

**UNIVERSIDAD NACIONAL DEL COMAHUE**

**Maestría en Estadística Aplicada**



**MODELACIÓN DE RESPUESTAS ORDINALES  
LONGITUDINALES MEDIANTE MODELOS LINEALES  
GENERALIZADOS MIXTOS**

**Trabajo de Tesis presentado por:**

**Autor**

**Natalia Rubio**

**Profesora de Matemáticas**

**(Universidad Nacional del Comahue, 2003)**

**Neuquén, 2016**

## **COMISIÓN ASESORA DE TESIS**

### **Director**

**Dr. Raúl E. Macchiavelli**

Ph.D., Statistics.  
Magister Scientiae, Biometría  
Ingeniero Agrónomo

### **Co-directores**

**Mg. Andrea Lina Lavallo**

**Mg. Gustavo Néstor Giménez**

## **JURADO DE TESIS**

**Miembro 1**

**Miembro 2**

**Miembro 3**

**Fecha de aprobación de la tesis: de xxxx de 2016**

**Director:** Dr. Raúl E. Macchiavelli

**Co- Directores:**

Mg. Andrea Lina Lavalle

Mg. Gustavo Néstor Giménez

**CERTIFICAN:**

Que la maestra Natalia Rubio ha realizado bajo mi dirección la tesis “**Modelación de respuestas ordinales longitudinales mediante Modelos Lineales Generalizados Mixtos**” la cual se encuentra finalizada y en condiciones de ser presentada.

Y para que así conste a todos los efectos del interesado se expide el presente certificado en Neuquén a los 29 días del mes de Marzo del año 2016.



*Dr. Raúl E. Macchiavelli*

*Mg. Andrea Lina Lavalle*

*Mg. Gustavo Néstor Giménez*

## AGRADECIMIENTOS

A toda mi familia y amigos que siempre me acompañaron a lo largo de mi carrera. En particular a mi hijo, **Iñaki**, que tuvo que esperar tardes y noches a que mamá terminara de estudiar. También quiero agradecer a mis compañeros de trabajo del Departamento de Estadística por el apoyo incondicional. Y un inmenso agradecimiento a mi director que siempre me acompañó desde lo académico en la investigación y desde lo personal ayudándome a no bajar los brazos. A mis dos co-directores que me enriquecieron en todo el crecimiento intelectual que con mucha dedicación estuvieron en todo momento.

## Modelación de respuestas ordinales longitudinales mediante Modelos Lineales Generalizados Mixtos

### Resumen

Los Modelos Lineales Generalizados Mixtos (MLGM) son una extensión del modelo lineal generalizado (MLG) adicionando los efectos aleatorios. En particular en estudios longitudinales un camino de modelación es considerar al individuo como efecto aleatorio. En este trabajo se analiza la aplicabilidad de dichos modelos para el caso de una variable de respuesta ordinal cuando los datos son de naturaleza longitudinal. Para ilustrar las técnicas abordadas en esta investigación se modela la variable ordinal *clase diamétrica* de distintos grupos funcionales de plantas expuestos a diferentes intensidades de perturbación a lo largo del tiempo, en dos bosques tropicales Mesoamericanos de Costa Rica: Tirimbina y Corinto. En una primera instancia, se aplican técnicas de análisis multivariado para la caracterización de la variable de respuesta ordinal, tanto para definir una recategorización de la misma como para realizar un análisis descriptivo de los efectos que intervienen en el diseño del problema. Se probaron cuatro modelos Log - Acumulados considerando el cumplimiento o no del supuesto de proporcionalidad e incorporando el efecto aleatorio del individuo. Del análisis de estos modelos se destacan ventajas y desventajas de cada uno. Para el caso concreto de estudio se propone un Modelo Log – Acumulado Mixto no proporcional que tiene en cuenta la ordinalidad de la variable, el carácter longitudinal de los datos y presenta la flexibilidad suficiente como para permitir la modelación sujeto específico.

**Palabras Claves:** *Modelos Lineales Generalizados Mixtos, Modelo Log – Acumulado, Bosques Tropicales*

## **Modeling longitudinal ordinal responses using Generalized Linear Mixed**

### **ABSTRACT**

The Generalized Mixed Linear Models (MLGM) are an extension of the Generalized Linear Model (GLM) adding the random effects. In longitudinal studies, a common way of modeling is to consider the individual as a random effect. In this work the applicability of such models for the case of an ordinal response variable is analyzed when the data are longitudinal. To illustrate the techniques discussed in this research, ordinal variable diameter class of different functional groups of plants exposed to different intensities of disturbance over time, are modeled in two Mesoamerican tropical forests of Costa Rica (Tirimbina and Corinto). First, multivariate analysis techniques are applied to the characterization of the ordinal, both to define a reclassification and as a descriptive analysis of the effects involved in the problem. Four Log - Accumulated models were tested, considering compliance or not the assumption of proportionality and incorporation or not of random effect of the individual. For each of these models the advantages and disadvantages are discussed. For this particular problem, a Mixed Accumulated not proportional model is proposed. It takes into account the ordinal variable, the longitudinal nature of the data and presents sufficient flexibility to allow subject - specific modeling.

**Keywords:** *Generalized Linear Mixed Models, Model Log - Accumulated, Rainforests.*

*INDICE GENERAL*

Resumen.....	V
Abstrac .....	VI
Introducción .....	1
<b>Capítulo I. Planteo del problema.....</b>	<b>5</b>
Objetivo General.....	10
Objetivos específicos.....	10
<b>Capítulo II. Marco teórico .....</b>	<b>11</b>
2.1 Modelo Lineal Generalizado con efectos fijos .....	11
2.2 Modelo Logit .....	14
2.3 Modelo Logit Multicategorico .....	16
2.3.1 Modelo Logit para variables Nominales .....	17
2.3.2 Modelo Logit para variables Ordinales .....	18
2.3.3 Estimación .....	20
2.4 Modelo Lineal Mixto para variables de respuesta con distribución normal.....	22

---

2.5	Medidas Repetidas - Datos Longitudinales .....	25
2.5.1	Notación en observaciones longitudinales .....	28
2.5.2	Modelos de análisis de datos longitudinales .....	29
2.6	Modelo Sujeto específico .....	33
2.7	Modelos Lineales Generalizados Mixtos (MLGM) .....	35
2.7.1	Modelo Logit Mixto para variable de respuesta nominal.....	38
2.7.2	Modelo Logit Acumulado Mixto para respuesta Ordinal .....	39
	<b>Ejemplo 1.</b> Evaluación acerca de la astringencia de vino .....	41
	<b>Ejemplo 2.</b> Relevadores de tamaño de frutos peras William´s .....	47
	<b>Capítulo III. Aplicación.....</b>	<b>51</b>
3.1	Análisis descriptivo Multivariado .....	53
3.2	Modelo Logit para variables ordinales .....	58
3.2.1	Bosque Corinto .....	59
3.2.2.	Bosque La Tirimbina.....	63
3.3.	Modelo Log – Acumulado Mixto para variables ordinales .....	67
3.3.1.	Análisis descriptivo .....	68
3.3.2	Modelo Lineal Generalizado Mixto .....	75
3.3.3.	Análisis de Proporcionalidad .....	84
3.3.4.	Estimación del modelo definitivo .....	87



**Capítulo IV. Conclusiones ..... 91**

**REFERENCIAS ..... 95**

Anexo1: Salidas completas del programