en la Figura 10. Estos coeficientes pueden generarse automáticamente utilizando las mismas técnicas explicadas con anterioridad (Figura 6).

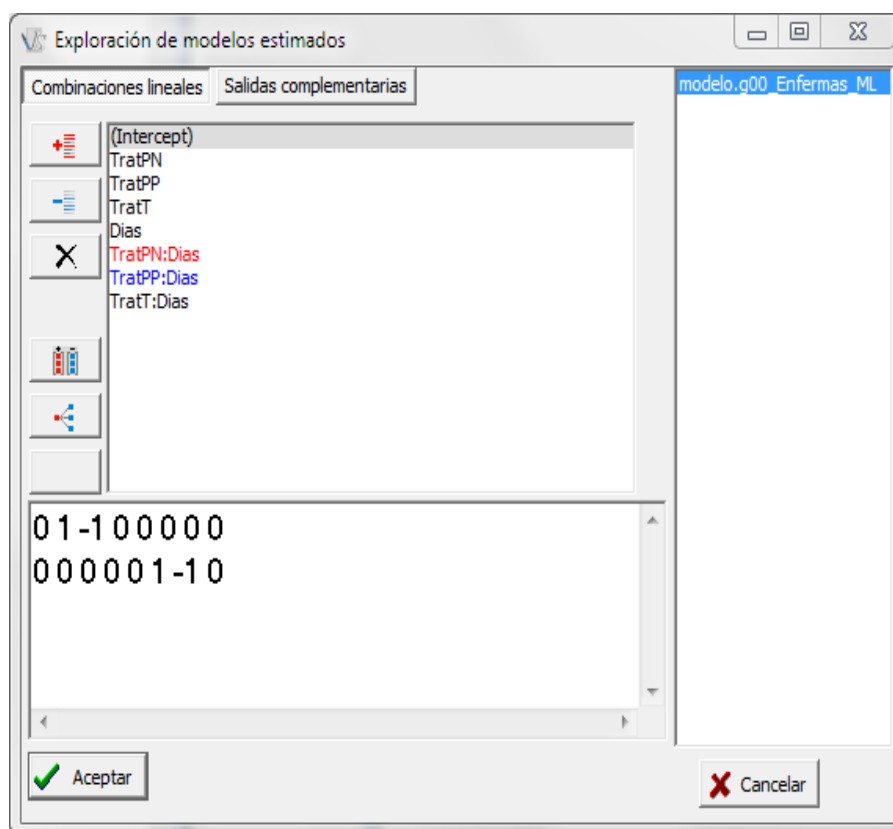


Figura 10: Ventana de exploración de modelo: solapa [Combinaciones lineales]

El resultado (parcial) de aceptar estas especificaciones se muestra en la siguiente salida:

Pruebas de hipótesis para combinaciones lineales

Comb. lineal	Estimación	E.E.	gl	Chi-Cuadrado	p-valor
Comb.1	0.5236	1.2488	1	0.1758	0.6750
Comb.2	-0.0013	0.0111	1	0.0141	0.9054
Total			2	0.7765	0.6782

Coefficientes de las combinaciones lineales

Parámetros	Comb.1	Comb.2
(Intercept)	0.0000	0.0000
TratPN	1.0000	0.0000
TratPP	-1.0000	0.0000
TratT	0.0000	0.0000

Dias	0.0000	0.0000
TratPN:Dias	0.0000	1.0000
TratPP:Dias	0.0000	-1.0000
TratT:Dias	0.0000	0.0000

Se puede verificar que tanto la combinación 1 como la 2, no son significativamente distintas de cero. Más aún, la prueba conjunta, que contrasta simultáneamente las dos combinaciones, resultó no significativa ($p=0.6782$). Estas pruebas confirman la presunción que los modelos de regresión para PN y PP son indistinguibles. Como ya habíamos establecido que la curva de progreso de la enfermedad para M y T tampoco eran distinguibles, podemos reducir el modelo uniendo M y T bajo el rótulo común *Control* y PN y PP bajo el rótulo común *Tratado*. Esto puede lograrse utilizando el editor de categorías (ver manejo de datos en el Manual del Usuario de InfoStat).

Una vez que tenemos los nuevos datos, donde se encuentran unidos los dos controles en una sola categoría y los tratados en otra categoría, procedemos a ajustar un modelo similar al mostrado en la (Figura 8 y Figura 9) (la formulación del modelo es idéntica, lo que cambian son las variables que clasifican a los datos). El resultado del ajuste de este modelo se muestra en el siguiente cuadro.

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.003_Enfermas_ML<-glmer(cbind(Enfermas
,as.numeric(as.character(Plantas)))-
Enfermas)~1+TratNuevo+Dias+TratNuevo:Dias+(1|Parcela)
,family=myFamily
,na.action=na.omit
,REML=F
,AGQ=1
,data=R.data03)
```

Resultados para el modelo: mlgm.modelo.003_Enfermas_ML

Variable dependiente: Enfermas

General

Familia	Enlace	nAGQ
binomial	logit	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
160	514.45	529.82	-252.22	142.95

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
TratNuevo	17.42	1	<0.0001
Días	2066.92	1	<0.0001
TratNuevo:Días	47.17	1	<0.0001

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.87	0.35	-16.55	<0.0001
TratNuevoTratado	-5.59	0.72	-7.72	<0.0001
Días	0.07	3.1E-03	21.39	<0.0001
TratNuevoTratado:Días	0.04	0.01	6.39	<0.0001

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Parcela	(Intercept)	0.44	0.66

Si al valor de la deviance de 142.95 lo dividimos por los grados de libertad del denominador (156) nos da un cociente de 0.92, lo que indica ausencia de sobredispersión. De acuerdo a la estimación de los parámetros los predictores lineales para controles y tratados son los siguientes:

Predictor lineal Control = $-5.868 + 0.067 \times \text{Días} + \text{parcela}$

Predictor lineal Tratado = $(-5.668 - 5.585) + (0.067 + 0.041) \times \text{Días} + \text{parcela}$

Los predictores lineales están en la escala definida por la función de enlace y describen rectas que difieren en la constante y en la pendiente según se trate del caso Control o Tratado. Asimismo, tanto para el Control como para el Tratado, hay tantas rectas paralelas como parcelas tenga cada condición, debido al efecto parcela, que en este caso afecta a la ordenada al origen. Estas rectas no son, sin embargo, las representaciones gráficas más apropiadas.

Para obtener curvas de progreso de la enfermedad en la escala original (proporciones para un problema binomial) debemos aplicar a los predictores lineales la función inversa de la función enlace.

Si $g(\cdot)$ representa la función de enlace logit, entonces $g^{-1}(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)}$ donde η representa al predictor lineal. Luego, las curvas de progreso de la enfermedad para *Control* y *Tratado* en la escala original para una parcela típica cuyo efecto aleatorio sea 0 serán:

$$\text{Progreso enfermedad}_{\text{Control}}(\text{Días}) = \frac{\exp(-5.868 + 0.067 \times \text{Días})}{1 + \exp(-5.868 + 0.067 \times \text{Días})}$$

$$\text{Progreso enfermedad}_{\text{Tratado}}(\text{Días}) = \frac{\exp((-5.868 - 5.585) + (0.067 + 0.041) \times \text{Días})}{1 + \exp((-5.868 - 5.585) + (0.067 + 0.041) \times \text{Días})}$$

Estas expresiones pueden usarse directamente en el graficador de funciones de InfoStat o, posiblemente más simple, sea graficar en función de *Días* y particionado por *Trat* (donde están unificadas las categorías) los valores ajustados y unir los puntos obtenidos con conectores (Figura 11). Para poder realizar este gráfico necesitamos haber guardado los valores [Ajustados] por el modelo, ésta es una opción que se pide en el momento de realizar estimación del modelo, en la solapa [Efectos fijos]. El grafico resultante (Figura 11) ha sido modificado, se han cambiado las escalas por defecto y se han agregado leyendas.

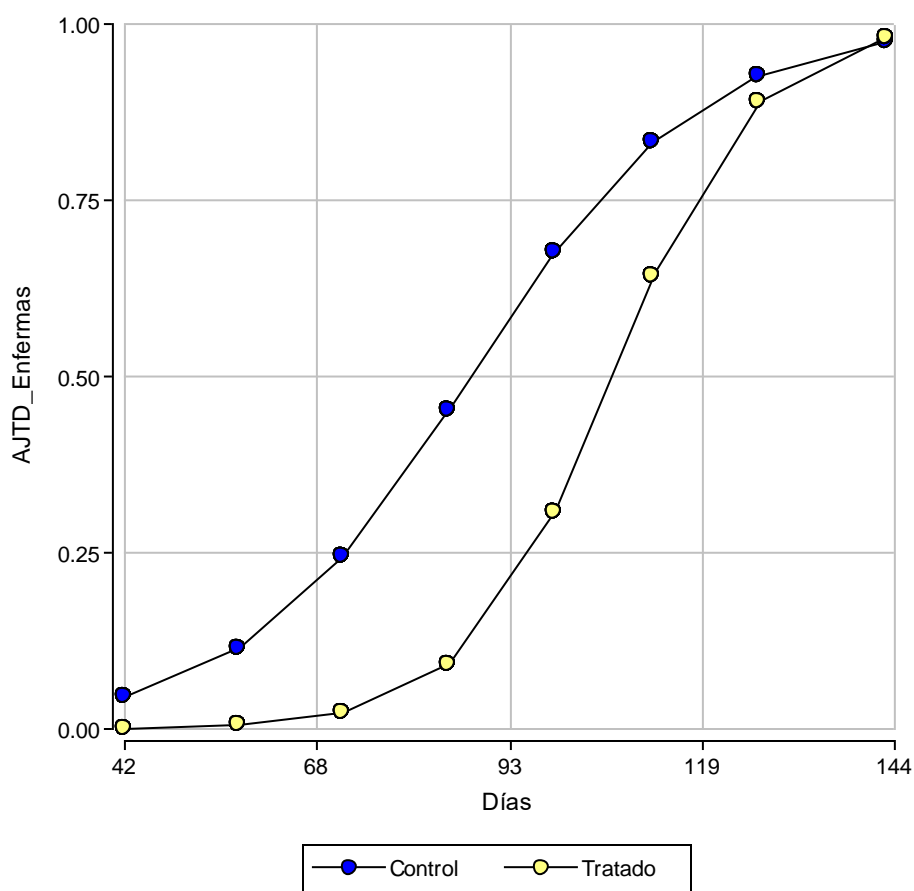


Figura 11: Curvas de evolución de enfermedad para Control ($M + T$) y Tratado ($PN + PP$)

Una pregunta frecuente en el análisis del progreso de una enfermedad es ¿cuándo se alcanza el 50% de individuos enfermos? Para modelos con efectos aleatorios, esta pregunta la debemos interpretar como “en un parcela típica (cuyo efecto aleatorio sea 0), ¿cuándo se alcanza el 50 de individuos enfermos?” Gráficamente se puede ver que para los controles este punto se alcanza alrededor del día 90 y para los tratados alrededor de día 105. Analíticamente, estos puntos corresponde al valor de $Días = -\beta_0 / \beta_1$. Para los controles $5.618/0.064 \cong 88$ días, mientras que para los tratados ese valor es $-(-5.618-4.817)/(0.064+0.034) \cong 106$ días.

Regresión logística simple: Prediciendo éxito en inseminación artificial

En este ejemplo se desea estudiar, como el del diámetro del cuerpo lúteo (DCL), afecta la probabilidad de que una vaca quede preñada o no después de una inseminación. Se tomó una muestra de 500 vacas inseminadas a las que se les midió el diámetro del cuerpo lúteo al momento de la inseminación. Posteriormente se registró la condición de la vaca (1=preñada, 0=no preñada). Los datos están en el archivo Preñez.IDB2. Se seleccionaron las variables Condición como respuesta y DCL como regresora (Figura 12).



Figura 12: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo Preñez.IDB2

Luego se declara en la solapa de efectos fijos a la covariable DCL y se pide un análisis con un modelo Bernoulli (Binary) y función de enlace logit (Figura 13).

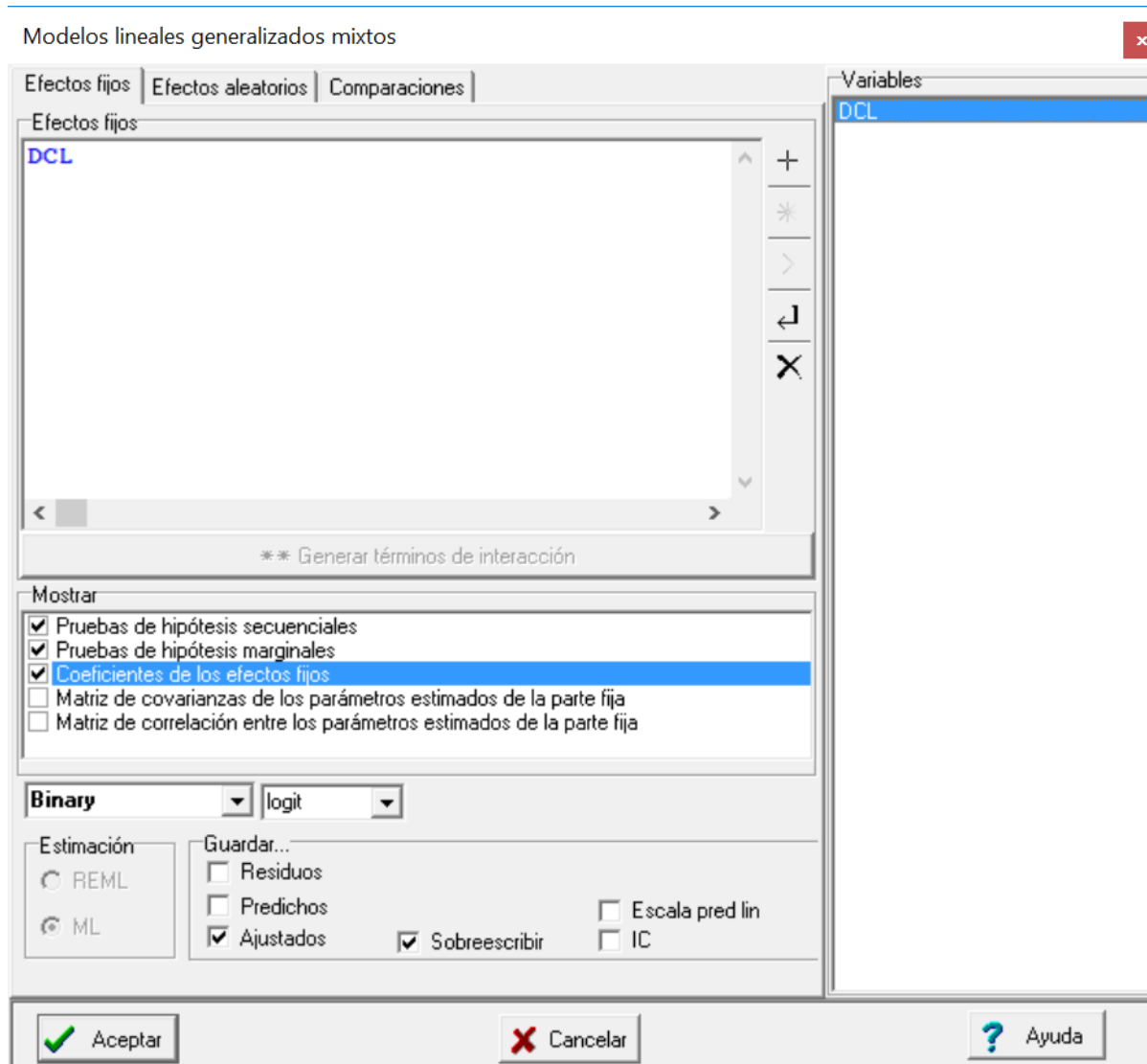


Figura 13: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variable regresora DCL (Diámetro del cuerpo lúteo) usando una distribución Bernoulli (Binary) y función de enlace logit para el análisis de los datos del archivo Preñez.IDB2.

A continuación se presentan los resultados del modelo solicitado:

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.020_Condicion_ML<-glm(Condicion~1+DCL
, family=myFamily
, na.action=na.omit
, data=R.data20)
```

Resultados para el modelo: `mlgm.modelo.020_Condicion_ML`

Variable dependiente: Condicion

General

Familia	Enlace	Convergencia	Escala
binomial	logit	Alcanzada	1.00

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
500	278.45	286.87	-137.22	274.45

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			499	297.89	
DCL	1	23.44	498	274.45	<0.0001

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
DCL	1	498	21.57	<0.0001

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.27	0.43	0.62	0.5325
DCL	1.58	0.34	4.64	<0.0001

El diámetro del cuerpo lúteo afecta significativamente el éxito de la inseminación artificial ($p < 0.0001$). Los datos observados (son solo 0 y 1) y la curva ajustadas para este ejemplo se presentan en la Figura 14. En este ejemplo, la pendiente de la línea recta (en la escala logit) está relacionada con el cociente de chances (odds ratio):

$$\exp(\beta_1) = \frac{\text{odds}(TCL+1)}{\text{odds}(TCL)} = \exp(1.58) = 4.85$$

Es decir, que las chances (odds) de quedar preñada se multiplican por 4.85 si el tamaño del cuerpo lúteo aumenta 1 cm. Debemos destacar que no se debe interpretar como que las probabilidades se multiplican por 4.85, ya que no tendría sentido pensar que si la probabilidad de quedar preñada por

ejemplo fuese de 0.60, esto multiplicado por 4.48 da 2.91, y este es un valor no posible para una probabilidad.

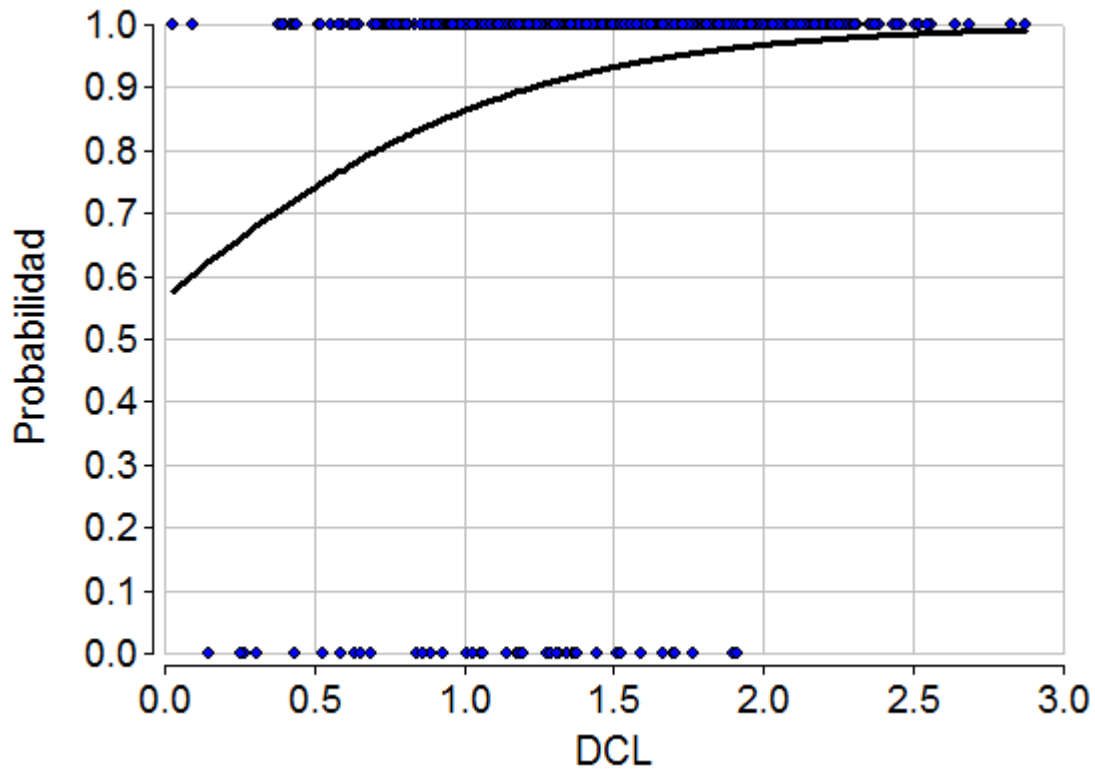


Figura 14: Datos observados (azul) y curva observada (negro) para los datos del ejemplo Preñez.IDB2.

Poisson: Ensayo clínico de progabide

Para probar la efectividad de una droga anti-epiléptica (progabide) se realizó un ensayo clínico con 59 pacientes. Los datos están analizados en Thall y Vail (1990), Breslow y Clayton (1993) y Diggle et al. (2002). Para cada paciente se registró el número de convulsiones durante un periodo de 8 semanas (línea de base). Luego se aleatorizaron en dos grupos: tratado (progabide) y placebo, además de darles a todos la quimioterapia tradicional. Se registró el número de convulsiones de cada paciente en cuatro periodos consecutivos de dos semanas. La pregunta de investigación es si la droga reduce la tasa de convulsiones epilépticas (cantidad de convulsiones/semana).

Las tasas para los pacientes tratados y controles pueden visualizarse en escala de raíz cuadrada:

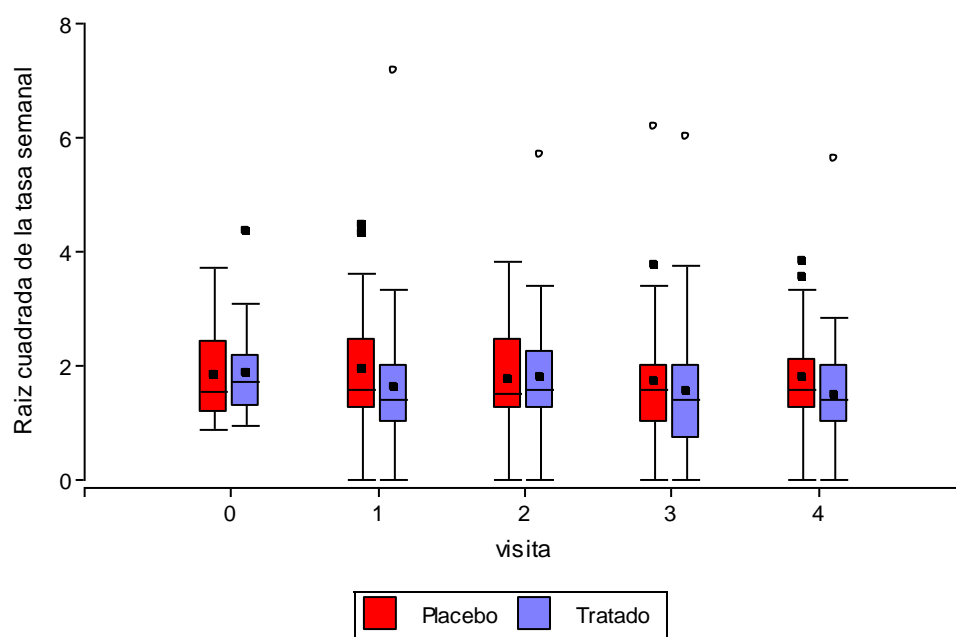


Figura 15: Diagramas de caja para la variable Raíz cuadrada de la tasa de convulsiones (cantidad de convulsiones por semana) para pacientes con placebo y tratados con progabide medidos en el día de inicio del experimento y durante 4 semanas. Archivo *Epilepsia.IDB2*.

Se puede ver que en los pacientes tratados hay una tasa levemente menor que en los pacientes con placebo (especialmente en las semanas 1 y 4).

Inicialmente plantearemos un modelo sin tener en cuenta la naturaleza longitudinal de esta información (es decir, suponiendo que todas las observaciones son independientes). El más simple es

un modelo loglineal que supone que los recuentos Y_{ij} del paciente i en la visita j siguen una distribución Poisson cuya media depende del periodo de observación (t_{ij}), del tratamiento (placebo/tratado), de la visita (pre-aleatorización o post-aleatorización) y de la interacción tratamiento \times visita:

$$Y_{ij} \sim \text{Poisson}(\mu_{ij})$$

$$\log(\mu_{ij}) = \log(t_{ij}) + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2j} + \beta_3 x_{1i} x_{2j}$$

donde

$$x_{1i} = \begin{cases} 0 & \text{si paciente } i \text{ placebo} \\ 1 & \text{si paciente } i \text{ tratado} \end{cases}$$

$$x_{2j} = \begin{cases} 0 & \text{si visita } j = 0 \\ 1 & \text{si visita } j > 0 \end{cases}$$

Debemos observar que para modelar tasas debemos usar $\log(t_{ij})$ como “offset”, para que los parámetros de la componente lineal puedan interpretarse en términos de tasas esperadas y no de recuentos esperados:

$$\text{tasa}_{ij} = \frac{\mu_{ij}}{t_{ij}}, \quad \log(\text{tasa}_{ij}) = \log(\mu_{ij}) - \log(t_{ij})$$

Esto es necesario porque los periodos de observación podrían ser diferentes tanto para cada paciente como para cada visita de seguimiento, y lo que nos interesa es la cantidad de convulsiones por semana (tasa semanal), no el total de convulsiones.

$$\log(\mu_{ij}) = \log(t_{ij}) + \log(\text{tasa}_{ij}) = \log(t_{ij}) + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2j} + \beta_3 x_{1i} x_{2j}$$

β_0 es el logaritmo de la tasa esperada de convulsiones por semana para un paciente del grupo placebo antes de la aleatorización (i.e. asignación de tratamientos).

β_1 es el logaritmo del cociente entre la tasa esperada de pacientes tratados y la tasa esperada de pacientes placebo antes de la aleatorización. Observar que x_{2j} y $x_{1i} x_{2j}$ valen cero antes de la aleatorización y por lo tanto el modelo para $\log(\text{tasa}_{ij})$ ante de la aleatorización será: $\log(\text{tasa}_{i0}) = \beta_0 + \beta_1 x_{1i}$. Luego, para un tratado, el logaritmo de la tasa esperada antes de la aleatorización es $\log(\text{tasa}_{\text{tratado};0}) = \beta_0 + \beta_1$, mientras que para un placebo es $\log(\text{tasa}_{\text{placebo};0}) = \beta_0$

. Luego, el logaritmo de la esperanza del cociente de tasas será $\log\left(\frac{\text{tasa}_{\text{tratado};0}}{\text{tasa}_{\text{placebo};0}}\right) = \beta_0 + \beta_1 - \beta_0 = \beta_1$

.

β_2 es el logaritmo del cociente entre la tasa esperada en pacientes con placebo después y antes de la aleatorización. Observar que el modelo para el logaritmo de la tasa esperada en el placebo antes de la aleatorización es:

$$\log(\text{tasa}_{\text{placebo};0}) = \beta_0 \text{ y,}$$

para después de la aleatorización es:

$$\log(\text{tasa}_{\text{placebo};1}) = \beta_0 + \beta_2.$$

Luego,

$$\log\left(\frac{\text{tasa}_{\text{placebo};1}}{\text{tasa}_{\text{placebo};0}}\right) = \log(\text{tasa}_{\text{placebo};1}) - \log(\text{tasa}_{\text{placebo};0}) = (\beta_0 + \beta_2) - \beta_0 = \beta_2.$$

$\beta_2 + \beta_3$ es el logaritmo del cociente entre la tasa esperada en pacientes tratados después y antes de la aleatorización.

$$\log(\text{tasa}_{\text{tratado};1}) = \beta_0 + \beta_1 + \beta_2 + \beta_3,$$

$$\log(\text{tasa}_{\text{tratado};0}) = \beta_0 + \beta_1,$$

$$\log\left(\frac{\text{tasa}_{\text{tratado};1}}{\text{tasa}_{\text{tratado};0}}\right) = \log(\text{tasa}_{\text{tratado};1}) - \log(\text{tasa}_{\text{tratado};0}) = (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1) = \beta_2 + \beta_3.$$

β_3 es la “ganancia de estar tratado” cuando se evalúa la esperanza del logaritmo de la tasa antes y después de la aleatorización en comparación con lo que ocurre en el placebo. Si β_3 es cero, entonces la esperanza del logaritmo del cociente de tasas antes y después de la aleatorización en pacientes

tratados es igual que en los pacientes que recibieron el placebo (β_2), si β_3 es significativamente menor que 0, entonces el tratamiento es efectivo.

Para ajustar este modelo en InfoStat debemos seleccionar la variable dependiente (convulsiones), la(s) variable(s) de clasificación y las covariables. Como en este caso tenemos variables binarias para tratamiento (placebo o tratado) y para visita (pre o post aleatorización), podríamos ponerlas tanto como variables de clasificación o covariables. En el ejemplo se declararon ambas como covariables para poder interpretar los resultados según lo descrito en los párrafos anteriores. Por otra parte, la variable “offset” siempre debe declararse como covariable, en este caso: $\log(\text{tiempo})$.

Debido a que estamos interesados en los coeficientes de un modelo de regresión, prestaremos atención a la tabla de coeficientes del modelo y las pruebas de hipótesis asociadas a esos coeficientes. La tabla de hipótesis secuenciales se omite explícitamente ya que esas pruebas para el caso que estamos tratando pueden llevar a confusión, especialmente en los modelos que incluyen efectos aleatorios sobre las pendientes.

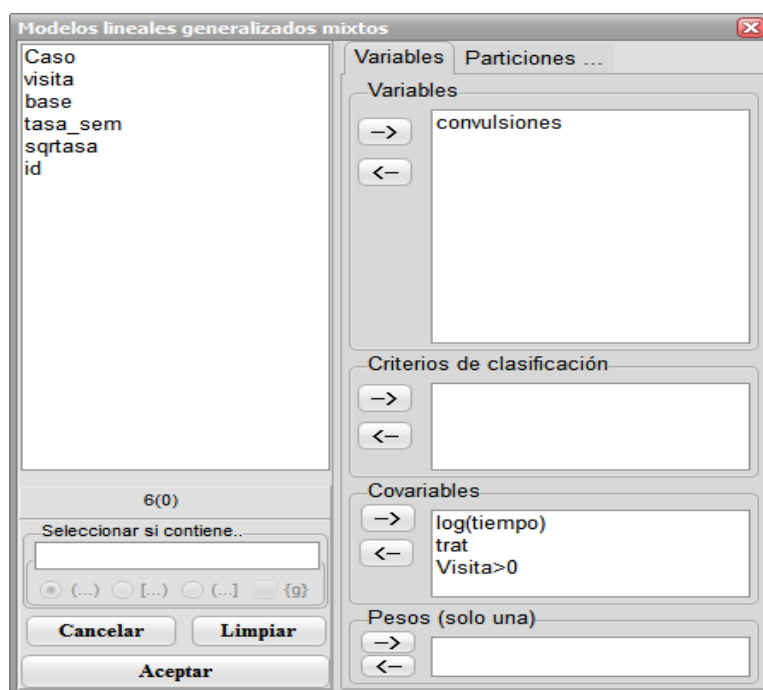


Figura 16: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo *Epilepsia.IDB2*

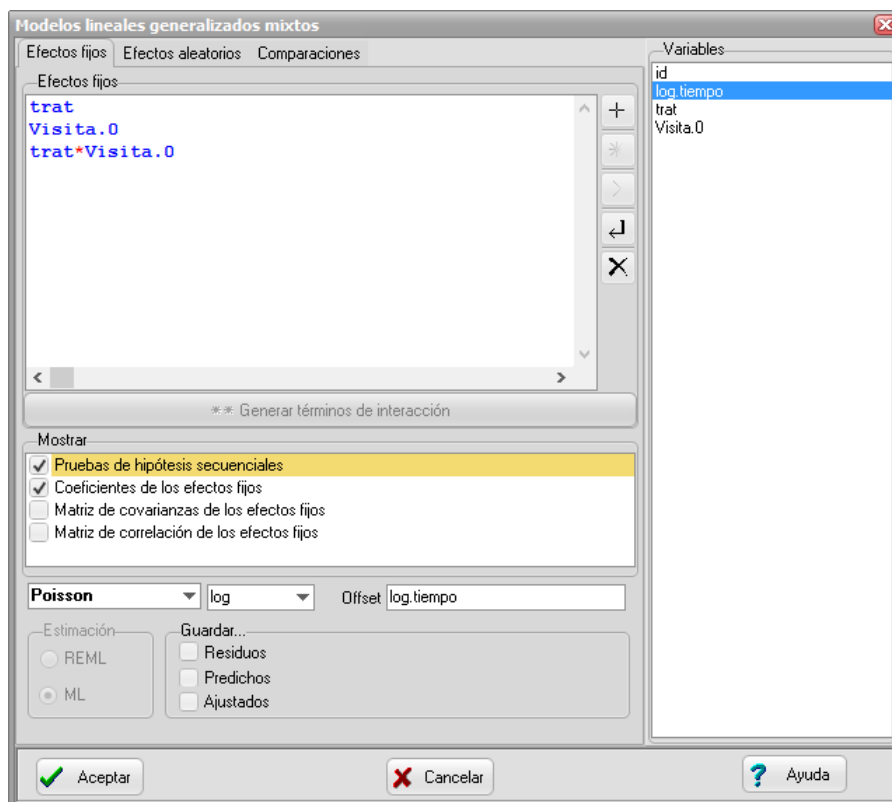


Figura 17: Ventana de Modelos lineales generalizados mixtos con la solapa *Efectos fijos* desplegada y declaración de variables regresoras para el análisis de los datos del archivo *Epilepsia.IDB2* usando una familia *Poisson*.

La salida correspondiente a este modelo incluye información sobre el ajuste al modelo. La “deviance” dividida por los grados de libertad residuales de este modelo (291, que surgen de la diferencia entre el total y los 4 grados de libertad usados para estimar los parámetros del modelo: β_0 , β_1 , β_2 y β_3), es 12.3, que es bastante mayor de 1, por lo que el modelo no ajusta bien (ya sabíamos de antemano que el modelo no consideraba la dependencia de las observaciones del mismo paciente).

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.000_convulsiones_ML<-
glm(convulsiones~1+trat+Visita.0+trat:Visita.0
,family=myFamily
,na.action=na.omit
,offset=log.tiempo
,data=R.data00)
```

Resultados para el modelo: `mlgm.modelo.000_convulsiones_ML`

Variable dependiente: *convulsiones*

General

Familia	Enlace	Convergencia	Escala
poisson	log	Alcanzada	1.00

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
295	4646.90	4661.65	-2319.45	3577.83

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			294	3584.13	
trat	1	0.68	293	3583.45	0.4088
Visita.0	1	3.08	292	3580.37	0.0794
trat:Visita.0	1	2.54	291	3577.83	0.1109

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
trat	1	291	0.32	0.5706
Visita.0	1	291	5.58	0.0188
trat:Visita.0	1	291	2.54	0.1121

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.35	0.03	39.57	<0.0001
trat	0.03	0.05	0.57	0.5702
Visita.0	0.11	0.05	2.36	0.0181
trat:Visita.0	-0.10	0.07	-1.59	0.1110

Una alternativa usada en algunos casos es un modelo similar al anterior, pero con un parámetro adicional para sobredispersión (obliga a un ajuste de la media y la varianza, pero no explica la causa de la falta de ajuste). En este ejemplo se obtiene esto seleccionando la familia "Quasi-Poisson".

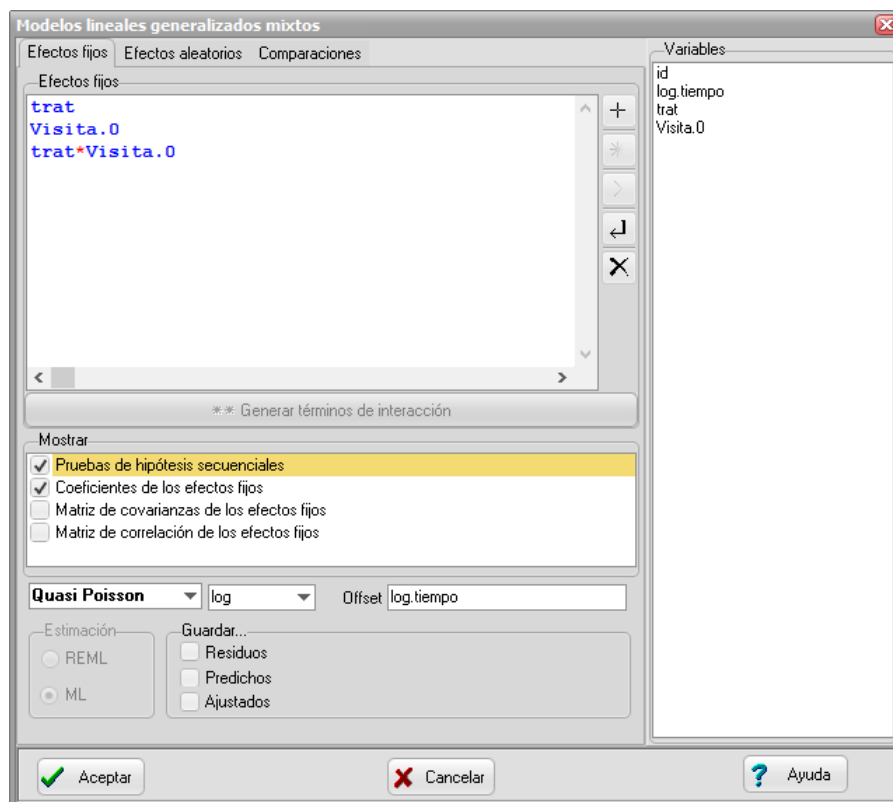


Figura 18: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variables regresoras para el análisis de los datos del archivo *Epilepsia.IDB2* usando una familia Quasi Poisson.

Especificación del modelo en R

```
mlgm.modelo.001_convulsiones_ML<-
glm(convulsiones~1+trat+Visita.0+trat:Visita.0
,family=myFamily
,na.action=na.omit
,offset=log.tiempo
,data=R.data00)
```

Resultados para el modelo: mlgm.modelo.001_convulsiones_ML

Variable dependiente: convulsiones

General

Familia	Enlace	Convergencia	Escala
quasipoisson	log	Alcanzada	19.70

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
---	-----	-----	--------	----------

295 3577.83
 AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			294	3584.13	
trat	1	0.68	293	3583.45	0.8524
Visita.0	1	3.08	292	3580.37	0.6927
trat:Visita.0	1	2.54	291	3577.83	0.7195

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
trat	1	291	0.02	0.8983
Visita.0	1	291	0.28	0.5949
trat:Visita.0	1	291	0.13	0.7198

Efectos fijos

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.35	0.15	8.91	<0.0001
trat	0.03	0.21	0.13	0.8983
Visita.0	0.11	0.21	0.53	0.5949
trat:Visita.0	-0.10	0.29	-0.36	0.7198

En esta salida anterior de los estimadores de los Efectos Fijos, los coeficientes corresponden a:

β_0 (Intercept)

β_1 trat

β_2 Visita.0

β_3 trat:Visita.0

En este caso el parámetro de escala se estima como 19.70, que se obtiene dividiendo el valor de chi-cuadrado de Pearson (no mostrado aquí) por los grados de libertad residuales del modelo (291). Se puede apreciar que el efecto de ninguno de los términos del modelo es significativo.

Ninguno de los modelos ajustados ha considerado explícitamente la naturaleza longitudinal de estos datos. Una manera de incorporarla es incluyendo el efecto del paciente. Por ejemplo, cada paciente podría tener una tasa “basal” diferente. Esto se logra con un modelo que incorpore un efecto aleatorio de paciente sobre el intercepto (β_0). Así, el modelo para el k -ésimo paciente queda:

$$Y_{ijk} | u_k \sim \text{Poisson}(\mu_{ijk})$$

$$\log(\mu_{ijk}) = \log(t_{ij}) + \beta_0 + u_k + \beta_1 x_{1i} + \beta_2 x_{2j} + \beta_3 x_{1i} x_{2j}$$

$$u_k \sim N(0, \sigma_u^2)$$

Este modelo se interpreta como que existe una tasa basal de convulsiones por semana específica para cada individuo, y la correlación entre las observaciones del mismo individuo proviene de compartir este efecto aleatorio sujeto-específico.

Para ajustar este modelo en InfoStat declaramos ID (variable que identifica al paciente) como variable de clasificación y luego la agregamos como efecto aleatorio (Figura 19), dejando la parte fija como en el primer modelo (Figura 18), en el que usamos una función de la familia Poisson.

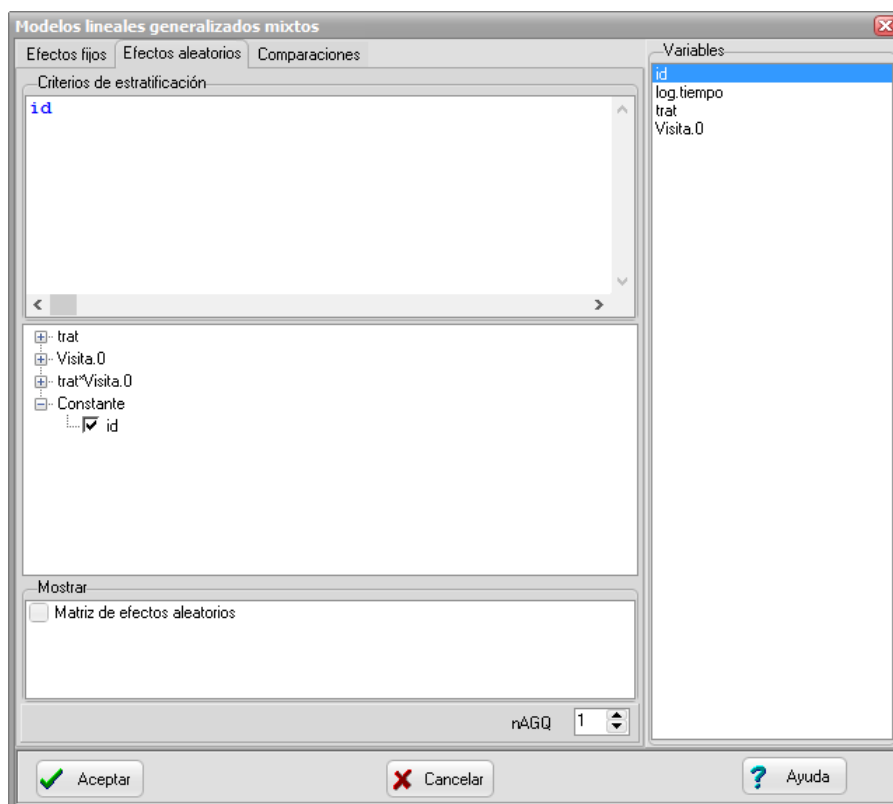


Figura 19: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y la identificación de cada individuo declarada para el análisis de los datos del archivo Epilepsia.IDB2.

Especificación del modelo en R

```
mlgm.modelo.009_convulsiones_ML<-
glmer(convulsiones~1+trat+Visita.0+trat:Visita.0+(1|id)
, family=myFamily
, na.action=na.omit
, REML=F
, AGQ=1
, offset=log.tiempo
, data=R.data09)
```

Resultados para el modelo: mlgm.modelo.009_convulsiones_ML

Variable dependiente: convulsiones

General

Familia	Enlace	nAGQ
poisson	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
295	2031.36	2049.79	-1010.68	705.29

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
trat	0.14	1	0.7104
Visita.0	3.08	1	0.0794
trat:Visita.0	2.54	1	0.1109

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
id	(Intercept)	0.61	0.78

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
β_0 (Intercept)	1.03	0.15	6.77	<0.0001
β_1 trat	-0.02	0.21	-0.11	0.9098
β_2 Visita.0	0.11	0.05	2.37	0.0177
β_3 trat:Visita.0	-0.10	0.06	-1.60	0.1097

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
id	(Intercept)	0.61	0.78

El modelo anterior produjo una disminución de la deviance condicional, y el cociente de deviance/GL es ahora de 2.42. Sin embargo, el modelo es incapaz de detectar el efecto de droga (*i.e.*, β_3 no es significativo).

Un modelo alternativo, que también incorpora efectos aleatorios en β_2 , implica que no sólo hay heterogeneidad entre sujetos en las tasas basales sino también en los cocientes de tasas de convulsiones antes y después de la aleatorización (es análogo a un modelo de regresión con intercepto y pendiente aleatorios):

$$\begin{aligned}
 Y_{ijk} | u_k, v_k &\sim \text{Poisson}(\mu_{ijk}) \\
 \log(\mu_{ijk}) &= \log(t_{ij}) + \beta_0 + u_k + \beta_1 x_{1i} + (\beta_2 + v_k) x_{2j} + \beta_3 x_{1i} x_{2j} \\
 [u_k, v_k] &\sim N\left(\underline{0}, \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)
 \end{aligned}$$

En InfoStat, para incorporar efectos aleatorios en dos parámetros e incluir una posible correlación entre éstos, necesitamos incorporar explícitamente el término “1+Visita” en la parte fija del modelo (Figura 20), de manera de que al incorporar el efecto aleatorio de paciente lo hagamos sobre estos dos términos del modelo simultáneamente y así se puede modelar la correlación entre ambos efectos aleatorios (Figura 21). El “1” es un símbolo que se utiliza para hacer referencia explícita a la constante de un modelo lineal.)

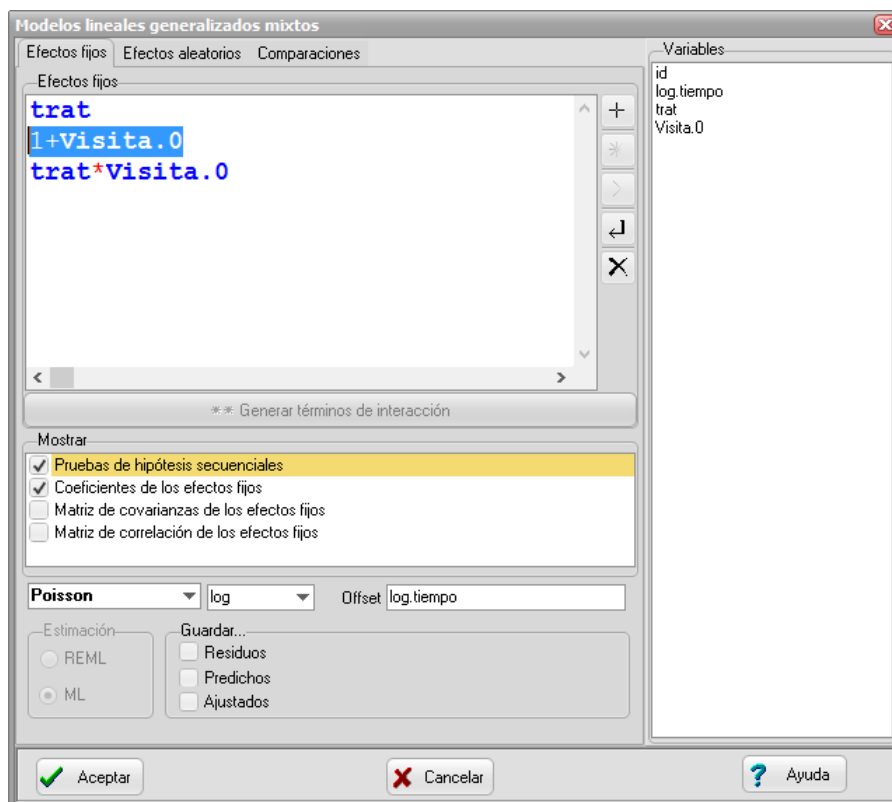


Figura 20: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variables regresoras incluyendo la constante (representada por el número 1) para el análisis de los datos del archivo *Epilepsia.IDB2*.

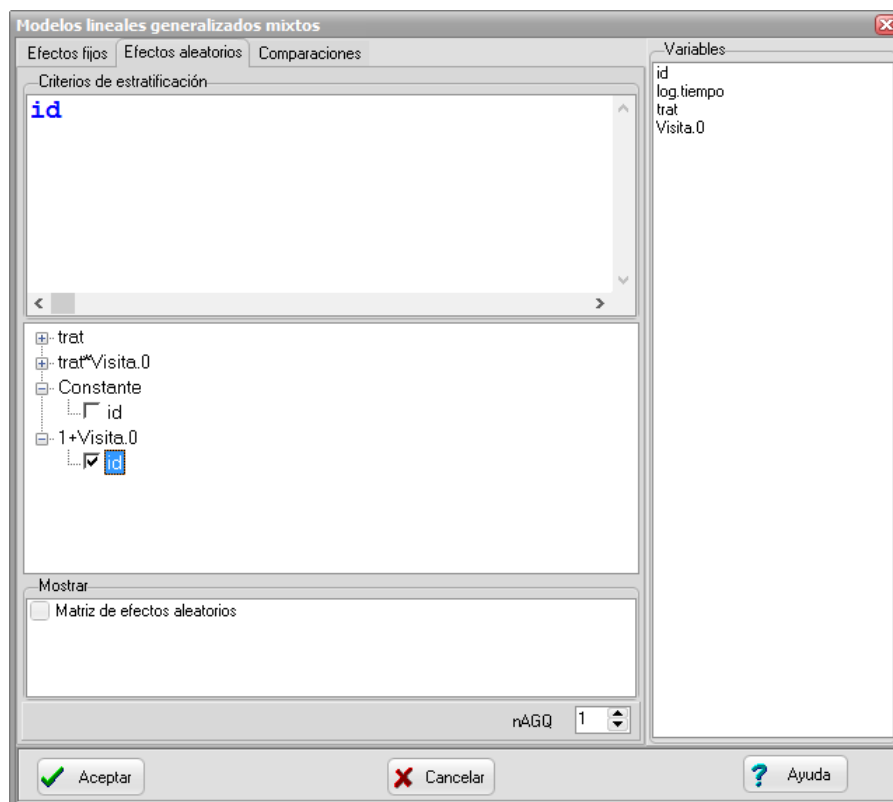


Figura 21: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y la identificación de cada individuo declarada dentro de 1+Visita para el análisis de los datos del archivo *Epilepsia.IDB2*.

Especificación del modelo en R

```
mlgm.modelo.011_convulsiones_ML<-
glmer(convulsiones~1+trat+Visita.0+trat:Visita.0+(1+Visita.0|id)
,family=myFamily
,na.action=na.omit
,REML=F
,AGQ=1
,offset=log.tiempo
,data=R.data09)
```

Resultados para el modelo: mlgm.modelo.011_convulsiones_ML

Variable dependiente: convulsiones

General

Familia	Enlace	nAGQ
poisson	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
295	1863.34	1889.15	-924.67	426.99

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
trat	0.05	1	0.8197
Visita.0	4.32	1	0.0377
trat:Visita.0	4.11	1	0.0426

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
β_0 (Intercept)	1.07130	0.14027	7.63754	<0.0001
β_1 trat	0.04948	0.19272	0.25675	0.7974
β_2 Visita.0	-0.00239	0.10909	-0.02195	0.9825
β_3 trat:Visita.0	-0.30716	0.15045	-2.04157	0.0412

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD	Corr 1
id	(Intercept)	0.50	0.71	
id	Visita.0	0.23	0.48	0.17

Podemos apreciar que este modelo ajusta mejor que el anterior (los valores de AIC y BIC son menores y el cociente deviance/GL es $426.99/291=1.47$). Se puede verificar que β_3 es significativo ($p=0.0412$ en la prueba de Wald), y su valor estimado tiene un signo negativo, por lo que se puede concluir que el tratamiento tiene el efecto de disminuir la tasa promedio de convulsiones para un sujeto dado.

Un sujeto del grupo placebo tiene aproximadamente la misma tasa promedio de convulsiones antes o después de la aleatorización (asignación de tratamientos β_2 no significativa), mientras que un sujeto tratado promedio tiene una tasa de convulsiones significativamente menor después de la aleatorización (asignación de tratamientos) (aproximadamente una reducción del 26%):

$$\exp(\hat{\beta}_2) = \exp(-0.02398) = 0.997$$

$$\exp(\hat{\beta}_2 + \hat{\beta}_3) = \exp(-0.02398 - 0.30716) = 0.734$$

Una estimación de la tasa esperada de convulsiones por semana para un sujeto “típico” (con efectos aleatorios $u=0$, $v=0$) del grupo tratado en el periodo post-aleatorización (es decir, una vez que está tomando el medicamento) es:

$$tasa_{(tratado,1)} = \exp(1.071 + 0.049 - 0.002 - 0.307) = 2.25$$

La tasa análoga para un sujeto “típico” del grupo placebo es:

$$tasa_{placebo,1} = \exp(1.071 - 0.002) = 2.91.$$

Por otra parte, la tasa promedio poblacional de convulsiones por semana para pacientes del grupo tratado en el periodo post-aleatorización se deriva de la siguiente manera. Recordemos que en el último modelo anterior (con efectos aleatorios) el número de convulsiones Y_{ijk} por paciente en un momento y tratamiento dados tiene, condicional a los efectos aleatorios, distribución Poisson $(Y_{ijk} | u_k, v_k \sim Poisson(\mu_{ijk}))$ con media μ_{ijk} . Supusimos además que:

$$[u_k, v_k] \sim N\left(\underline{0}, \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right) \text{ y asumimos que}$$

$$\log(\mu_{ijk} / t_{ij}) = \beta_0 + u_k + \beta_1 x_{1i} + (\beta_2 + v_k) x_{2j} + \beta_3 x_{1i} x_{2j}$$

En forma matricial $\log(\tau_k) = \mathbf{x}_k \boldsymbol{\beta} + \mathbf{z}'_k \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix}$, de donde el modelo para la tasa τ_k será:

$$\tau_k = \exp\left(\mathbf{x}_k \boldsymbol{\beta} + \mathbf{z}'_k \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix}\right) = \exp(\mathbf{x}_k \boldsymbol{\beta}) \exp\left(\mathbf{z}'_k \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix}\right). \text{ Si tomamos esperanza de } \tau_k, \text{ nos queda}$$

$$E(\tau_k) = \exp(\mathbf{x}_k \boldsymbol{\beta}) E\left(\exp\left(\mathbf{z}'_k \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix}\right)\right). \text{ Podemos reconocer en } E\left(\exp\left(\mathbf{z}'_k \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix}\right)\right) \text{ la función}$$

generatriz de momentos de un vector normal bi-variado $\begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix}$ evaluada en \mathbf{z}_k . Desarrollando la

función generatriz resulta $m(\mathbf{z}_k) = \exp\left(\mathbf{z}_k E \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix} + \frac{\mathbf{z}_k \Sigma \mathbf{z}_k}{2}\right)$ y dado que $E \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix} = \mathbf{0}$, la expresión

se reduce a $m(z_k) = \exp\left(\frac{z'_k \Sigma z_k}{2}\right)$ donde Σ es $\text{cov}(\mu_k, \nu_k) = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$ y

$z'_k = [1, 1] \forall k$. Luego la expresión final para la esperanza es:

$$E(\tau_k) = \exp(\mathbf{x}_k \boldsymbol{\beta}) [1 \quad 1] \frac{\begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

En nuestro ejemplo la tasa promedio poblacional de un individuo tratado será entonces:

$$\text{tasa}_{(\text{prom. pobl. tratado})} = \exp\left(1.071 + 0.049 - 0.002 - 0.307 + \frac{1}{2} [1 \quad 1] \begin{bmatrix} 0.500 & 0.057 \\ 0.057 & 0.232 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = 3.43$$

mientras que para pacientes del grupo placebo en el mismo periodo es:

$$\text{tasa}_{(\text{prom. pobl. placebo})} = \exp\left(1.071 - 0.002 + \frac{1}{2} [1 \quad 1] \begin{bmatrix} 0.500 & 0.057 \\ 0.057 & 0.232 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = 4.45.$$

Otra alternativa para modelar este conjunto de datos con sobredispersión es usar como distribución condicional a la Binomial Negativa con función de enlace Log (Figura 22). En este caso sólo es necesario agregar un efecto aleatorio de sujeto sobre la tasa basal para lograr el ajuste adecuado (Figura 23).

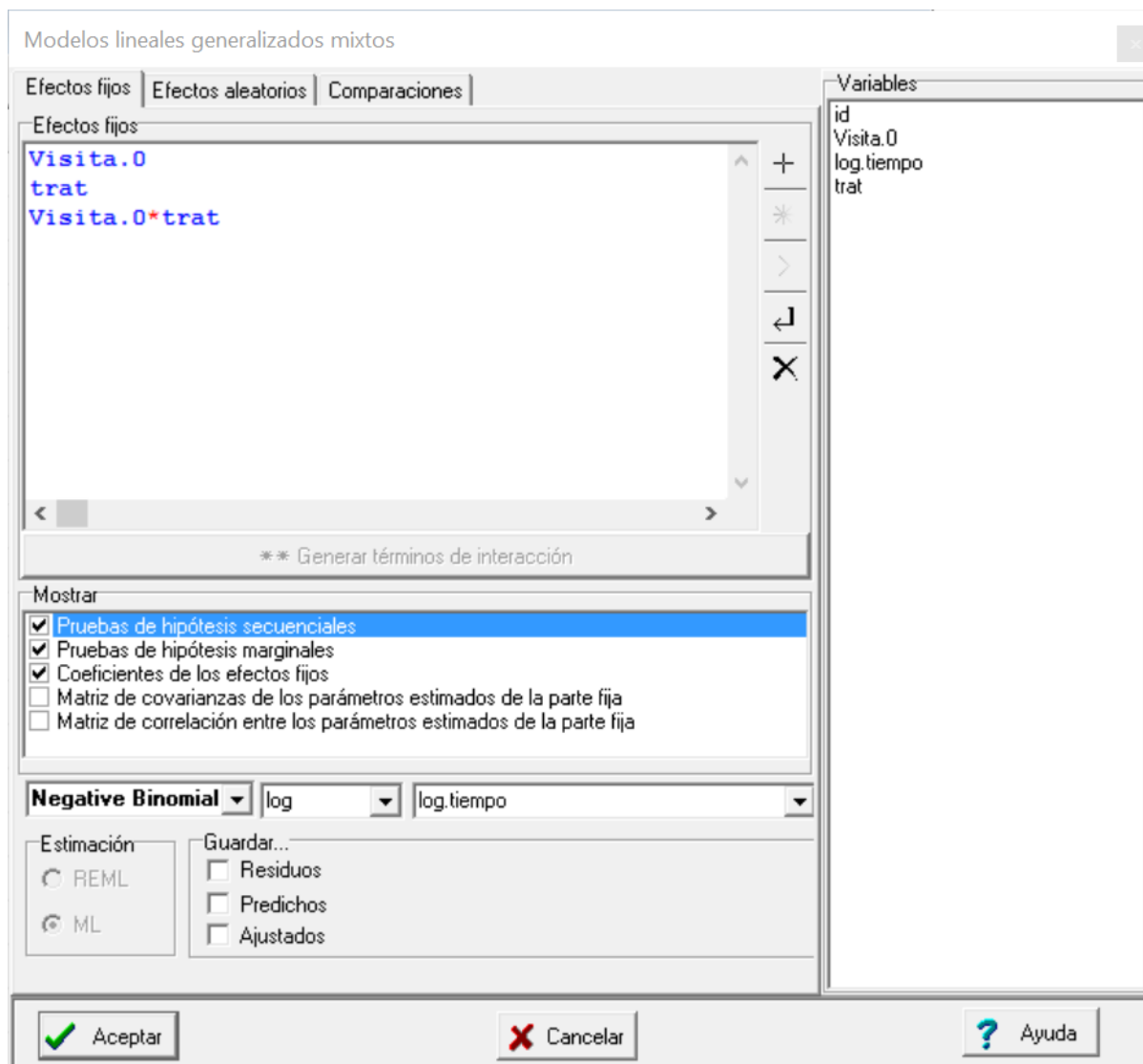


Figura 22: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variables regresoras para el análisis de los datos del archivo Epilepsia.IDB2 usando el modelo Binomial negativo.

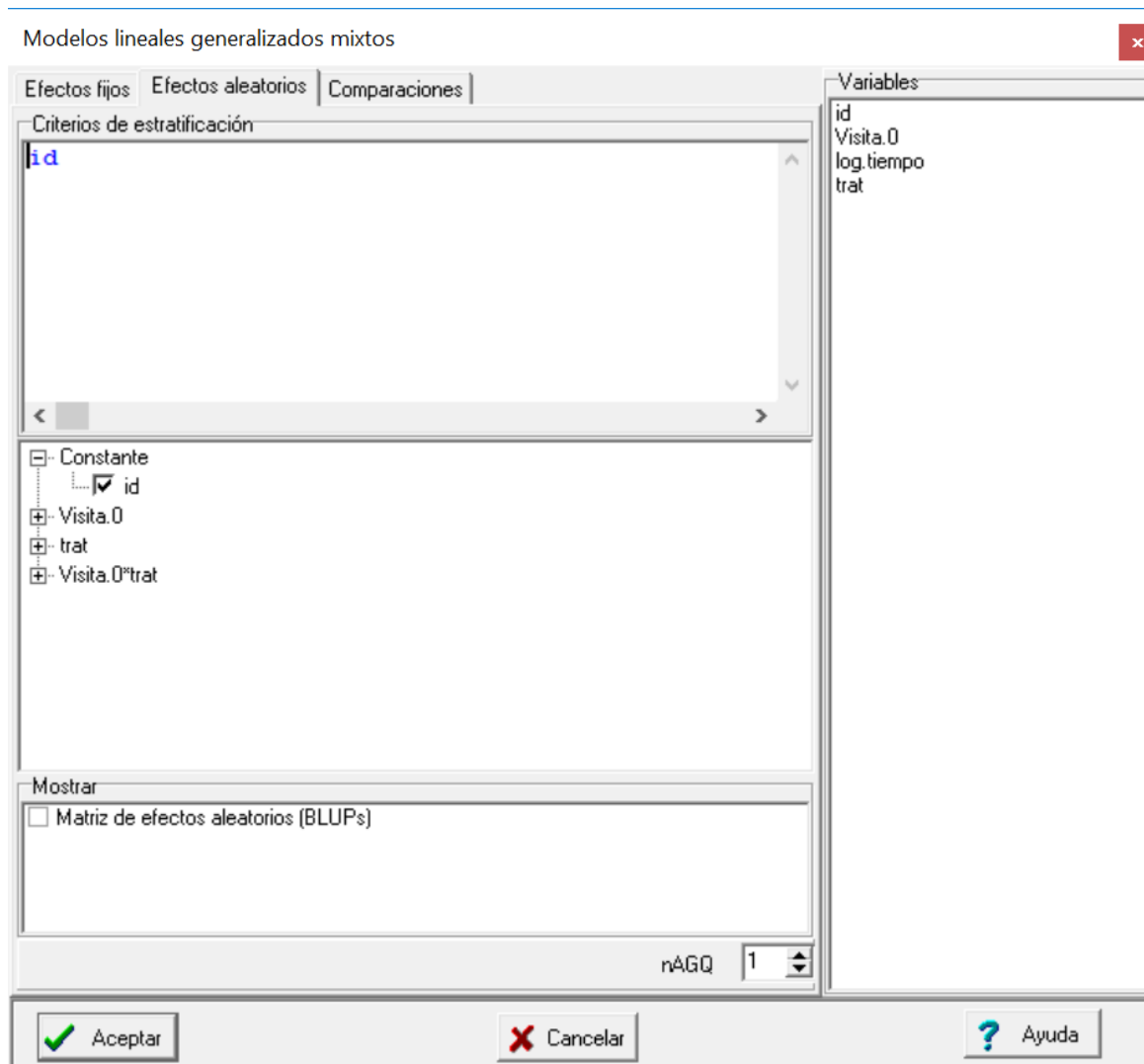


Figura 23: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y declaración de efecto aleatorio de id para el análisis de los datos del archivo Epilepsia.IDB2 usando el modelo Binomial negativo.

Podemos apreciar que este modelo ajusta mejor que los anteriores usando Poisson (los valores de AIC y BIC son menores y el cociente deviance/GL es $274.08/291=0.94$). Se puede verificar que β_3 es significativo ($p=0.0272$ en la prueba de Wald), y su valor estimado tiene un signo negativo (-0.31), por lo que se puede concluir que el tratamiento tiene el efecto de disminuir la tasa promedio de convulsiones para un sujeto dado.

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.000_convulsiones_ML<-
glmer.nb(convulsiones~1+Visita.0+trat+Visita.0:trat+(1|id)
,link=myFamily$link
,na.action=na.omit
,REML=F
,AGQ=1
,offset=log.tiempo
,data=R.data00)
```

Resultados para el modelo: mlgm.modelo.000_convulsiones_ML

Variable dependiente: convulsiones

General

Familia	Enlace	nAGQ
Negative Binomial(6.7803)	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
295	1789.46	1811.58	-888.73	274.08

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
Visita.0	3.87	1	0.0493
trat	0.41	1	0.5239
Visita.0:trat	4.88	1	0.0272

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.08	0.18	6.16	<0.0001
Visita.0	0.02	0.10	0.23	0.8149
trat	0.07	0.24	0.30	0.7644
Visita.0:trat	-0.31	0.14	-2.21	0.0272

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
id	(Intercept)	0.66	0.81

Poisson con sobredispersión: Comparación de herbicidas

Los datos malezas.IDB2 representan un experimento para comparar 3 herbicidas, realizado en un diseño completamente aleatorizado con 6 repeticiones. La variable malezas es el recuento de malezas en un marco de 1 m² aleatoriamente ubicado en cada parcela luego de dos semanas de aplicar el herbicida.

Como los datos son recuentos, primero se ajustará un modelo lineal generalizado con distribución Poisson y enlace logarítmico (Figura 24). Hay un único efecto fijo, los tratamientos.

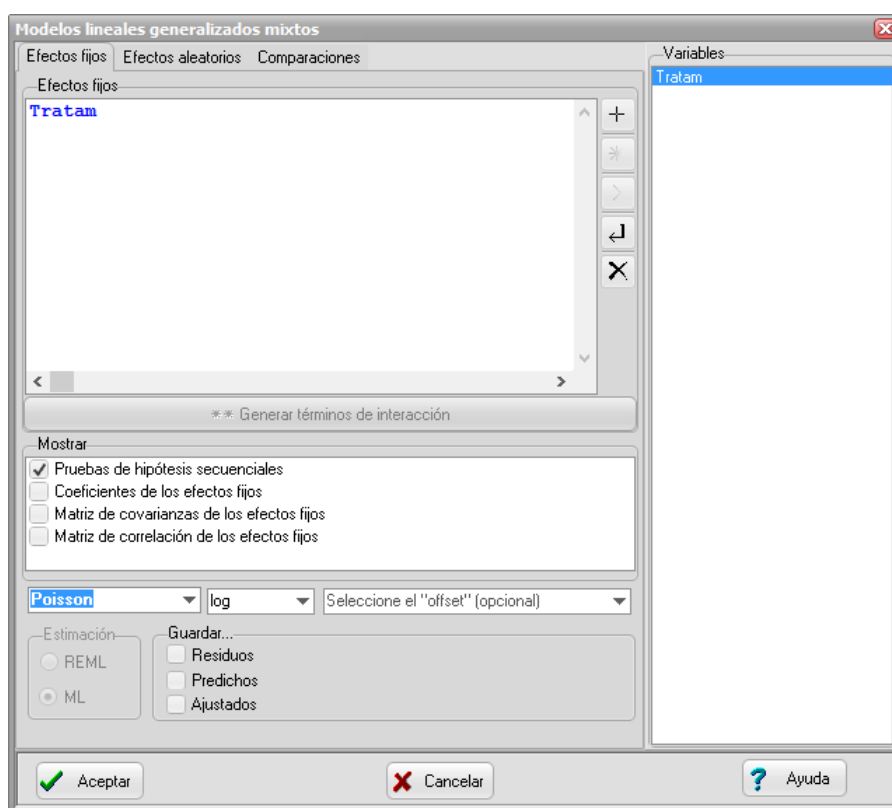


Figura 24: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y la declaración de “Tratam” como único término fijo del modelo. La familia seleccionada es Poisson. Datos: Malezas.IDB2.

La comparación de las medias de tratamiento la podemos hacer con una prueba LSD (Figura 25).

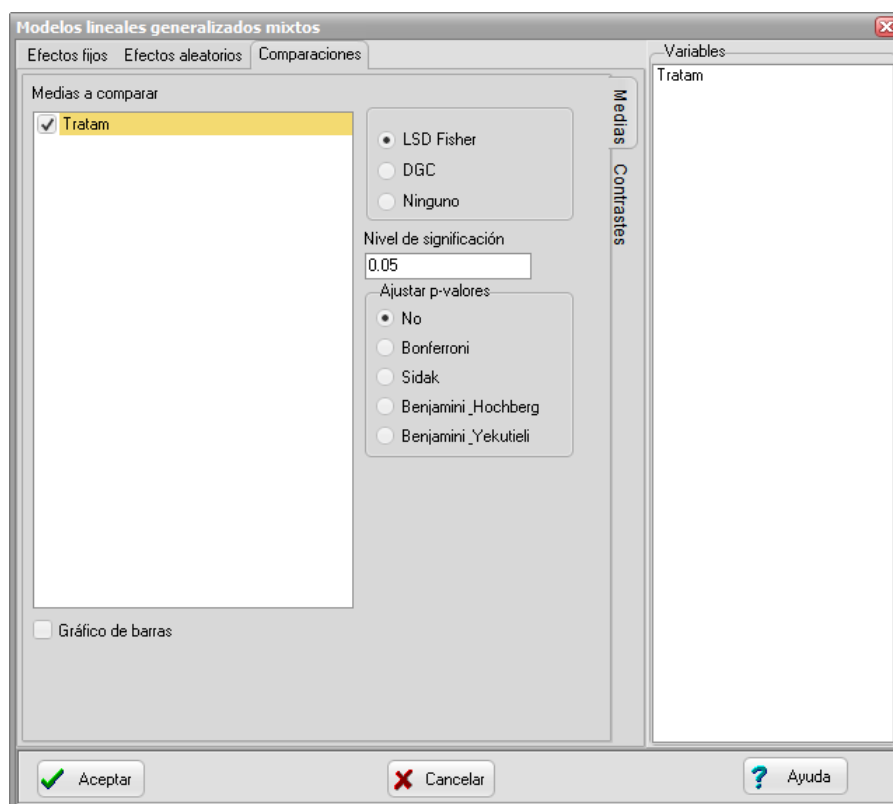


Figura 25: Ventana de Modelos lineales generalizados mixtos con la solapa Comparaciones desplegada y la declaración de “Tratam” como único término fijo del modelo para el cual se piden la comparación de medias mediante LSD. Datos: Malezas.IDB2.

La salida para estas especificaciones se presenta a continuación.

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.003_Malezas_ML<-glm(Malezas~1+Tratam
, family=myFamily
, na.action=na.omit
, offset=NULL
, data=R.data02)
```

Resultados para el modelo: mlgm.modelo.003_Malezas_ML

Variable dependiente: Malezas

General

Familia	Enlace	Convergencia	Escala
poisson	log	Alcanzada	1.00

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
18	149.77	152.44	-71.89	89.31

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			17	135.70	
Tratam	2	46.38	15	89.31	<0.0001

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
Tratam	2	15	20.20	0.0001

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.85	0.27	3.17	0.0015
Tratam2	0.73	0.33	2.24	0.0252
Tratam3	1.64	0.29	5.61	<0.0001

Malezas - Medias ajustadas y errores estándares para Tratam

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

Tratam	PredLin	E.E.	Media	E.E.	
3	2.48	0.12	12.00	1.41	A
2	1.58	0.19	4.83	0.90	B
1	0.85	0.27	2.33	0.62	C

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

Podemos observar que hay diferencias significativas entre todos los tratamientos. Como este es un modelo lineal generalizado de efectos fijos, debemos observar el cociente Deviance/gl, cuyo valor esperado asintótico debería ser cercano 1 si el modelo ajusta bien. En este problema el cociente es $89.31/15 = 5.95$, por lo que hay sobredispersión y las conclusiones obtenidas por este modelo no serían válidas (es muy posible que los errores estándar de las medias estén subestimados).

Una alternativa para situaciones en las que el modelo lineal generalizado presenta sobredispersión es el uso de un modelo relacionado, que incluye un parámetro adicional que multiplica la varianza esperada (recordemos que en datos Poisson la varianza debe ser igual a la media, por lo que el modelo con un parámetro de sobredispersión multiplicativo permitiría explicar varianzas mayores que la media). En este caso no se usa una distribución de probabilidades, y entonces se plantean las ecuaciones de “cuasi verosimilitud”, que son similares a las ecuaciones de verosimilitud para una distribución Poisson (es posible hacer lo mismo para la distribución binomial). En InfoStat se denomina *quasi Poisson* (Figura 26), y es importante notar que, si bien no es posible realizar inferencias por verosimilitud, las otras pruebas que se presentan (Wald) son aproximadamente válidas para muestras grandes.

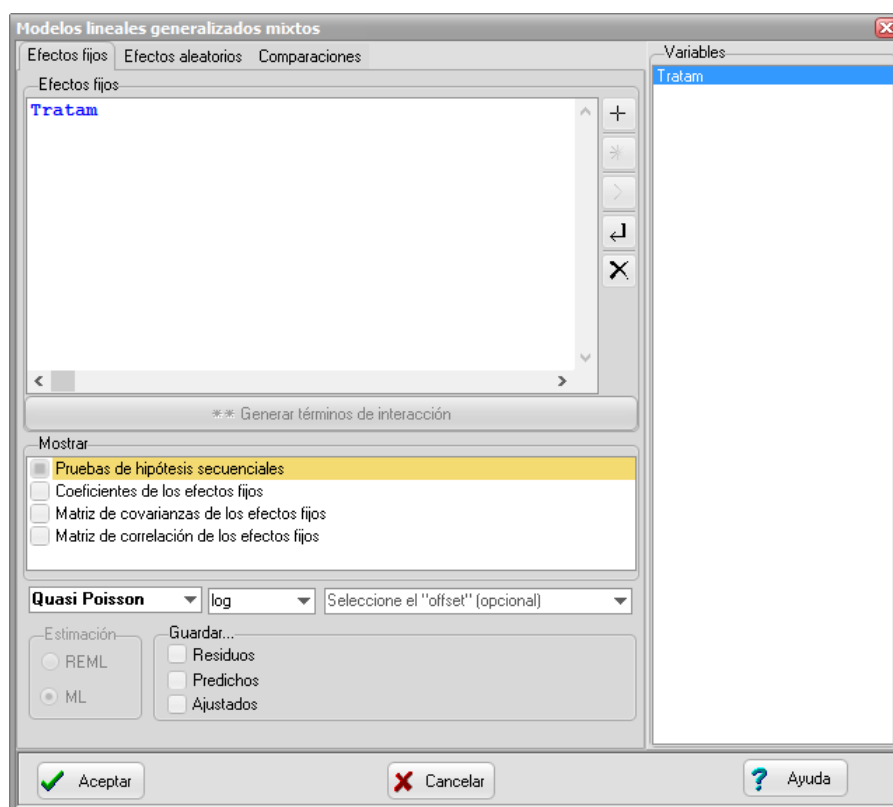


Figura 26: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y la declaración de “Tratam” como único término fijo del modelo. La familia seleccionada es *Quasi-Poisson*. Datos: Malezas.IDB2.

Los resultados son mucho más conservadores que los obtenidos con un modelo Poisson, debido al hecho de que los errores estándar son más grandes (el parámetro de sobredispersión estimado por chi-cuadrado de Pearson / gl es 6.46 y figura como “Escala” en la tabla General). Los errores estándar de los estimadores de efectos fijos y de los predictores lineales son raíz(6.46)=2.54 veces más grandes.

Modelos lineales generalizados mixtos**Especificación del modelo en R**

```
mlgm.modelo.001_Malezas_ML<-glm(Malezas~1+Tratam
, family=myFamily
, na.action=na.omit
, offset=NULL
, data=R.data00)
```

Resultados para el modelo: mlgm.modelo.001_Malezas_ML

Variable dependiente: Malezas

General

Familia	Enlace	Convergencia	Escala
quasipoisson	log	Alcanzada	6.46

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
18				89.31

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			17	135.70	
Tratam	2	46.38	15	89.31	0.0276

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
Tratam	2	15	3.13	0.0732

Efectos fijos

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.85	0.68	1.25	0.2313
Tratam2	0.73	0.83	0.88	0.3924
Tratam3	1.64	0.74	2.21	0.0434

Malezas - Medias ajustadas y errores estándares para Tratam

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

Tratam	PredLin	E.E.	Media	E.E.	
3	2.48	0.30	12.00	3.59	A
2	1.58	0.47	4.83	2.28	A B
1	0.85	0.68	2.33	1.58	B

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

Otra alternativa que permite explicar mejor la sobredispersión es mediante la introducción de un efecto aleatorio de unidad experimental (Stroup, 2013). En este caso los datos se pueden explicar mediante un modelo lineal generalizado mixto:

$$\log(\lambda_{ij}) = \mu + \tau_i + u_{ij}$$

$$y_{ij} | u_{ij} \text{ i.d. } \text{Poisson}(\lambda_{ij})$$

$$u \text{ i.d. } N(0, \sigma_u^2)$$

Se puede probar que la varianza (marginal) de Y_{ij} es mayor que λ_{ij} por lo que este es un modelo con sobredispersión. La ventaja que tiene este modelo es que provienen de una verosimilitud válida, por lo que es posible realizar inferencias basadas en métodos de verosimilitud. Además, la variabilidad entre unidades experimentales es una forma muy clara de explicar la sobredispersión, ya que es un principio fundamental de la experimentación, postulado por Fisher hace más de 80 años.

Para ajustar este modelo agregamos “parcela” como criterio de clasificación, y la seleccionamos como criterio de estratificación en la solapa de efectos aleatorios (Figura 27).

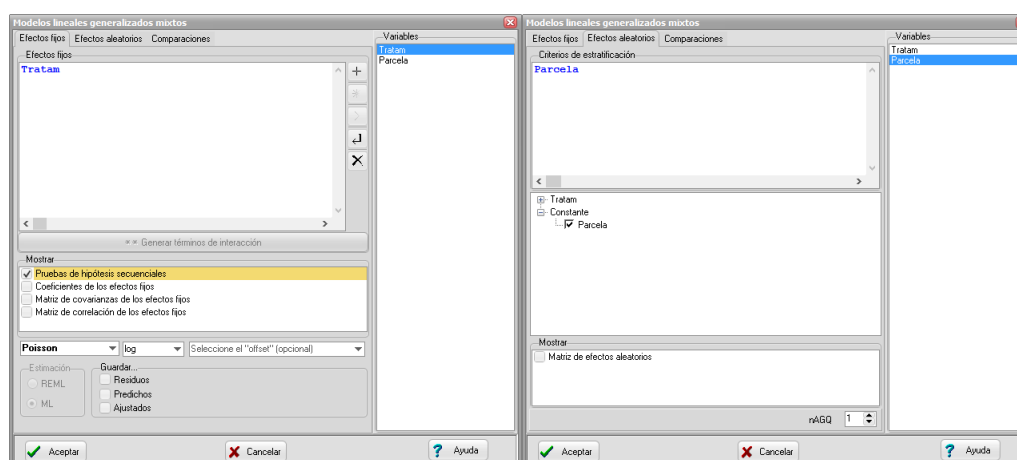


Figura 27: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y la declaración de "Tratam" como único término fijo del modelo y la declaración de la Poisson (derecha). Solapa de efectos aleatorios indicando que la inclusión del efecto aleatorio Parcela. Datos: Malezas.IDB2.

Los resultados indican un AIC bastante menor que el modelo Poisson sin efecto aleatorio, y los errores estándar son más grandes, por lo que los resultados de la prueba LSD no son tan liberales como en el caso de modelo Poisson de efectos fijos. El cociente de Deviance/GL es $8.09/15=0.54$, bastante menor que 1 indicando que no hay sobredispersión.

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.006_Malezas_ML<-glmer(Malezas~1+Tratam+(1|Parcela)
, family=myFamily
, na.action=na.omit
, REML=F
, AGQ=1
, offset=NULL
, data=R.data02)
```

Resultados para el modelo: mlgm.modelo.006_Malezas_ML

Variable dependiente: Malezas

General

Familia	Enlace	nAGQ
poisson	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
18	105.64	109.20	-48.82	8.09

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
Tratam	6.61	2	0.0367

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
--------	-------	-------	---------	---------

Tratam	2	15	3.80	0.0462
--------	---	----	------	--------

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.57	0.44	1.30	0.1950
Tratam2	0.84	0.57	1.48	0.1400
Tratam3	1.52	0.55	2.75	0.0059

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Parcela	(Intercept)	0.57	0.76

Malezas - Medias ajustadas y errores estándares para Tratam

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

Tratam	PredLin	E.E.	Media	E.E.	
3	2.09	0.35	8.11	2.85	A
2	1.41	0.37	4.11	1.53	A B
1	0.57	0.44	1.77	0.78	B

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

Otra alternativa para modelar este conjunto de datos con sobredispersión es usar como distribución condicional a la Binomial Negativa con función de enlace Log y efectos fijos de tratamiento. Podemos apreciar que este modelo ajusta bien, mejorando sobre el Poisson de efectos fijos (el valor de AIC es 105.4, comparable al Poisson de efecto aleatorio, y el cociente deviance/GL es $19.79/15=1.32$).

Modelos lineales generalizados mixtos**Especificación del modelo en R**

```
mlgm.modelo.007_Malezas_ML<-glm.nb(Malezas~1+Tratam
,link=myFamily$link
,na.action=na.omit
,offset(NULL)
,data=R.data02)
```

Resultados para el modelo: mlgm.modelo.007_Malezas_ML

Variable dependiente: Malezas

General

Familia	Enlace	Convergencia	Escala
Negative Binomial (1.7075)	log	Alcanzada	1.00

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
18	105.40	108.96	-48.70	19.79

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			17	30.01	
Tratam	2	10.22	15	19.79	0.0060

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
Tratam	2	15	4.95	0.0224

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.85	0.41	2.06	0.0393
Tratam2	0.73	0.55	1.33	0.1845
Tratam3	1.64	0.53	3.09	0.0020

Malezas - Medias ajustadas y errores estándares para Tratam

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

Tratam	PredLin	E.E.	Media	E.E.	
3	2.48	0.33	12.00	4.01	A
2	1.58	0.36	4.83	1.76	A B
1	0.85	0.41	2.33	0.96	B

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

Cuando se usa la Binomial negativa como alternativa para modelar la sobredispersión no es necesario declarar un efecto aleatorio de unidad experimental (en este caso parcela), ya que el parámetro de agregación fue suficiente para explicar la sobredispersión. Si en este ejemplo usando la binomial negativa declaramos además el efecto de la parcela, se obtiene un modelo menos parsimonioso ya que los criterios de ajuste AIC y BIC suben. A continuación, se presenta la salida del modelo binomial negativo agregando el efecto aleatorio de parcela. Se puede ver que el AIC pasa de 105.40 a 107.63,

mientras que BIC pasa de 108.96 a 112.09, indicando que no es necesario el efecto aleatorio de parcela si se usa este modelo para estos datos.

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.008_Malezas_ML<-glmer.nb(Malezas~1+Tratam+(1|Parcela)
,link=myFamily$link
,na.action=na.omit
,REML=F
,AGQ=1
,offset=NULL
,data=R.data02)
```

Resultados para el modelo: mlgm.modelo.008_Malezas_ML

Variable dependiente: Malezas

General

Familia	Enlace	nAGQ
Negative Binomial(972.6398..	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
18	107.63	112.09	-48.82	8.11

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
Tratam	6.61	2	0.0367

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
Tratam	2	15	3.80	0.0462

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.57	0.44	1.30	0.1948
Tratam2	0.84	0.57	1.48	0.1401

Tratam3	1.52	0.55	2.75	0.0060
---------	------	------	------	--------

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Parcela	(Intercept)	0.57	0.76

Malezas - Medias ajustadas y errores estándares para Tratam*Inversa de la función de enlace con efecto aleatorio=0**LSD Fisher (Alfa=0.05)**Procedimiento de corrección de p-valores: No*

Tratam	PredLin	E.E.	Media	E.E.	
3	2.09	0.35	8.11	2.85	A
2	1.41	0.37	4.11	1.53	A B
1	0.57	0.44	1.77	0.78	B

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

Regresión Poisson con efectos aleatorios: Análisis de hábitat de escorpiones

Para estudiar las variables ambientales que afectan la abundancia de escorpiones se realizó un plan de muestreo en dos sitios (bosque maduro y secundario) del Parque Provincial y Reserva Forestal Chancaní (31°22'13,21" S; 65°27'13,75" O), ubicado en el oeste de la provincia de Córdoba, Argentina. En cada sitio se delimitaron 15 transectas fijas de 50 m de largo por 6 m de ancho cada una, separadas

entre sí por 70 m (Nime et al. 2013; Nime et al. 2016).



El muestreo se llevó a cabo en el mes de noviembre de 2009 durante 16 noches (de luna nueva a luna llena), alternando el lugar de muestreo (7 noches en cada sitio). Para cada noche se obtuvo el porcentaje (0% luna nueva - 100% luna llena) de luna visible mediante el programa MoonTool (<http://www.fourmilab.ch/moontoolw>). Las

variables ambientales se midieron en cada transecta y cada noche de muestreo: temperatura del aire (a 120 cm por encima del nivel del suelo, °C), humedad relativa (%), presión atmosférica (Pa) y la velocidad del viento (a 2 m sobre el nivel del suelo y en la superficie del suelo, m/s) (Nime et al. 2013). También se registró la hora de evaluación de cada transecta para evaluar su posible efecto.

Mediante observación con luz UV se realizó el conteo de los escorpiones de cada una de las 7 especies presentes en la Reserva de Chancaní (Archivo Escorpiones.IDB2). En este ejemplo se usará como variable respuesta al conteo total de escorpiones (Total), el efecto de Sitio como variable clasificatoria y todas las variables ambientales como covariables (Figura 28).

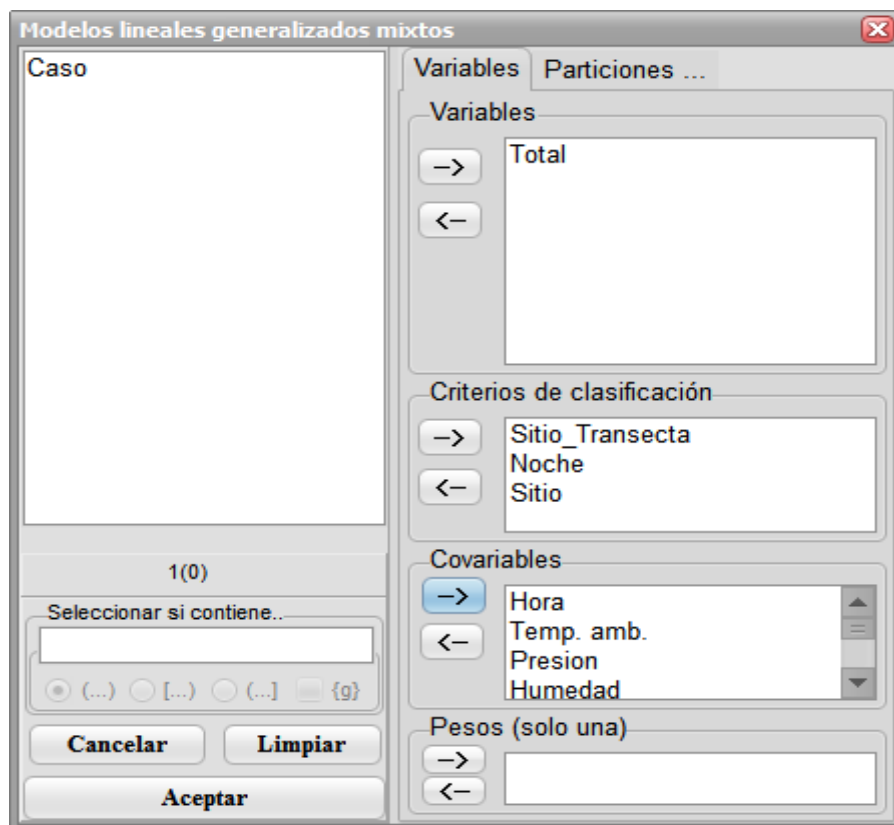


Figura 28: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo *Escorpiones.IDB2*.

En la solapa de efectos fijos se declararon todas las covariables y el efecto de sitio en último lugar. Esto permite interpretar las pruebas hipótesis secuenciales como el efecto de sitio dado que han sido ya descontados los efectos de las covariables. Se seleccionó la distribución Poisson para modelar los conteos y como función de enlace “log” (Figura 29). En este ejemplo no hay una variable para el *Offset*, ya que el esfuerzo de muestreo fue el mismo por cada transecta. Debido que hay varios recuentos en cada transecta y que por ello, pueden estar correlacionados, que se consideró el efecto de transecta como aleatorio. A su vez, cada transecta es evaluada repetidamente en el tiempo, induciendo correlaciones temporales. Una estrategia para modelar esta correlación es la incorporación del efecto de transecta como aleatorio. Para modelar completamente la estructura del diseño de muestreo y dado que todas las transectas de un mismo sitio eran evaluadas en la misma noche, en la solapa de efectos aleatorios se agregó también el efecto de Noche dentro de transecta (Figura 30).

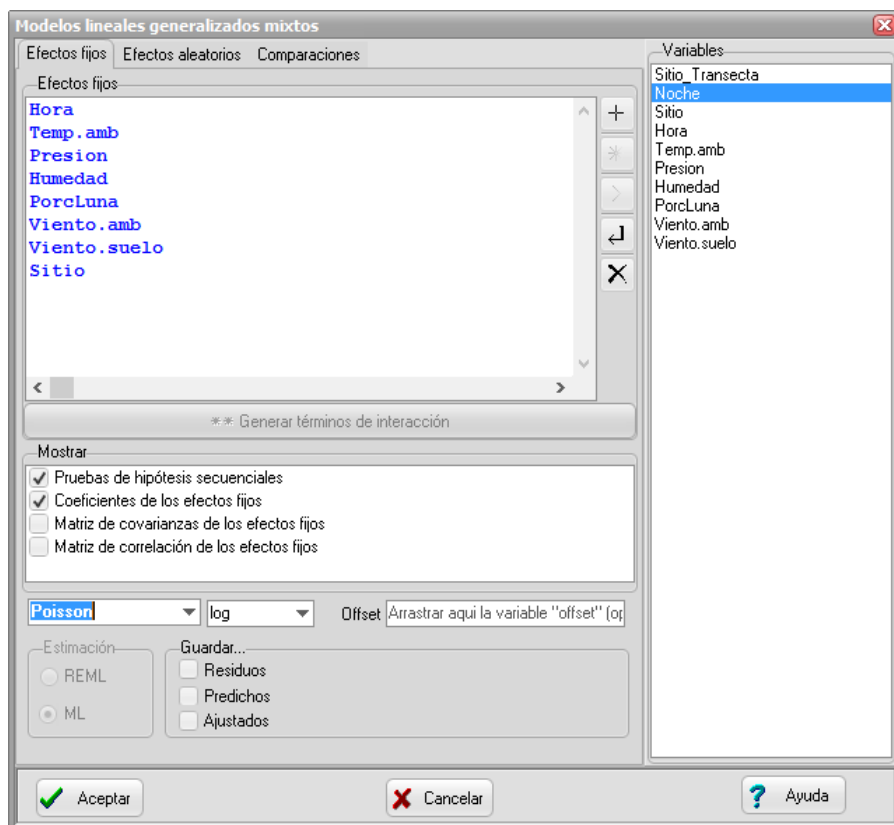


Figura 29: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración del efecto de sitio y todas las covariables ambientales, usando familia Poisson y función de enlace log para el análisis de los datos del archivo Escorpiones.IDB2.

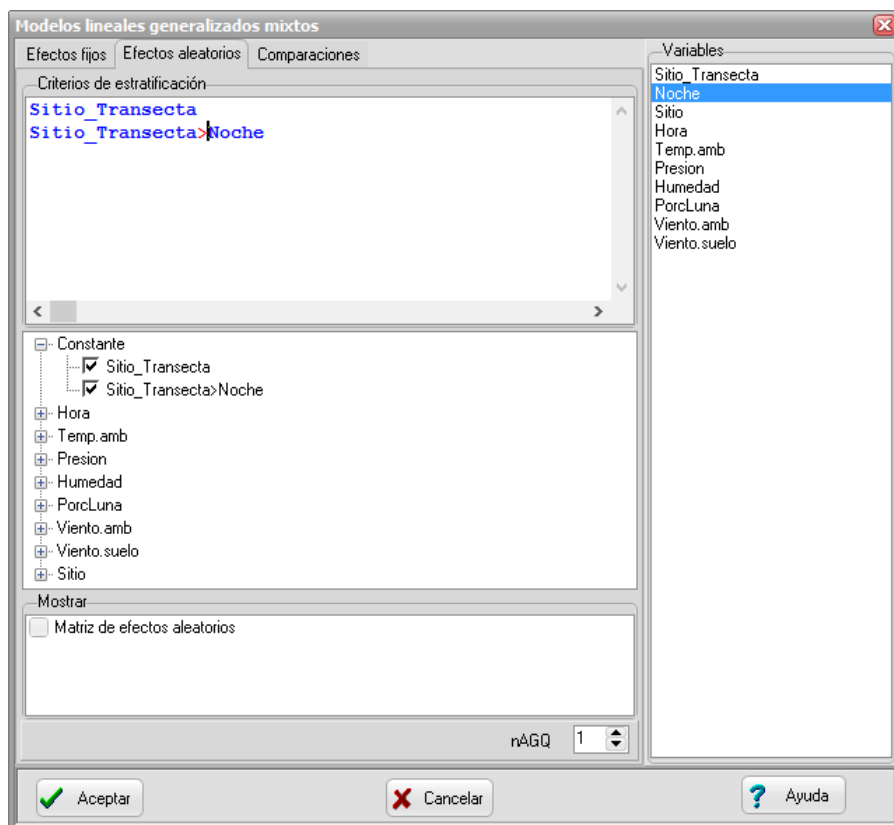


Figura 30: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorio desplegada y declaración del efecto de transecta y de noche dentro de transecta para el análisis de los datos del archivo *Escorpiones.IDB2*.

La salida correspondiente a las especificaciones anteriores se presenta a continuación:

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.029_Total_ML<-
glmer(Total~1+Hora+Temp.amb+Presion+Humedad+PorcLuna+Viento.amb+Viento.suel
o+Sitio+(1|Sitio_Transecta)+(1|Sitio_Transecta:Noche)
,family=myFamily
,na.action=na.omit
,REML=F
,AGQ=1
,offset=
,data=R.data15)
```

Resultados para el modelo: mlgm.modelo.029_Total_ML

Variable dependiente: Total

General

Familia	Enlace	nAGQ
poisson	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
210	674.94	711.75	-326.47	205.62

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
Hora	2.89	1	0.0893
Temp.amb	37.43	1	<0.0001
Presion	0.14	1	0.7051
Humedad	0.26	1	0.6095
PorcLuna	29.42	1	<0.0001
Viento.amb	0.02	1	0.8783
Viento.suelo	0.38	1	0.5398
Sitio	8.90	1	0.0028

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-17.99	3.23	-5.57	<0.0001
Hora	0.08	0.12	0.66	0.5123
Temp.amb	0.06	0.04	1.55	0.1200
Presion	0.01	1.1E-03	13.48	<0.0001
Humedad	0.02	0.01	2.43	0.0150
PorcLuna	-0.01	1.9E-03	-6.20	<0.0001
Viento.amb	0.01	0.02	0.50	0.6147
Viento.suelo	-0.02	0.04	-0.46	0.6459
SitioSecundario	-0.58	0.19	-3.09	0.0020

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Sitio_Transecta:Noche	(Intercept)	0.01	0.07
Sitio_Transecta	(Intercept)	0.13	0.36

Luego, se evaluó un modelo con efectos aleatorios de transecta únicamente (eliminando el posible efecto noche):

Modelos lineales generalizados mixtos**Especificación del modelo en R**

```
mlgm.modelo.030_Total_ML<-
glmer (Total~1+Hora+Temp.amb+Presion+Humedad+PorcLuna+Viento.amb+Viento.suel
o+Sitio+(1|Sitio_Transecta)
,family=myFamily
,na.action=na.omit
,REML=F
,AGQ=1
,offset=
,data=R.data15)
```

Resultados para el modelo: `mlgm.modelo.030_Total_ML`

Variable dependiente: *Total*

General

Familia	Enlace	nAGQ
poisson	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
210	672.95	706.42	-326.47	208.92

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
Hora	3.45	1	0.0631
Temp.amb	45.68	1	<0.0001
Presion	0.24	1	0.6250
Humedad	0.35	1	0.5562
PorcLuna	31.63	1	<0.0001
Viento.amb	0.02	1	0.8783
Viento.suelo	0.38	1	0.5398
Sitio	8.89	1	0.0029

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-18.03	3.22	-5.59	<0.0001
Hora	0.08	0.12	0.66	0.5091
Temp.amb	0.06	0.04	1.56	0.1188
Presion	0.01	1.1E-03	13.23	<0.0001
Humedad	0.02	0.01	2.44	0.0146
PorcLuna	-0.01	1.8E-03	-6.24	<0.0001
Viento.amb	0.01	0.02	0.51	0.6086
Viento.suelo	-0.02	0.04	-0.47	0.6390
SitioSecundario	-0.58	0.19	-3.09	0.0020

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Sitio Transecta	(Intercept)	0.13	0.36

Este último modelo tiene un valor de AIC (y BIC) menor que el anterior (AIC=672.95; BIC=706.42 vs AIC=674.94; BIC=711.75), indicando que en este caso no es necesario el efecto de noche dentro de transecta.

Debido a que este es un estudio observacional, se realizó un proceso de selección de regresoras que no resultaron importante como predictoras para lograr un modelo más parsimonioso. Para esto se usó selección hacia atrás (*Backwards selection*). Se usaron los p-valores asociados a los coeficientes de los efectos fijos, que indican el efecto marginal de cada una de las covariables, que como entraron todas como lineales (en la escala logarítmica), tienen un grado de libertad cada una. Estas pruebas (Wald test) no coinciden necesariamente con la pruebas secuenciales (cociente de verosimilitud). Los resultados obtenidos son los siguientes:

Modelos lineales generalizados mixtos**Especificación del modelo en R**

```
mlgm.modelo.031_Total_ML<-
glmer(Total~1+Humedad+PorcLuna+Sitio+(1|Sitio_Transecta)
,family=myFamily
,na.action=na.omit
,REML=F
,AGQ=1
,offset=
,data=R.data15)
```

Resultados para el modelo: mlgm.modelo.031_Total_ML

Variable dependiente: Total

General

Familia	Enlace	nAGQ
poisson	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
210	665.60	682.34	-327.80	211.26

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
Humedad	10.47	1	0.0012
PorcLuna	65.02	1	<0.0001
Sitio	12.50	1	0.0004

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.48	0.25	1.88	0.0602
Humedad	0.02	0.01	3.57	0.0004
PorcLuna	-0.01	1.6E-03	-8.13	<0.0001
SitioSecundario	-0.67	0.18	-3.83	0.0001

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Sitio Transecta	(Intercept)	0.13	0.36

Total - Medias ajustadas y errores estándares para Sitio

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Estimaciones para la siguiente combinación de valores de las covariables

Covariable Valor

Humedad Media

PorcLuna Media

Procedimiento de corrección de p-valores: No

Sitio	PredLin	E.E.	Media	E.E.	
Maduro	0.67	0.12	1.95	0.23	A
Secundario	-1.8E-03	0.14	1.00	0.14	B

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

El sitio correspondiente al bosque maduro presentó un mayor promedio de conteo de escorpiones. Las variables ambientales que más explican la presencia de escorpiones son el porcentaje de luna, la presión y la humedad relativa.

Análisis de recuentos en un diseño de parcelas divididas

Se realizó un ensayo para evaluar la efectividad de tres herbicidas (A, B, C) con y sin riego en un cultivo de papa. El diseño del experimento fue parcelas divididas, con cada repetición del experimento en una localidad diferente (5 localidades). El factor riego (sí / no) se aleatorizó a nivel de parcela completa, y el factor herbicida (A, B o C) se aleatorizó a nivel de subparcela. La variable respuesta fue la cantidad de malezas en un marco de 0.25 m². Los datos se encuentran en el archivo EvaluacionHerbicidas.IDB2.

Para hacer el análisis se declara la variable respuesta recuento y las variables de clasificación (Figura 31).

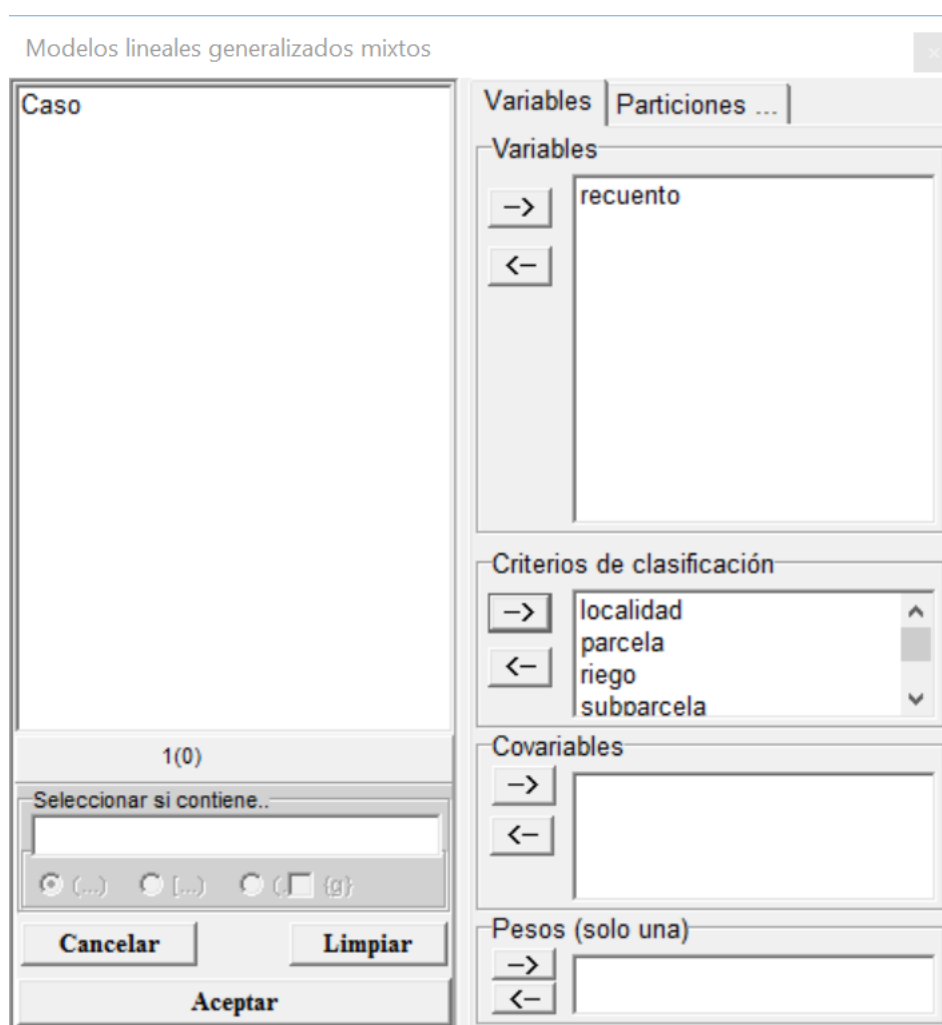


Figura 31: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo EvaluacionHerbicidas.IDB2.

Especificación del modelo en R

```
mlgm.modelo.005_recuento_ML<-
glmer(recuento~1+riego+herbicida+riego:herbicida+(1|localidad)+(1|parcela)
,family=myFamily
,na.action=na.omit
,REML=F
,AGQ=1
,offset=NULL
,data=R.data05)
```

Resultados para el modelo: mlgm.modelo.005_recuento_ML

Variable dependiente: *recuento*

General

Familia	Enlace	nAGQ
poisson	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
30	216.38	227.59	-100.19	61.27

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
riego	12.22	1	0.0005
herbicida	80.30	2	<0.0001
riego:herbicida	29.85	2	<0.0001

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.36	0.18	19.09	<0.0001
riegoSI	-0.75	0.27	-2.81	0.0049
herbicidaB	-0.63	0.14	-4.52	<0.0001
herbicidaC	-0.59	0.14	-4.30	<0.0001
riegoSI:herbicidaB	-1.12	0.33	-3.42	0.0006
riegoSI:herbicidaC	-1.65	0.39	-4.18	<0.0001

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
parcela	(Intercept)	0.12	0.35
localidad	(Intercept)	9.3E-10	3.1E-05

recuento - Medias ajustadas y errores estándares para riego

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

riego	PredLin	E.E.	Media	E.E.	
NO	2.96	0.17	19.23	3.21	A
SI	1.28	0.22	3.61	0.81	B

Medias con una letra común no son significativamente diferentes ($p > 0.05$)**recuento - Medias ajustadas y errores estándares para herbicida**

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

herbicida	PredLin	E.E.	Media	E.E.	
A	2.99	0.13	19.86	2.65	A
B	1.80	0.19	6.04	1.13	B
C	1.57	0.22	4.83	1.04	B

Medias con una letra común no son significativamente diferentes ($p > 0.05$)**recuento - Medias ajustadas y errores estándares para riego*herbicida**

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

riego	herbicida	PredLin	E.E.	Media	E.E.	
NO	A	3.36	0.18	28.87	5.09	A
NO	C	2.77	0.19	15.99	3.05	B
NO	B	2.73	0.19	15.41	2.96	B
SI	A	2.62	0.20	13.67	2.73	B
SI	B	0.86	0.32	2.37	0.76	C
SI	C	0.38	0.39	1.46	0.56	C

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

En este modelo Poisson con la parte aleatoria contemplando la estructura de parcelas divididas del diseño, se presenta una sobredispersión, ya que la deviance de 61.27 y los GL residuales de la parte fija son 24, dando un cociente de 2.55.

Otra alternativa para modelar la sobredispersión es el uso de la Binomial negativa. En este caso solo declaramos en la parte aleatorio la Localidad y las parcelas principales dentro de localidad para contemplar la estructura del diseño en parcelas divididas usado.

Especificación del modelo en R

```
mlgm.modelo.007_recuento_ML<-
glmer.nb(recuento~1+riego+herbicida+riego:herbicida+(1|localidad)+(1|parcela)
,link=myFamily$link
,na.action=na.omit
,REML=F
,AGQ=1
,offset=NULL
,data=R.data05)
```

Resultados para el modelo: mlgm.modelo.007_recuento_ML

Variable dependiente: *recuento*

General

Familia	Enlace	nAGQ
Negative Binomial (5.7414)	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
30	197.71	210.32	-89.85	28.55

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
riego	12.27	1	0.0005
herbicida	33.37	2	<0.0001
riego:herbicida	11.07	2	0.0039

Efectos fijos

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.38	0.22	15.27	<0.0001
riegoSI	-0.71	0.33	-2.17	0.0303
herbicidaB	-0.64	0.30	-2.14	0.0327
herbicidaC	-0.58	0.30	-1.93	0.0533
riegoSI:herbicidaB	-1.12	0.50	-2.24	0.0254
riegoSI:herbicidaC	-1.64	0.55	-2.99	0.0028

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
parcela	(Intercept)	0.03	0.18

localidad (Intercept) 5.1E-11 7.1E-06

recuento - Medias ajustadas y errores estándares para riego

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

riego	PredLin	E.E.	Media	E.E.	
NO	2.97	0.15	19.48	2.90	A
SI	1.34	0.22	3.83	0.83	B

Medias con una letra común no son significativamente diferentes (p > 0.05)

recuento - Medias ajustadas y errores estándares para herbicida

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

herbicida	PredLin	E.E.	Media	E.E.	
A	3.02	0.17	20.55	3.42	A
B	1.82	0.22	6.18	1.34	B
C	1.62	0.24	5.07	1.20	B

Medias con una letra común no son significativamente diferentes (p > 0.05)

recuento - Medias ajustadas y errores estándares para riego*herbicida

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

riego	herbicida	PredLin	E.E.	Media	E.E.	
NO	A	3.38	0.22	29.28	6.48	A
NO	C	2.80	0.23	16.37	3.81	A B
NO	B	2.74	0.24	15.42	3.65	B
SI	A	2.67	0.24	14.42	3.53	B
SI	B	0.91	0.36	2.48	0.88	C
SI	C	0.45	0.41	1.57	0.64	C

Medias con una letra común no son significativamente diferentes (p > 0.05)

Este modelo presenta un AIC y BIC menores al Poisson y un cociente de deviance cercano a 1 (28.55/24=1.18). El parámetro de agregación k es de 5.7414; a medida que este parámetro aumenta la distribución se aproxima más a la Poisson (la sobredispersión disminuye). Esto puede apreciarse a partir de la relación media/varianza para la distribución binomial negativa ($\text{var} = \mu + \mu^2/k$).

Con este modelo Binomial negativo, el análisis detecto una interacción significativa entre riego y herbicida ($p=0.0039$). Para estudiar esta interacción se realizó un gráfico de puntos de los recuentos medios (Figura 32). En este gráfico se pueden ver las letras indicando las diferencias significativas entre

medias ($p=0.05$) usando la prueba LSD de Fisher. Al observar el comportamiento del perfil de respuesta con los diferentes herbicidas y el riego, aparentemente no se aprecia una interacción importante. Sin embargo, con riego hay diferencias entre herbicida A y herbicida C, mientras que sin riego estas diferencias no se detectan, siendo esto lo que provoca la interacción significativa.

Si se observa el mismo gráfico, pero ahora en la escala del predictor lineal (Figura 33), ahora la presencia de interacción es más marcada. Debemos recordar que en esta escala (log) es en la que se define la interacción. Por ejemplo, en la escala log (predictor lineal), la diferencia en la media sin y con riego del herbicida A es $3.38 - 2.67 = 0.71$, mientras que la misma diferencia para el herbicida B es $2.74 - 0.91 = 1.83$. Si exponenciamos cada media, esta diferencia se transforma en un cociente, y podemos ver que pasar de sin riego a con riego con el herbicida A, la media disminuye de 29.28 a 14.42, una disminución del 51%. Para el herbicida B, la media disminuye de 15.42 a 2.48, una disminución del 84%. Estos mismos cocientes $29.28/14.42$ y $15.42/2.48$ también puede obtenerse haciendo $\exp(0.71)$ y $\exp(1.83)$, es decir exponenciando la diferencias correspondientes en la escala log. Se debe destacar que el enlace logarítmico significa que las comparaciones de medias se están haciendo en esta escala, y cuando pasamos a la escala original de los recuentos, estas comparaciones ya no son diferencias sino cocientes, por lo que deben expresarse porcentualmente. Para las mismas combinaciones de tratamientos, si realizáramos las diferencias entre las medias en la escala de los recuentos observaríamos 14.86 (sin vs. Con riego en herbicida A) y 12.94 (sin vs. Con riego en herbicida B). Estas son las diferencias que observamos en el gráfico de medias en la escala de recuentos.

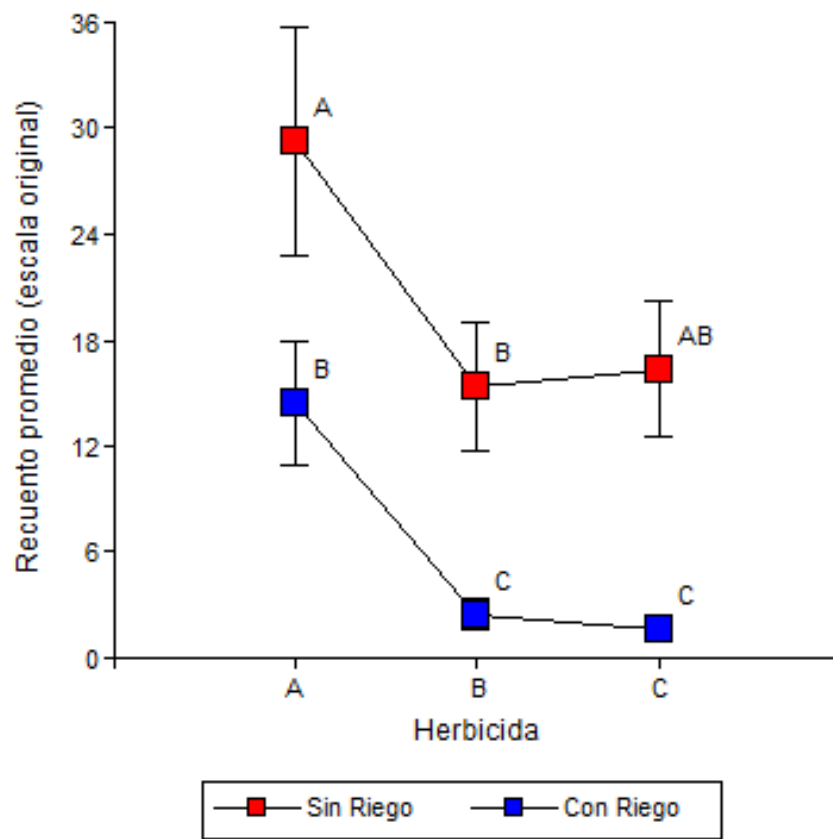


Figura 32: Gráficos de puntos de Recuentos promedio para los niveles del factor Herbicida y los niveles del factor Riego.

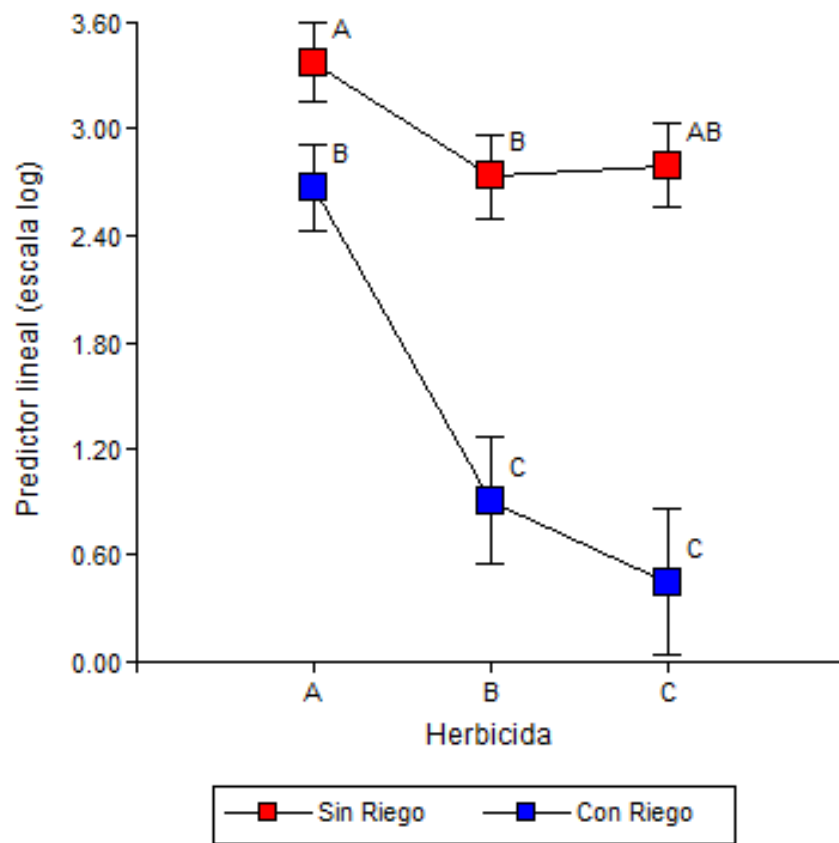


Figura 33: Gráficos de puntos de Predictor lineal para los niveles del factor *Herbicida* y los niveles del factor *Riego*.

Análisis de la incidencia de una enfermedad en un diseño alfa latice

El objetivo de este trabajo fue la fenotipificación de una población de mapeo de 160 líneas endocriadas de girasol para la búsqueda de resistencia a la marchitez causada por *Verticillium dahliae* (Verticilosis) en girasol. Los datos son parte de la tesis doctoral del Ing.Agr. Juan Francisco Montecchia, sobre la Caracterización Genómica de fuentes de resistencia a la marchitez causada por *Verticillium dahliae* (Verticilosis) en girasol (EEA-INTA Balcarce; Grupo de Genómica de girasol de INTA Castelar y Grupo de Patología Vegetal de la EEA-Balcarce). Los datos corresponden dos ensayos realizados en los años (2015 y 2016) en las que se realizó un diseño Alfa Lattice con dos repeticiones y control del efecto fila y columna dentro de cada repetición. Los datos se encuentran en el archivo `AlfaLatticeFilaColumnaBinomial.IDB2`.

Cada genotipo se evaluó en una micro-parcela de alrededor de 20 plantas con espaciamiento de 0.7m. Parcelas con menos de 5 plantas logradas fueron excluidas del ensayo. En el resto de las parcelas, todas las plantas logradas (*plantas expuestas* en el archivo de datos) fueron expuestas a la infección y fueron oportunamente evaluadas para detectar la presencia de síntomas de la enfermedad (*infectadas* en el archivo de datos). El propósito de este análisis es estimar la incidencia de la enfermedad para cada genotipo. Para ello ajustaremos un modelo binomial mixto (Figura 34), con efecto fijo de genotipo y efectos aleatorios de año, repetición (dentro de año), fila (dentro de repetición-año) y columna (dentro de repetición-año).

En la solapa *Efectos fijos* se debe declarar a *genotipo*, y debido a la naturaleza de la variable respuesta pedimos un ajuste con una familia *Binomial*, con enlace *logit* y declaramos la variable *Plantas expuestas* para indicar el número de plantas que se observan en cada parcela, que como habíamos adelantado no es el mismo para todas las parcelas (Figura 35).

En el módulo de modelos lineales generalizados mixtos todos los factores, no importando el orden en que se ingresan con considerados como efectos aleatorios cruzados. Sin embargo, en nuestro ejemplo hay factores aleatorios anidados y cruzados dentro del nivel de repeticiones-años (fijas y columnas). La forma más fácil de especificar esta estructura de

efectos aleatorios es codificar cada factor aleatorio de manera tal que la nueva codificación contenga la estructura de anidamiento y cruzamiento. Esta codificación ya se ha realizado en el archivo de datos. Por ejemplo, para indicar que repetición esta anidada en año, se combinaron los identificadores de año con los identificadores de repetición para producir la variable $Rep(año)$, mientras que para generar el código de filas anidadas dentro de repeticiones y años, se combinaron los identificadores de $Rep(año)$ con los identificadores de filas (de cada repetición) para lograr lo que se llama $Fila(Rep(año))$ en el archivo de datos (Figura 36).

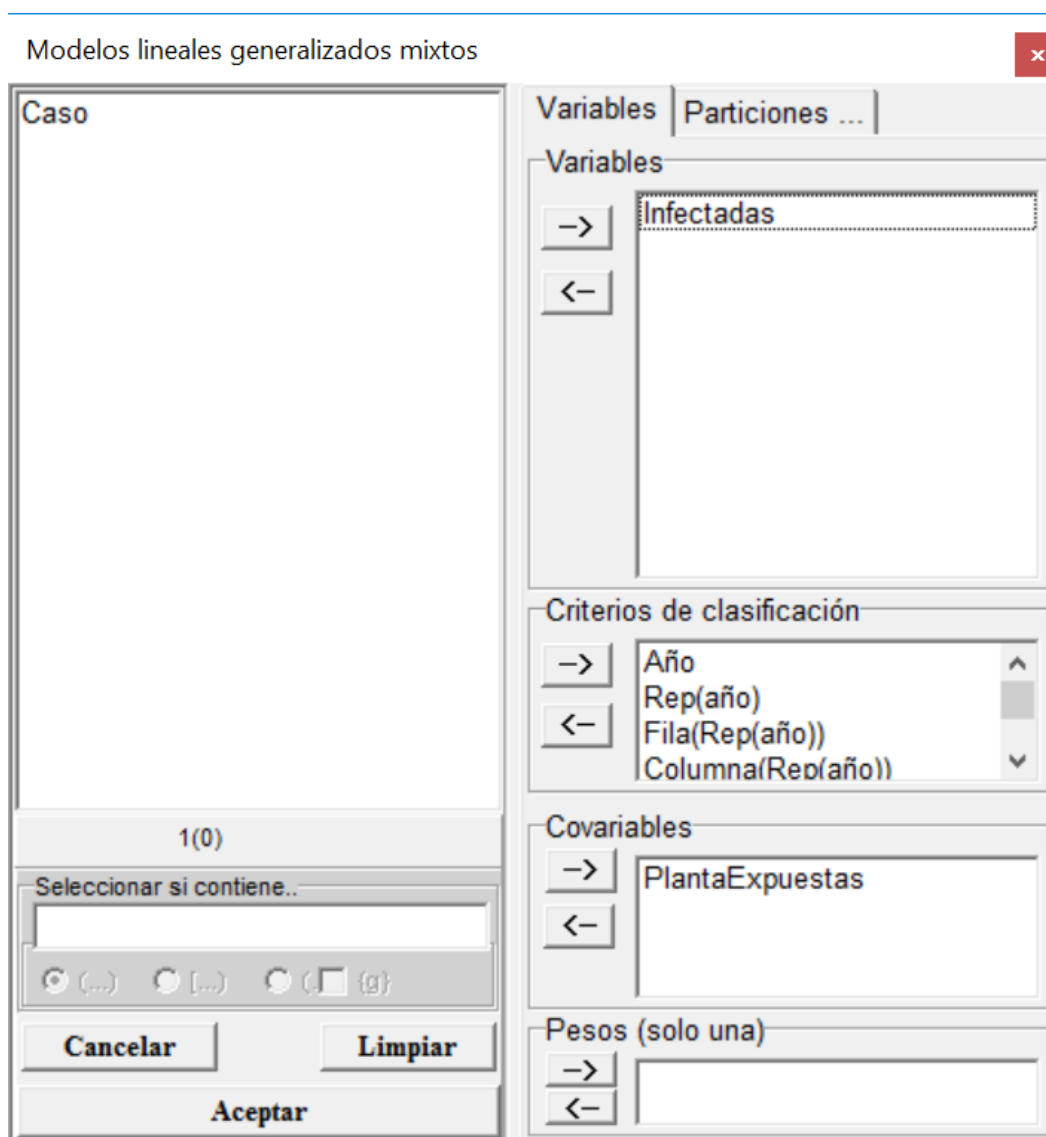


Figura 34: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo *AlfaLaticeFilaColumnaBinomial.IDB2*.

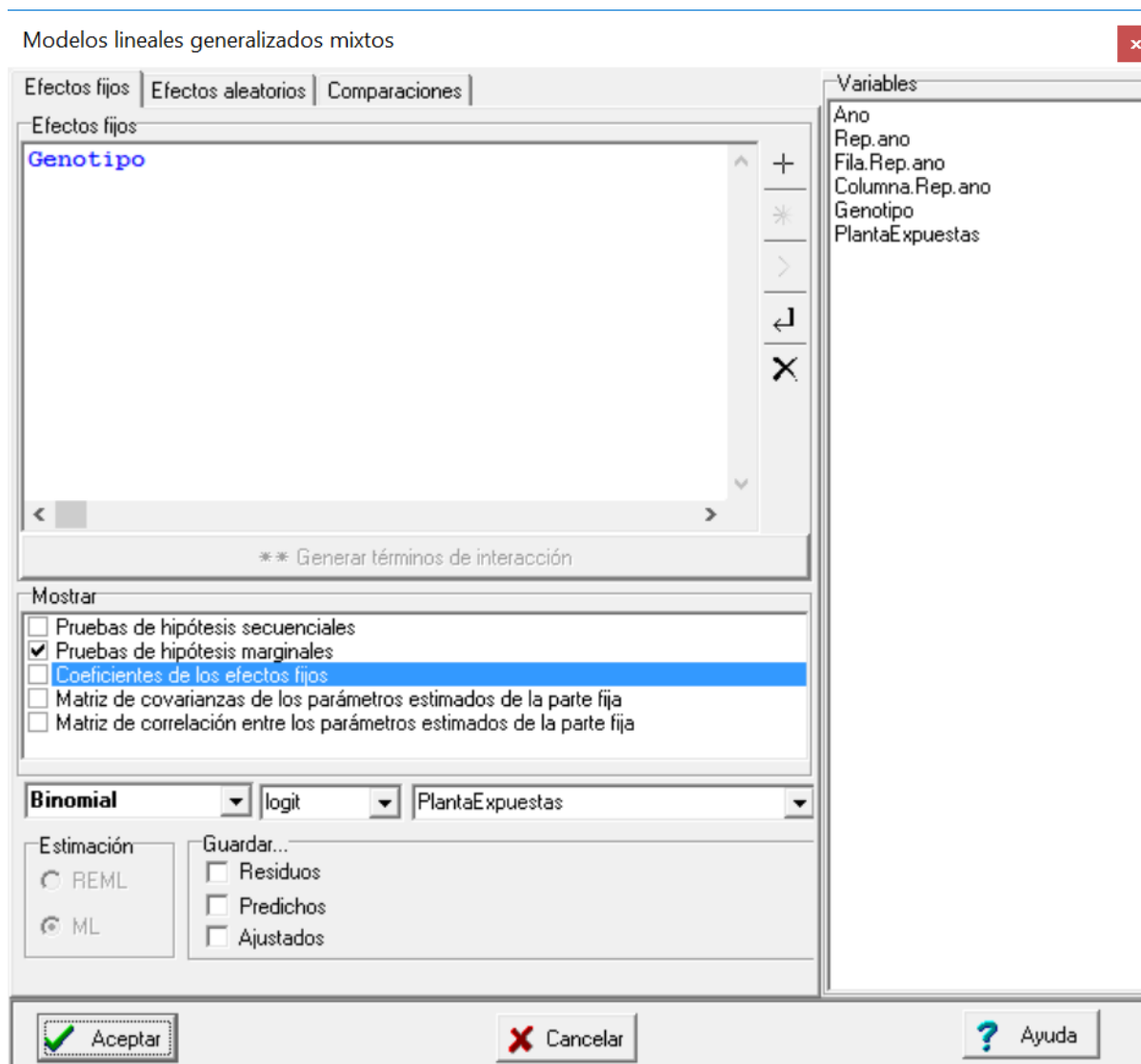


Figura 35: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada con la declaración de efectos fijos de Genotipo, usando familia Binomial, función de enlace logit y la identificación del número de ensayos (PlantasExpuestas) para el análisis de los datos del archivo AlfaLatticeFilaColumnaBinomial.IDB2.

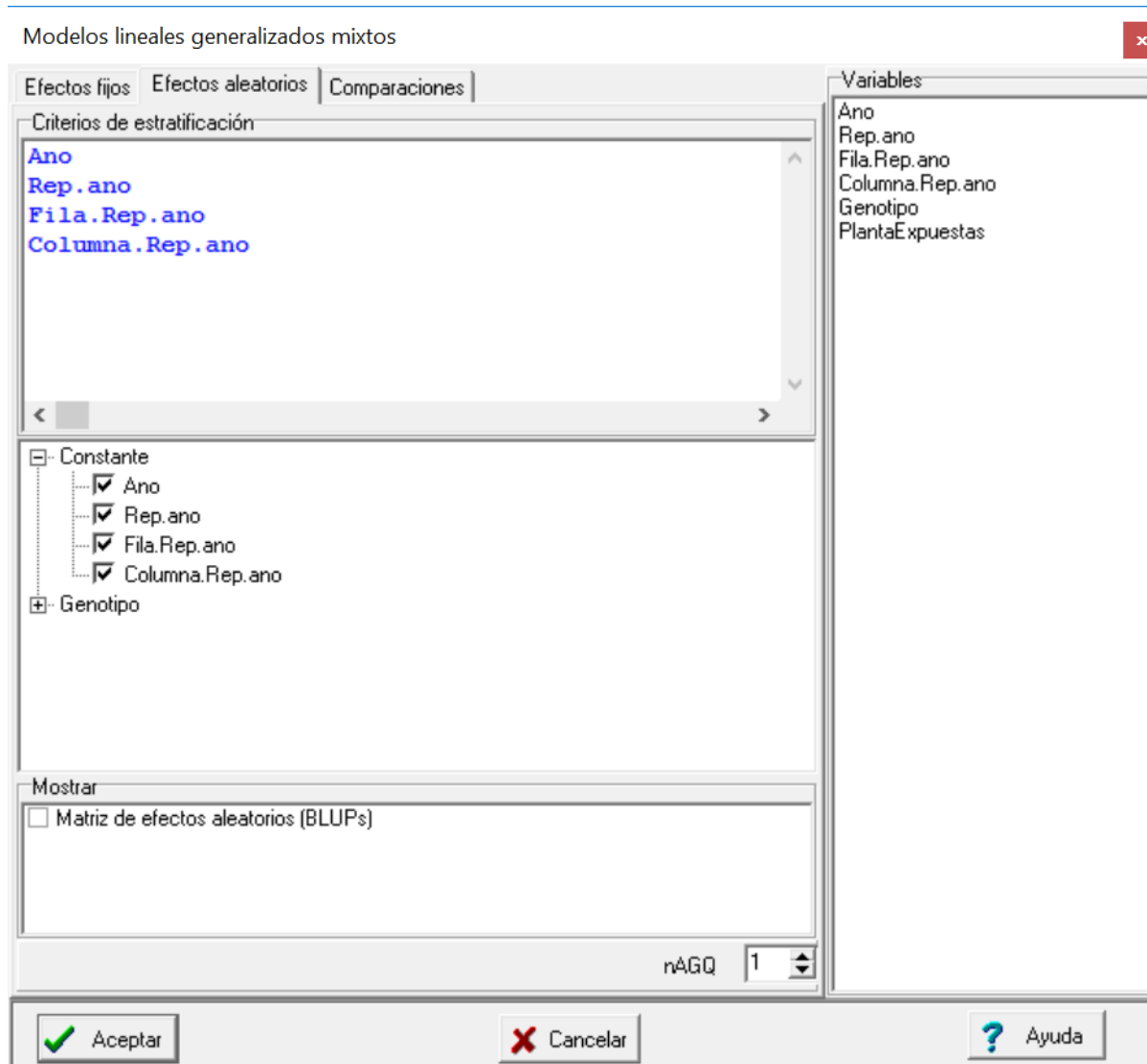


Figura 36: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada con la declaración de efectos de Año, Rep dentro de Año, fila dentro de Rep dentro de Año y Columna dentro de Rep dentro de Año para el análisis de los datos del archivo AlfaLaticeFilaColumnaBinomial.IDB2.

A l aceptar las especificaciones anteriores se obtendrá la siguiente salida:

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.010_Infectadas_ML<-glmer(cbind(Infectadas
,as.numeric(as.character(PlantaExpuestas)) -
Infectadas)~1+Genotipo+(1|Año)+(1|Rep.año)+(1|Fila.Rep.año)+(1|Columna.Rep.
año)
```



```
,family=myFamily
,na.action=na.omit
,REML=F
,AGQ=1
,data=R.data10)
```

Resultados para el modelo: `mlgm.modelo.010_Infectadas_ML`

Variable dependiente: *Infectadas*

General

Familia	Enlace	nAGQ
binomial	logit	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
659	2499.76	3218.28	-1089.88	693.80

AIC y BIC menores implica mejor

Pruebas de hipótesis marginales (Wald) para los efectos fijos

Source	numDF	denDF	F-value	p-value
Genotipo	155	503	5.78	<0.0001

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Columna.Rep.ano	(Intercept)	0.25	0.50
Fila.Rep.ano	(Intercept)	0.36	0.60
Rep.ano	(Intercept)	0.58	0.76
Ano	(Intercept)	0.80	0.89

Infectadas - Medias ajustadas y errores estándares para Genotipo

Inversa de la función de enlace con efecto aleatorio=0

DGC (Alfa=0.05)

Genotipo	PredLin	E.E.	Media	E.E.	
28	1.48	0.86	0.81	0.13	A
8	0.83	0.95	0.70	0.20	B
99	0.73	1.05	0.68	0.23	B
30	0.62	0.85	0.65	0.19	B
128	0.59	0.88	0.64	0.20	B
114	0.54	0.85	0.63	0.20	B
10	0.53	0.89	0.63	0.21	B
26	0.52	0.85	0.63	0.20	B
137	0.26	0.89	0.56	0.22	B
142	0.23	0.98	0.56	0.24	B
111	0.19	0.85	0.55	0.21	B
5	0.18	0.84	0.55	0.21	B

Modelos Lineales Generalizados Mixtos en InfoStat

86	0.14	0.87	0.53	0.22	B
33	0.12	0.84	0.53	0.21	B
66	0.06	0.87	0.51	0.22	B
109	-0.02	0.85	0.50	0.21	B
133	-0.09	1.09	0.48	0.27	B
143	-0.10	0.83	0.48	0.21	B
55	-0.15	0.83	0.46	0.21	B
87	-0.15	0.84	0.46	0.21	B
151	-0.21	0.98	0.45	0.24	B
38	-0.21	0.83	0.45	0.21	B
94	-0.24	0.88	0.44	0.22	B
57	-0.30	0.98	0.43	0.24	B
140	-0.30	0.89	0.43	0.22	B
88	-0.32	0.84	0.42	0.21	B
76	-0.35	0.85	0.41	0.21	B
152	-0.40	0.90	0.40	0.22	B
41	-0.41	0.83	0.40	0.20	B
156	-0.47	0.86	0.39	0.20	B
113	-0.48	0.84	0.38	0.20	B
150	-0.54	0.76	0.37	0.18	B
64	-0.54	0.85	0.37	0.20	B
1	-0.60	0.84	0.36	0.19	B
44	-0.61	0.88	0.35	0.20	B
115	-0.62	0.82	0.35	0.19	B
12	-0.63	0.88	0.35	0.20	B
19	-0.63	0.84	0.35	0.19	B
127	-0.66	0.91	0.34	0.21	B
13	-0.68	0.90	0.34	0.20	B
106	-0.68	0.85	0.34	0.19	B
139	-0.77	0.86	0.32	0.19	B
75	-0.78	0.84	0.31	0.18	B
32	-0.86	0.88	0.30	0.18	B
147	-0.88	0.86	0.29	0.18	B
18	-0.90	0.86	0.29	0.18	B
24	-0.92	0.86	0.29	0.18	B
144	-0.93	0.84	0.28	0.17	B
31	-0.95	0.83	0.28	0.17	B
48	-1.02	0.88	0.26	0.17	B
120	-1.03	0.84	0.26	0.16	B
119	-1.09	0.87	0.25	0.16	B
15	-1.11	0.93	0.25	0.17	B
17	-1.12	0.86	0.25	0.16	B
93	-1.13	0.87	0.24	0.16	B
49	-1.14	0.83	0.24	0.15	B
39	-1.20	0.83	0.23	0.15	B
96	-1.23	0.92	0.23	0.16	B
36	-1.24	0.92	0.22	0.16	B
118	-1.24	0.87	0.22	0.15	B
2	-1.27	0.93	0.22	0.16	B
77	-1.28	0.85	0.22	0.15	B
52	-1.33	0.85	0.21	0.14	B
20	-1.35	0.87	0.21	0.14	B
45	-1.36	0.84	0.20	0.14	B
11	-1.39	0.98	0.20	0.16	B
50	-1.39	0.80	0.20	0.13	B
92	-1.39	0.86	0.20	0.14	B
95	-1.41	1.13	0.20	0.18	B
34	-1.48	0.95	0.19	0.14	B

Modelos Lineales Generalizados Mixtos en InfoStat

122	-1.50	0.85	0.18	0.13	B	
46	-1.52	1.00	0.18	0.15	B	
145	-1.53	0.86	0.18	0.12	B	
29	-1.56	0.89	0.17	0.13	B	
105	-1.56	0.84	0.17	0.12	B	
97	-1.61	0.86	0.17	0.12	B	
23	-1.61	0.85	0.17	0.12	B	
22	-1.64	0.85	0.16	0.11	B	
84	-1.67	0.98	0.16	0.13	B	
7	-1.67	0.88	0.16	0.12	B	
134	-1.67	0.88	0.16	0.12	B	
63	-1.69	0.98	0.16	0.13	B	
53	-1.70	0.86	0.16	0.11	B	
79	-1.72	0.88	0.15	0.11	B	
6	-1.73	0.92	0.15	0.12	B	
78	-1.74	0.89	0.15	0.11	B	
71	-1.78	0.87	0.14	0.11	B	
27	-1.80	0.95	0.14	0.12	B	
59	-1.81	0.85	0.14	0.10	B	
21	-1.85	0.86	0.14	0.10	B	
35	-1.88	1.01	0.13	0.12	B	
124	-1.88	0.87	0.13	0.10	B	
47	-1.97	0.85	0.12	0.09		C
70	-1.98	0.85	0.12	0.09		C
68	-1.98	0.91	0.12	0.10		C
4	-2.03	0.85	0.12	0.09		C
138	-2.08	0.90	0.11	0.09		C
81	-2.08	0.91	0.11	0.09		C
98	-2.13	0.85	0.11	0.08		C
126	-2.25	0.90	0.09	0.08		C
103	-2.30	0.93	0.09	0.08		C
60	-2.32	1.00	0.09	0.08		C
72	-2.32	0.97	0.09	0.08		C
42	-2.33	0.90	0.09	0.07		C
73	-2.35	0.91	0.09	0.07		C
43	-2.36	0.95	0.09	0.08		C
25	-2.36	0.91	0.09	0.07		C
101	-2.42	1.00	0.08	0.07		C
125	-2.42	1.02	0.08	0.08		C
154	-2.44	0.94	0.08	0.07		C
141	-2.48	0.92	0.08	0.07		C
117	-2.56	0.90	0.07	0.06		C
153	-2.57	0.89	0.07	0.06		C
129	-2.57	0.95	0.07	0.06		C
104	-2.66	1.11	0.07	0.07		C
148	-2.68	0.95	0.06	0.06		C
136	-2.68	0.96	0.06	0.06		C
14	-2.68	1.08	0.06	0.07		C
116	-2.75	0.98	0.06	0.06		C
85	-2.75	0.89	0.06	0.05		C
62	-2.77	0.92	0.06	0.05		C
131	-2.83	0.85	0.06	0.04		C
74	-2.87	0.91	0.05	0.05		C
3	-2.90	1.08	0.05	0.05		C
112	-2.91	1.02	0.05	0.05		C
102	-2.92	1.02	0.05	0.05		C
54	-2.94	1.00	0.05	0.05		C
100	-3.05	0.98	0.05	0.04		C

58	-3.17	1.01	0.04	0.04	C
149	-3.20	1.00	0.04	0.04	C
56	-3.23	0.95	0.04	0.03	C
110	-3.24	1.04	0.04	0.04	C
90	-3.26	0.90	0.04	0.03	C
89	-3.27	1.09	0.04	0.04	C
135	-3.31	0.95	0.04	0.03	C
37	-3.32	0.77	0.03	0.03	C
40	-3.32	1.09	0.03	0.04	C
123	-3.38	0.92	0.03	0.03	C
83	-3.48	0.99	0.03	0.03	C
130	-3.48	1.07	0.03	0.03	C
82	-3.56	0.99	0.03	0.03	C
61	-3.57	0.92	0.03	0.02	C
51	-3.61	1.08	0.03	0.03	C
107	-3.64	0.96	0.03	0.02	C
67	-3.65	1.06	0.03	0.03	C
155	-3.71	0.95	0.02	0.02	C
69	-3.74	1.11	0.02	0.03	C
121	-3.79	0.99	0.02	0.02	C
91	-3.83	1.10	0.02	0.02	C
132	-3.85	1.08	0.02	0.02	C
108	-3.95	1.06	0.02	0.02	C
16	-3.97	1.08	0.02	0.02	C
9	-4.02	1.06	0.02	0.02	C
146	-4.12	1.07	0.02	0.02	C
80	-4.28	1.06	0.01	0.01	C
65	-4.44	1.08	0.01	0.01	C

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

El modelo ajustado no presenta problemas de sobredispersión. La deviance sobre los grados de libertad residuales da un valor de $693/503 = 1.38$. Se encontraron efectos significativos de Genotipo ($p < 0.0001$) y la prueba DGC permitió diferenciar tres grupos, uno formado solo por el genotipo 28 que fue el de mayor incidencia estimada (81%). El segundo grupo con incidencias estimadas entre 70% y 10% y el grupo tercero con valores de incidencia entre 1% y 9% es de los genotipos recomendables con respecto a su resistencia a la enfermedad.

Usando la estimación de parámetros de los efectos aleatorios podemos observar que el efecto año, con una varianza de 0.80 es el más importante, seguido por el efecto de las repeticiones dentro de año con una varianza de 0.58. Como es de esperar en este tipo de diseños, los valores de varianza más bajos corresponden a los efectos aleatorios de fila y columna anidados en los años y las repeticiones.

Estimación de los componentes de varianza

En áreas como el mejoramiento genético animal o vegetal es de particular interés el cálculo de componentes de varianza. Estos son usados para obtener heredabilidades, respuestas a la selección, coeficientes de variabilidad genética aditiva, coeficientes de diferenciación genética, etc. Los modelos lineales generalizados mixtos pueden usarse para estimar los componentes de varianza para variables con distribución normal, por medio del estimador de máxima verosimilitud restringida (REML).

En muchos estudios de genética de poblaciones se trabaja con varias poblaciones que a su vez están representadas por uno o más individuos de distintas familias. En este caso se cuenta con dos factores en el modelo, las poblaciones y las familias dentro de cada población. Para ejemplificar el uso de componentes de varianza se usan los datos que se presentan en el archivo [Compvar.IDB2](#) (Navarro et ál. 2005). Estos datos provienen de un ensayo de siete poblaciones de cedro (*Cedrela odorata* L.) con un total de 115 familias y fueron usados para ejemplificar el uso de componentes de varianza en el tutorial Modelos Lineales Mixtos: Aplicaciones en InfoStat (Di Rienzo et al. 2011). Para algunas familias se cuenta con repeticiones y para otras no. Además, el número de familias dentro de cada población no es el mismo. Las variables registradas son el largo promedio de las semillas (largo), el diámetro, el largo del tallo y número de hojas de plantines de cedro.

Si consideramos a las poblaciones como una muestra aleatoria de un conjunto grande de poblaciones, entonces la inferencia estará orientada a este conjunto grande de poblaciones. El efecto de las poblaciones estudiadas es aleatorio, y el interés será la estimación de los componentes de varianza debida a poblaciones y a familias dentro de poblaciones. Los datos se encuentran en el archivo [Compvar.IDB2](#):

Primero se selecciona el menú *Estadísticas*, submenú *Modelos lineales generalizados mixtos* y escogemos *Estimación*. Al realizar esta selección aparecerá la ventana de selección de variables, donde especificamos como variables dependientes a *Largo* y como criterios de clasificación a *Población* y *Familia* (Figura 37). En la ventana de efectos fijos no ponemos ningún efectos, y seleccionamos la familia *Normal* y la función de enlace *Identidad* (Figura 38). A diferencia del caso de Modelos Lineales Generales y Mixtos, donde si declaramos diferentes efectos aleatorios por defectos estos quedan anidados en el orden en que los ingresamos, en el caso de los Modelos Lineales Generalizados Mixtos, los factores aleatorios siempre son cruzados. Para declarar entonces el efecto de la Familia dentro de la Población, estas dos variables se han combinado previamente al análisis. Para esto se procede a ir

al menú Datos, opción Cruzar variables para formar una nueva variable, y luego se declara Población y Familia, y al aceptar se crea la variable combinada. Otra opción es no combinarlas, y en el momento de declarar los efectos aleatorios, primero se coloca Población, y luego Población>Familia (Figura 39).

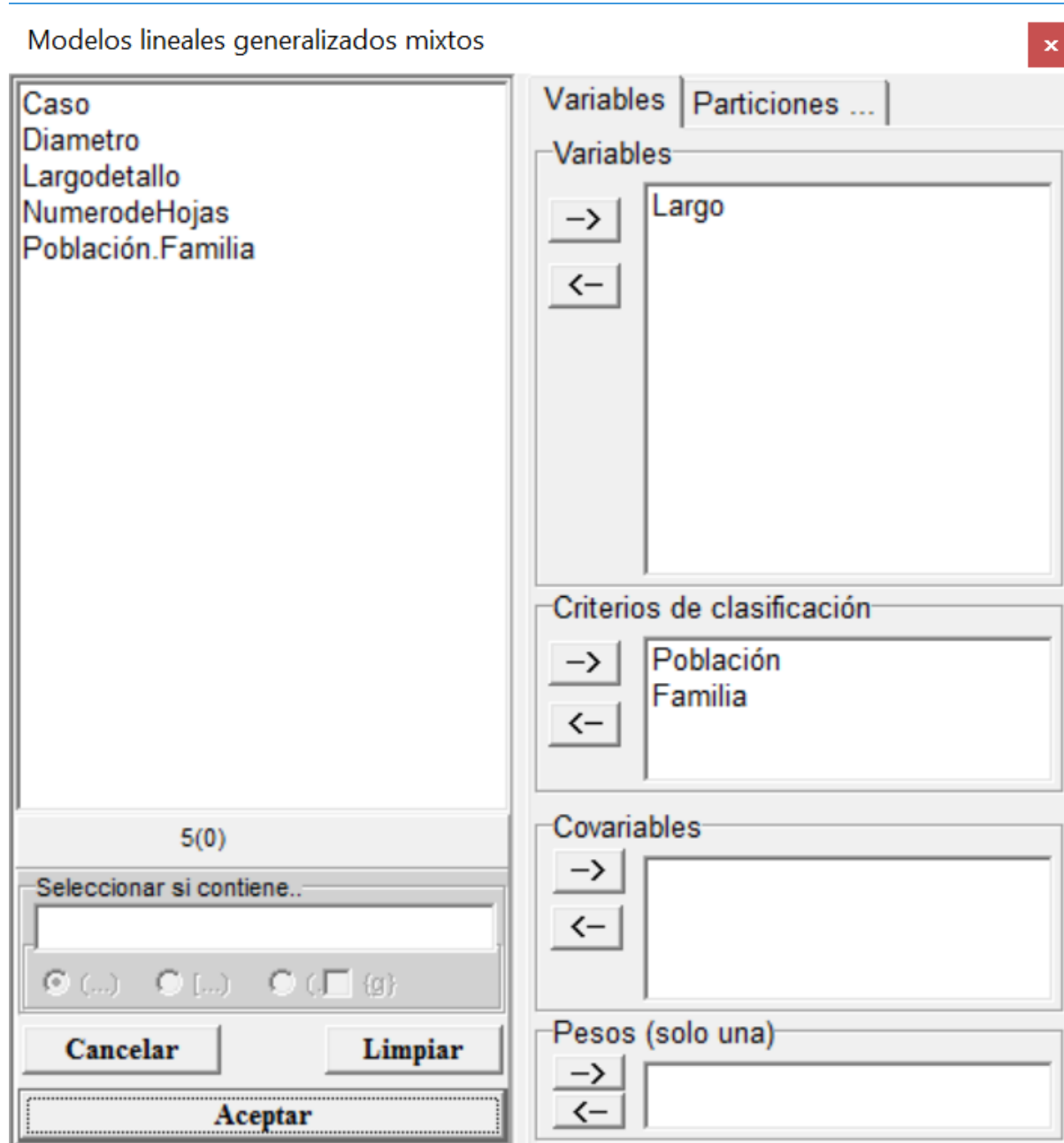


Figura 37: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo CompVar.IDB2.

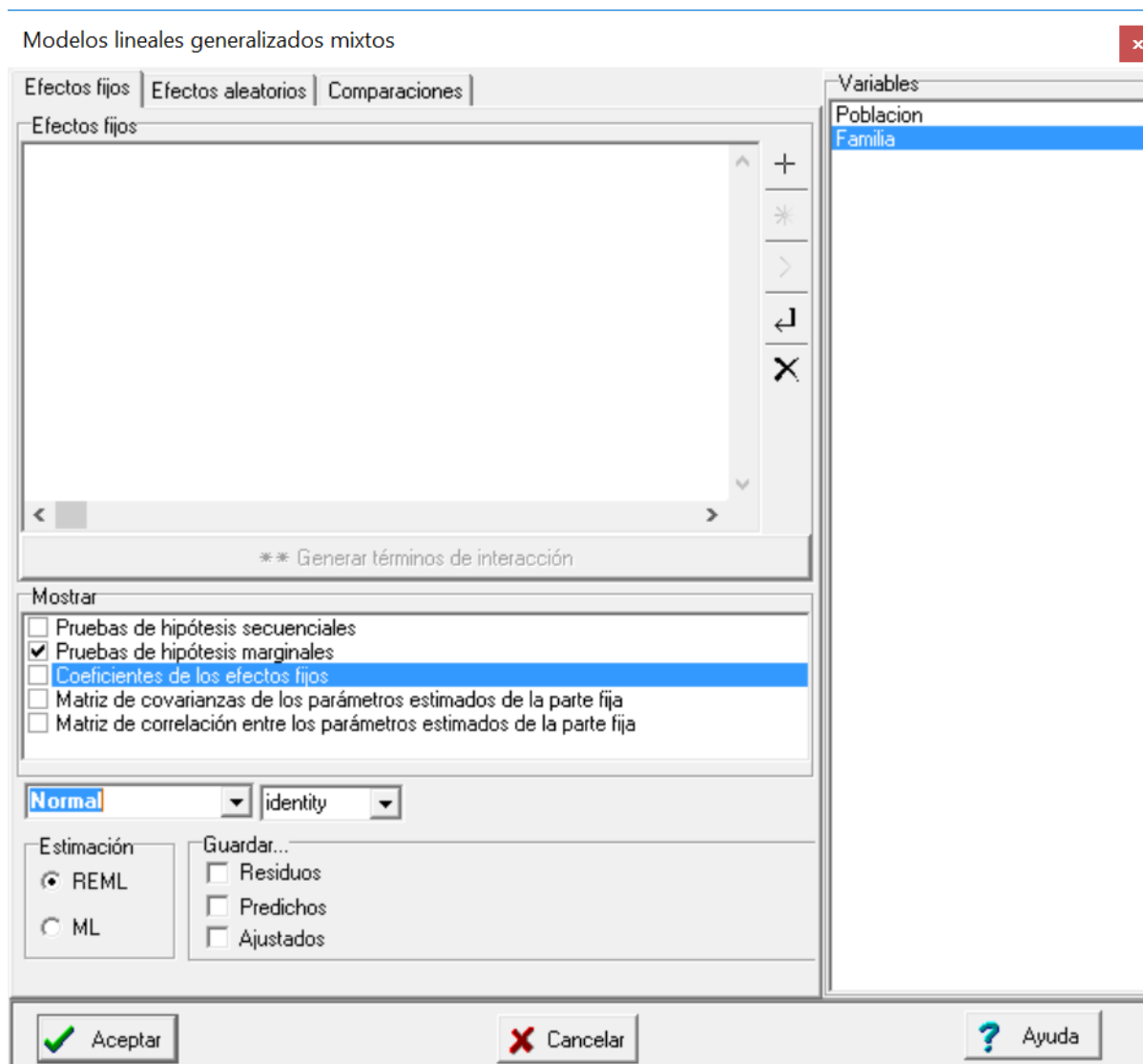


Figura 38: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y sin declaración de efectos fijos, usando familia Normal y función de enlace identidad para el análisis de los datos del archivo CompVar.IDB2.

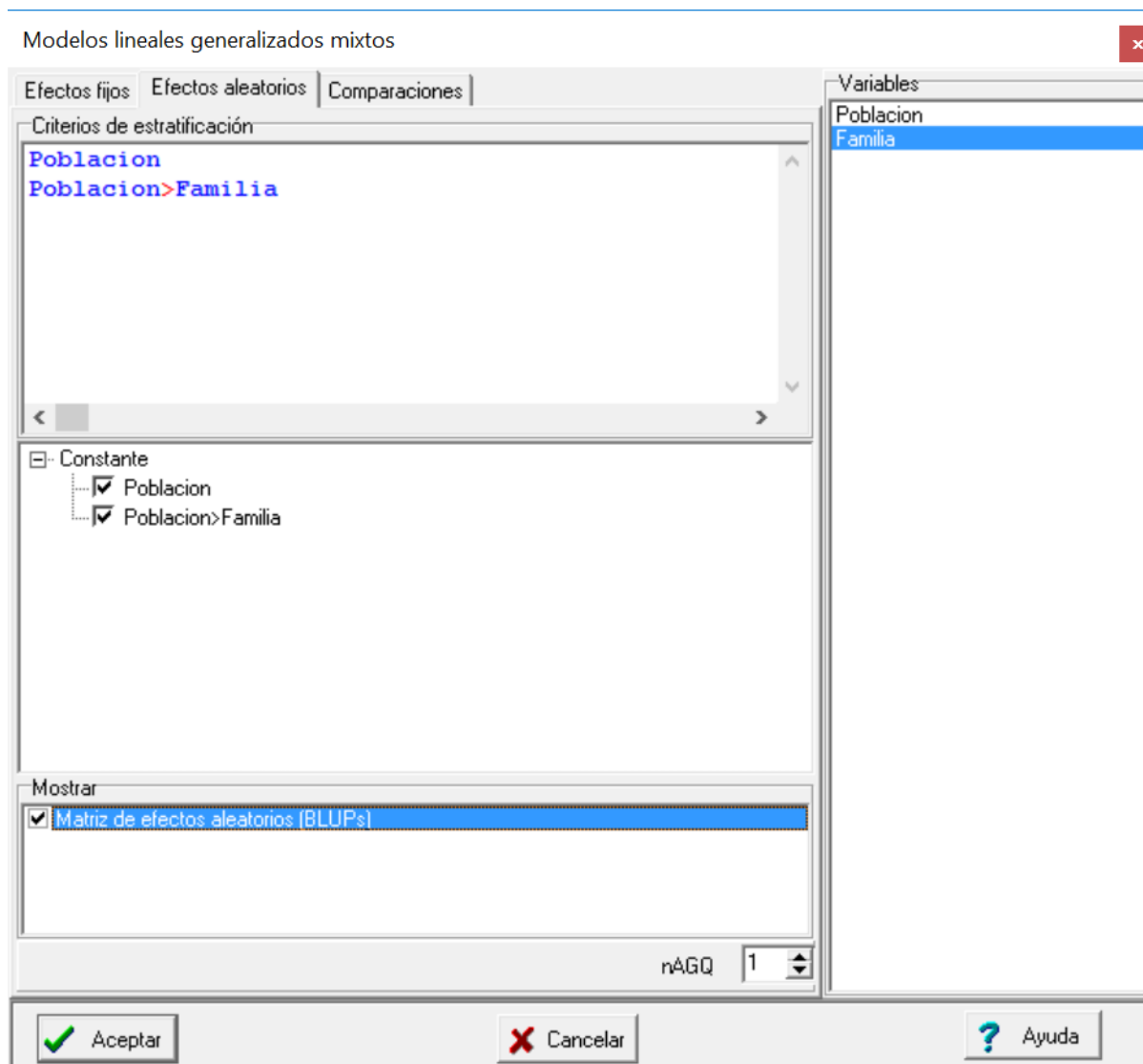


Figura 39: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y declarando los efectos de población y de familia dentro de población (Poblacion>Familia) para el análisis de los datos del archivo CompVar.IDB2.

A continuación se presenta la salida obtenida con las especificaciones anteriores:

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.011_Largo_REML<-
lmer(Largo~1+(1|Poblacion)+(1|Poblacion:Familia)
,na.action=na.omit
,REML=T
,control=lmerControl(optimizer="bobyqa")
,data=R.data11)
```


Resultados para el modelo: mlgm.modelo.011_Largo_REML*Variable dependiente: Largo***General**

Familia	Enlace	nAGQ
gaussian	identity	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
214	2016.47	2029.93	-1004.23	2008.47

*AIC y BIC menores implica mejor***Parámetros de los efectos aleatorios**

RndEff	Param	Var	Var%	SD
Poblacion:Familia	(Intercept)	219.01	15.42	14.80
Poblacion	(Intercept)	737.47	51.93	27.16
Residual		463.71	32.65	21.53

Se puede observar que el porcentaje de varianza explicado por la población es alto (51.93%). La varianza de familia dentro de población es menos de un tercio de la varianza de población (15.42%). La varianza no explicada en este problema es importante, ya que duplica a la de familia dentro de población.

Gamma: Disfunción eréctil

Fernández et al. (2010) analizaron datos provenientes de un estudio para comparar características de los latidos del corazón en pacientes con disfunción eréctil (P) y sujetos control (C). En cada paciente se midió HB, el cociente entre baja frecuencia (LF) y alta frecuencia (HF) obtenidas de la descomposición espectral del electrocardiograma, luego de tres estados (15 minutos acostado, 15 minutos sentado, y 15 minutos parado). Análisis preliminares indican que el uso de la distribución normal (aún luego de una transformación logarítmica) no explican adecuadamente estos datos, dada la naturaleza asimétrica de la distribución de los cocientes $HB=LF/HF$. La distribución gamma es preferible en estos casos, ya que brinda la flexibilidad necesaria en la forma manteniéndose en el marco de la familia exponencial de dos parámetros. Esto permite plantear un modelo lineal generalizado.

Se debe observar que cada paciente tiene tres observaciones, cada una tomadas luego de estar 15 minutos en una posición, y que el orden de las posiciones no puede ser aleatorizado debido a los efectos que tiene cada una sobre las posteriores. Por lo tanto será necesario incorporar correlación entre observaciones del mismo sujeto. En el marco de modelos lineales generalizados mixtos, esto se logra indirectamente mediante la adición de un efecto aleatorio de sujeto al modelo.

El interés principal es comparar los promedios de HB en pacientes con y sin disfunción eréctil en los tres estados (posiciones). Para ello ajustamos un modelo lineal generalizado mixto con los datos del archivo Disfunción eréctil.IDB2, seleccionando las variables como en la Figura 40, asumiendo distribución gamma y enlace log (Figura 41), que es el más usado para esta distribución (el enlace canónico es el inverso). El predictor lineal contiene efectos (fijos) de condición (paciente o control), estado (acostado, parado, sentado) y su interacción; y efecto aleatorio de sujeto (Figura 42).

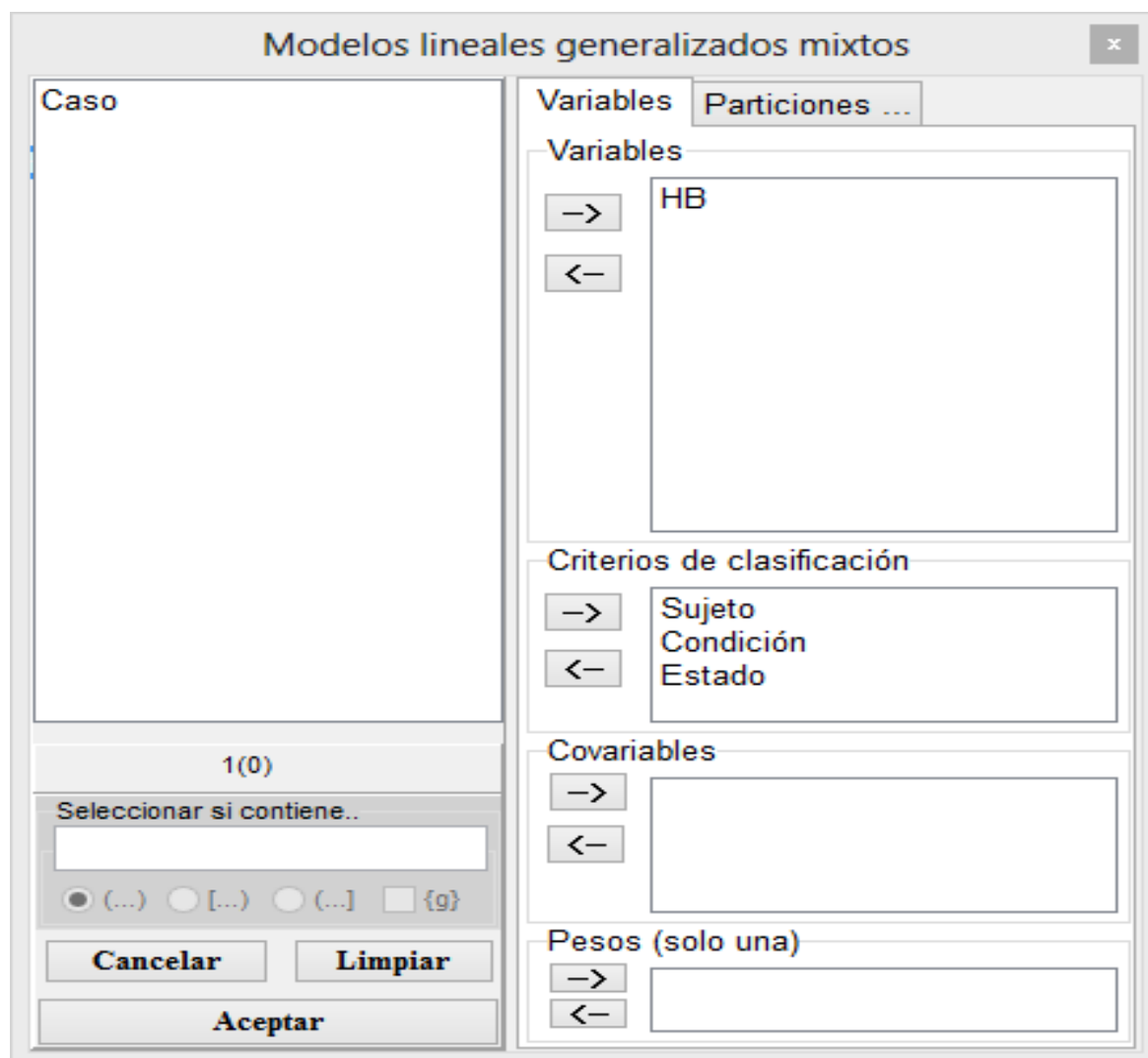


Figura 40: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo Disfunción erétil.IDB2.

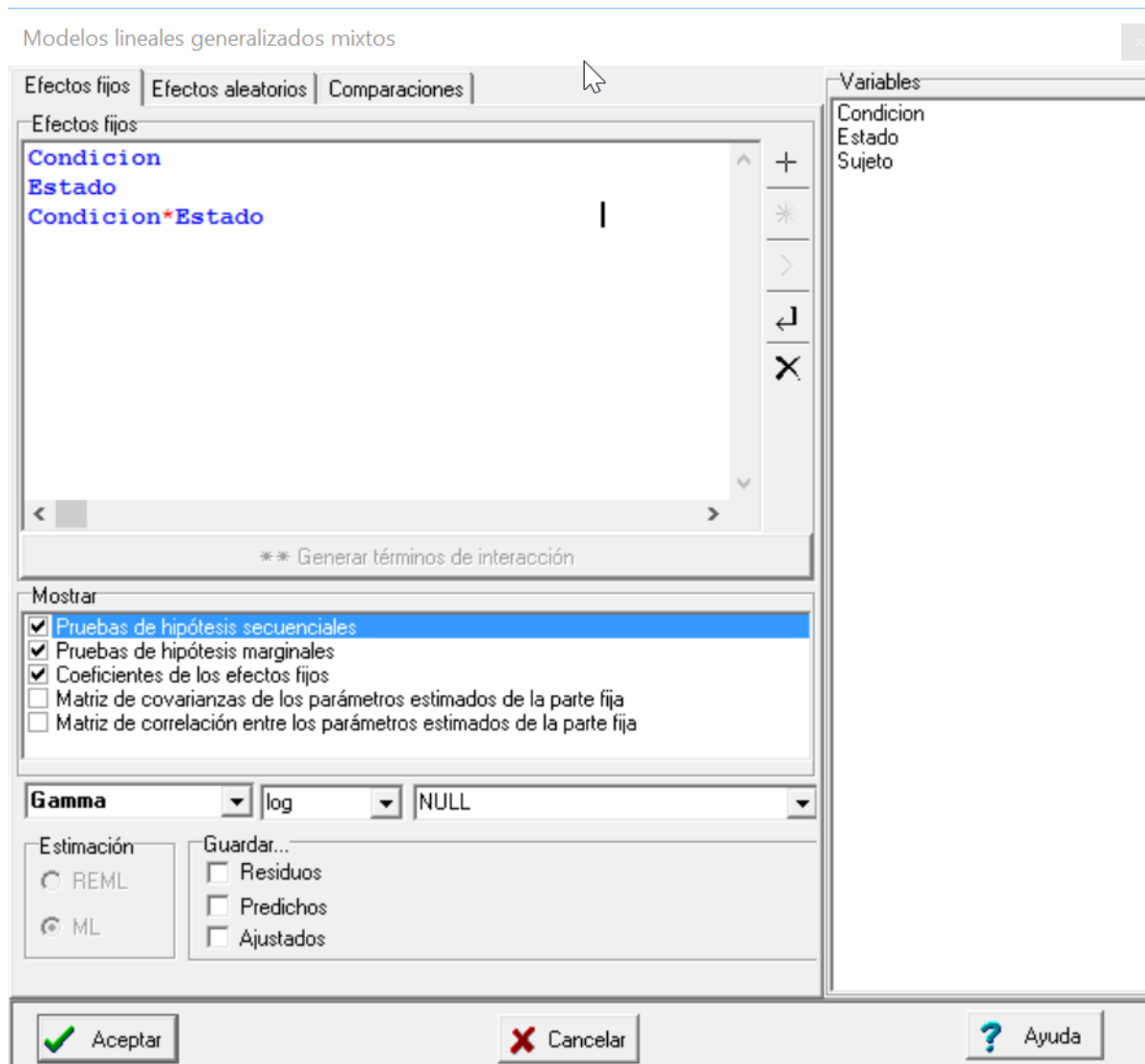


Figura 41: Ventana de Modelos lineales generalizados mixtos mostrando las solapa Efectos fijos con los efectos Condición, Estado y su interacción declarados. Datos en Disfunción eréctil.idb2.

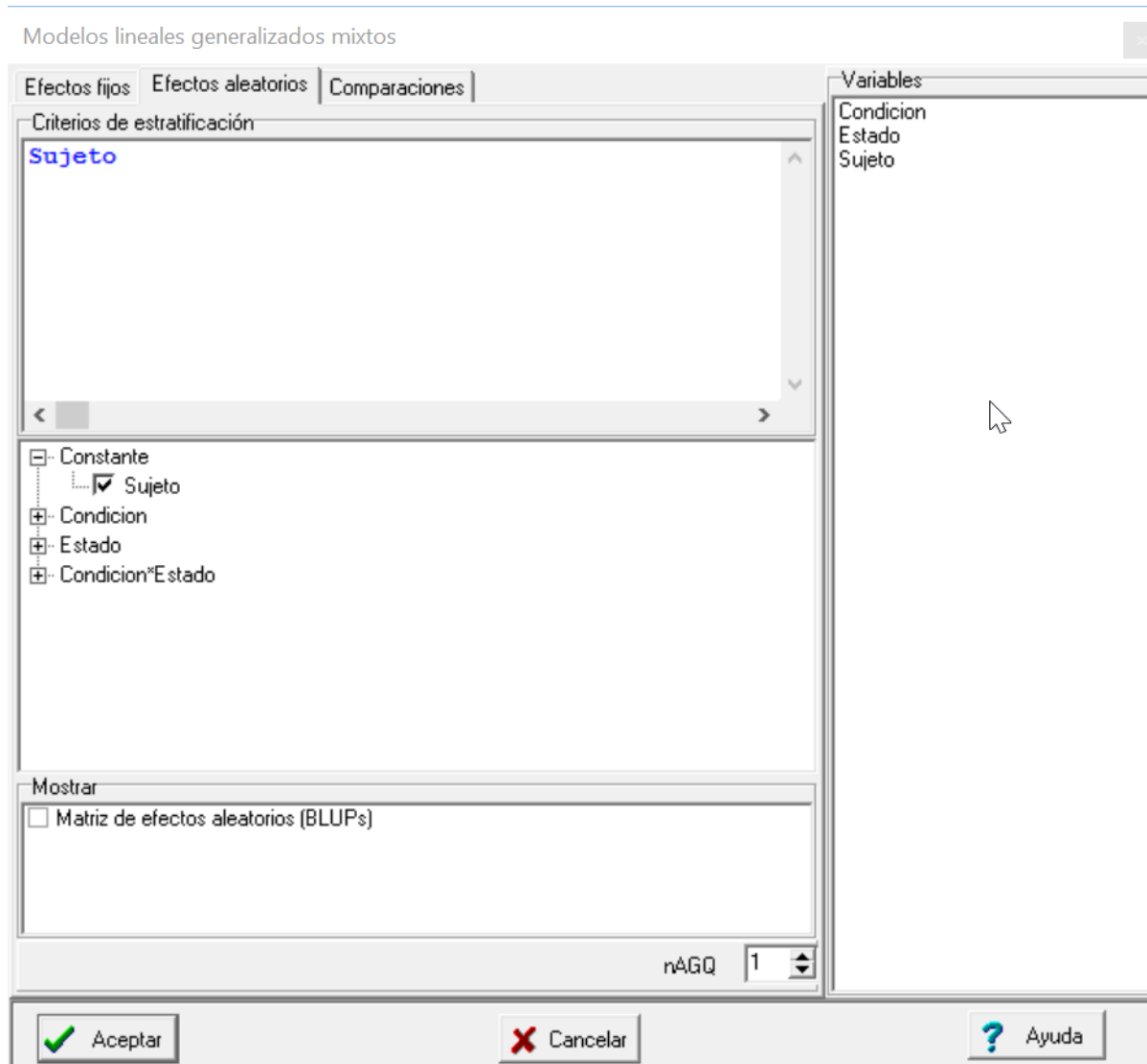


Figura 42: Ventana de Modelos lineales generalizados mixtos mostrando las solapas Efectos aleatorios desplegada con la identificación del sujeto. Datos en *Disfunción erétil.idb2*.

El resultado del análisis se muestra a continuación:

Modelos lineales generalizados mixtos

Especificación del modelo en R

```
mlgm.modelo.006_HB_ML<-
glmer(HB~1+Condición+Estado+Condición:Estado+(1|Sujeto)
,family=myFamily
,na.action=na.omit
,REML=F
,AGQ=1
,data=R.data06)
```

Resultados para el modelo: mlgm.modelo.006_HB_ML

Variable dependiente: HB

General

Familia	Enlace	nAGQ
Gamma	log	1

Medidas de ajuste del modelo

N	AIC	BIC	logLik	Deviance
87	369.72	389.44	-176.86	12.50

AIC y BIC menores implica mejor

Pruebas de hipótesis secuenciales para los efectos fijos

Term	Chi-square	df	p-value
Condicion	1.87	1	0.1711
Estado	72.22	2	<0.0001
Condicion:Estado	5.03	2	0.0810

Efectos fijos

	Estimate	Std. Error	t value	Pr(> z)
(Intercept)	0.70	0.16	4.25	<0.0001
CondicionP	0.33	0.25	1.30	0.1940
EstadoParado	1.23	0.13	9.27	<0.0001
EstadoSentado	0.34	0.13	2.57	0.0103
CondicionP:EstadoParado	-0.20	0.21	-0.93	0.3507
CondicionP:EstadoSentado	0.27	0.21	1.29	0.1964

Parámetros de los efectos aleatorios

RndEff	Param	Var	SD
Sujeto	(Intercept)	0.15	0.39
Residual		0.20	0.45

HB - Medias ajustadas y errores estándares para Condicion

Inversa de la función de enlace con efecto aleatorio=0

LSD Fisher (Alfa=0.05)

Procedimiento de corrección de p-valores: No

Condicion	PredLin	E.E.	Media	E.E.	
P	1.57	0.17	4.82	0.80	A
C	1.22	0.15	3.39	0.50	A

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

HB - Medias ajustadas y errores estándares para Estado*Inversa de la función de enlace con efecto aleatorio=0**LSD Fisher (Alfa=0.05)**Procedimiento de corrección de p-valores: No*

Estado	PredLin	E.E.	Media	E.E.	
Parado	1.99	0.13	7.32	0.92	A
Sentado	1.34	0.13	3.81	0.48	B
Acostado	0.86	0.13	2.37	0.30	C

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

HB - Medias ajustadas y errores estándares para Condicion*Estado*Inversa de la función de enlace con efecto aleatorio=0**LSD Fisher (Alfa=0.05)**Procedimiento de corrección de p-valores: No*

Condicion	Estado	PredLin	E.E.	Media	E.E.	
P	Parado	2.06	0.19	7.82	1.49	A
C	Parado	1.92	0.16	6.85	1.13	A B
P	Sentado	1.64	0.19	5.13	0.96	B
C	Sentado	1.04	0.17	2.83	0.47	C
P	Acostado	1.03	0.19	2.79	0.54	C D
C	Acostado	0.70	0.16	2.01	0.33	D

Medias con una letra común no son significativamente diferentes ($p > 0.05$)

Podemos observar que no se encontró una interacción significativa entre los factores condición y estado ($p=0.0810$). El factor condición no resultó significativo ($p=0.1711$) mientras que si lo fue el factor estado ($p<.0001$). La prueba LSD de Fisher permitió encontrar diferencias entre los tres niveles del factor estado, siendo el HB de parado mayor al de sentado, y este a su vez mayor que el HB de acostado.

Referencias

Agresti A. 2002. Categorical Data Analysis. New Jersey: John Wiley & Sons.

Fernández E. A., Souza Neto E.P., Abry P., Macchiavelli R., Balzarini M., Cuzin B., Baude C., Frutoso J., Gharib C. (2010) Assessing erectile neurogenic dysfunction from heart rate variability through a Generalized Linear Mixed Model framework. *Computer Methods and Programs in Biomedicine* 99: 49–56.

Bates D., Maechler M., Matrix LT. 2011. "Package 'lme4'." /packages/lme4/lme4. <http://mirrors.dotsrc.org/pub/pub/cran/web/packages/lme4/lme4.pdf> (Accessed February 26, 2012).

Di Rienzo J.A., Guzmán A.W., Casanoves F. 2002. "A multiple-comparisons method based on the distribution of the root node distance of a binary tree." *Journal of agricultural, biological, and environmental statistics* 7(2): 129–142.

Robles W., Pantoja A., Abreu E., Pena J., Ortiz J., Lugo M.D., Cortes M., Macchiavelli R. 2007. "Effects of cultural practices on the incidence of aphids and virosis on *Carica papaya* L." *Manejo Integrado de Plagas y Agroecología* 77: 38-42.

Breslow N., Clayton D. 1993. Approximate inference in generalized linear mixed models. *J. American Statistical Association* 88: 125-134.

Diggle P., Heagerty P., Liang K., Zeger S. 2002. *Analysis of Longitudinal Data*. 2da. Ed. Oxford Univ. Press: New York.

Thall P., Vail S. 1990. Some covariance models for longitudinal count data with overdispersion. *Biometrics* 46: 657-71.

Nime M., Casanoves F., Vrech D., Mattoni C. (2013). Relationship environmental variables and surface activity of scorpions in the Arid Chaco ecoregion of Argentina. *Invertebrate Biology* 132: 145-155.

Nime M., Casanoves F., Mattoni C. (2016). Microhabitat use and behavior differ across sex-age classes in the scorpion *Brachistosternus ferrugineus* (Scorpiones: Bothriuridae). *Journal of Arachnology* 44(2): 235-244.

Figura 1: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo PapayaBinomial.IDB2.....	5
Figura 2: Solapas con las opciones para especificación de un modelo lineal generalizado mixto.	5
Figura 3: Campos para la selección de la distribución, la función de enlace y le covariable que indica el número de experimentos para el análisis de los datos del archivo PapayaBinomial.IDB2.	6
Figura 4: Ventana con la solapa Efectos fijos desplegada para los datos del archivo PapayaBinomial.IDB2 con el menú contextual indicando donde colocar la variable número de ensayos.	7
Figura 5: Ventana desplegada con la solapa Comparaciones para los datos del archivo PapayaBinomial.IDB2.....	10
Figura 6: Especificación de contrastes para comparar tratamientos. Archivo PapayaBinomial.IDB2.	11
Figura 7: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables y covariables para el análisis de los datos del archivo PapayaBinomial.IDB2.	14
Figura 8: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de efectos de Trat, Dias y su interacción, usando una familia Binomial y enlace logit para el análisis de los datos del archivo PapayaBinomial.IDB2.	15
Figura 9: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y declaración de efectos de Parcela y enlace logit para el análisis de los datos del archivo PapayaBinomial.IDB2.....	16
Figura 10: Ventana de exploración de modelo: solapa [Combinaciones lineales]	19
Figura 11: Curvas de evolución de enfermedad para Control (M + T) y Tratado (PN + PP).....	23
Figura 12: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo Preñez.IDB2	24

Figura 13: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variable regresora DCL (Diámetro del cuerpo lúteo) usando una distribución Bernoulli (Binary) y función de enlace logit para el análisis de los datos del archivo Preñez.IDB2.	25
Figura 14: Datos observados (azul) y curva observada (negro) para los datos del ejemplo Preñez.IDB2.	27
Figura 15: Diagramas de caja para la variable Raíz cuadrada de la tasa de convulsiones (cantidad de convulsiones por semana) para pacientes con placebo y tratados con progabide medidos en el día de inicio del experimento y durante 4 semanas. Archivo Epilepsia.IDB2.....	28
Figura 16: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo Epilepsia.IDB2	31
Figura 17: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variables regresoras para el análisis de los datos del archivo Epilepsia.IDB2 usando una familia Poisson.	32
Figura 18: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variables regresoras para el análisis de los datos del archivo Epilepsia.IDB2 usando una familia Quasi Poisson.	34
Figura 19: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y la identificación de cada individuo declarada para el análisis de los datos del archivo Epilepsia.IDB2.	36
Figura 20: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variables regresoras incluyendo la constante (representada por el número 1) para el análisis de los datos del archivo Epilepsia.IDB2.....	39
Figura 21: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y la identificación de cada individuo declarada dentro de 1+Visita para el análisis de los datos del archivo Epilepsia.IDB2.	40

Figura 22: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración de variables regresoras para el análisis de los datos del archivo Epilepsia.IDB2 usando el modelo Binomial negativo.	44
Figura 23: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y declaración de efecto aleatorio de id para el análisis de los datos del archivo Epilepsia.IDB2 usando el modelo Binomial negativo.....	45
Figura 24: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y la declaración de “Tratam” como único término fijo del modelo. La familia seleccionada es Poisson. Datos: Malezas.IDB2.	47
Figura 25: Ventana de Modelos lineales generalizados mixtos con la solapa Comparaciones desplegada y la declaración de “Tratam” como único término fijo del modelo para el cual se piden la comparación de medias mediante LSD. Datos: Malezas.IDB2.	48
Figura 26: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y la declaración de “Tratam” como único término fijo del modelo. La familia seleccionada es Quasi-Poisson. Datos: Malezas.IDB2.....	50
Figura 27: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y la declaración de “Tratam” como único término fijo del modelo y la declaración de la Poisson (derecha). Solapa de efectos aleatorios indicando que la inclusión del efecto aleatorio Parcela. Datos: Malezas.IDB2.....	53
Figura 28: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo Escorpiones.IDB2.	59
Figura 29: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y declaración del efecto de sitio y todas las covariables ambientales, usando familia Poisson y función de enlace log para el análisis de los datos del archivo Escorpiones.IDB2.	60
Figura 30: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorio desplegada y declaración del efecto de transecta y de noche dentro de transecta para el análisis de los datos del archivo Escorpiones.IDB2.	61

Figura 31: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo EvaluacionHerbicidas.IDB2.	66
Figura 32: Gráficos de puntos de Recuentos promedio para los niveles del factor Herbicida y los niveles del factor Riego.	72
Figura 33: Gráficos de puntos de Predictor lineal para los niveles del factor Herbicida y los niveles del factor Riego.	73
Figura 34: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo AlfaLatticeFilaColumnaBinomial.IDB2.	75
Figura 35: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada con la declaración de efectos fijos de Genotipo, usando familia Binomial, función de enlace logit y la identificación del número de ensayos (PlantasExpuestas) para el análisis de los datos del archivo AlfaLatticeFilaColumnaBinomial.IDB2.	76
Figura 36: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada con la declaración de efectos de Año, Rep dentro de Año, fila dentro de Rep dentro de Año y Columna dentro de Rep dentro de Año para el análisis de los datos del archivo AlfaLatticeFilaColumnaBinomial.IDB2.	77
Figura 37: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo CompVar.IDB2.	83
Figura 38: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos fijos desplegada y sin declaración de efectos fijos, usando familia Normal y función de enlace identidad para el análisis de los datos del archivo CompVar.IDB2.	84
Figura 39: Ventana de Modelos lineales generalizados mixtos con la solapa Efectos aleatorios desplegada y declarando los efectos de población y de familia dentro de población (Poblacion>Familia) para el análisis de los datos del archivo CompVar.IDB2.	85
Figura 40: Ventana de selección de variables del módulo de Modelos lineales generalizados mixtos con la declaración de variables para el análisis de los datos del archivo Disfunción eréctil.IDB2.	88

Figura 41: Ventana de Modelos lineales generalizados mixtos mostrando las solapa Efectos fijos con los efectos Condición, Estado y su interacción declarados. Datos en Disfunción eréctil.idb2. 89

Figura 42: Ventana de Modelos lineales generalizados mixtos mostrando las solapas Efectos aleatorios desplegada con la identificación del sujeto. Datos en Disfunción eréctil.idb2. 90

Funciones de enlace en modelos lineales generalizados

El modelo lineal que se estima es de la forma general $\text{link}(\mu) = \mathbf{X}\beta$

Logit

Logit	Inversa	Derivada de la inversa
$\log\left(\frac{\mu}{1-\mu}\right)$	$\frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)}$	$\frac{\exp(\mathbf{X}\beta)}{(1 + \exp(\mathbf{X}\beta))^2}$

Probit

Probit	Inversa	Derivada de la inversa
$\phi^{-1}(\mu)$	$\phi(\mathbf{X}\beta)$	$\frac{1}{\sqrt{2\pi}} \exp\left(-(\mathbf{X}\beta)^2\right)$

Complementary log-log

Cloglog	Inversa	Derivada de la inversa
$\log(-\log(1-\mu))$	$1 - \exp(-\exp(\mathbf{X}\beta))$	$(1-\mu)\exp(\mathbf{X}\beta)$

Log

log	Inversa	Derivada de la inversa
$\log(\mu)$	$\exp(\mathbf{X}\beta)$	$\exp(\mathbf{X}\beta) = \mu$

Sqrt

Sqrt	Inversa	Derivada de la inversa
$\text{sqrt}(\mu)$	$(\mathbf{X}\beta)^2$	$2\mathbf{X}\beta$

Identity

Identity	Inversa	Derivada de la inversa
μ	$\mathbf{X}\beta$	1

$1/\mu^2$

$1/u^2$	Inversa	Derivada de la inversa
μ^{-2}	$\mathbf{X}\beta^{-\frac{1}{2}}$	$-\frac{1}{2}\mathbf{X}\beta^{-\frac{3}{2}}$

Inverse

Inverse	Inversa	Derivada de la inversa
$-\mu^{-1}$	$-(\mathbf{X}\boldsymbol{\beta})^{-1}$	$(\mathbf{X}\boldsymbol{\beta})^{-2}$

Método delta para calcular errores estándar de medias en escala original (enlace inverso) conociendo el error estándar de $X\hat{\beta}$

$$s.e.(\hat{\mu}) = s.e.(\text{link}^{-1}(X\hat{\beta})) = \left| \frac{d \text{link}^{-1}(X\hat{\beta})}{d X\hat{\beta}} \right| s.e.(X\hat{\beta})$$

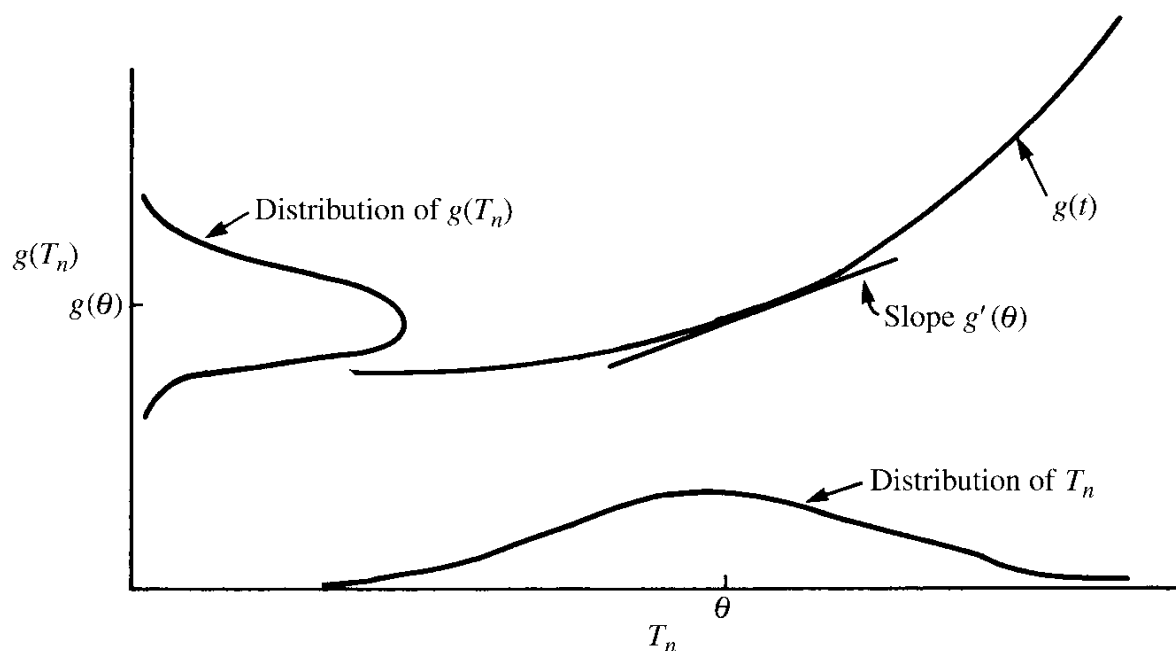


FIGURE 3.1 Depiction of delta method. (Agresti 2002), p. 74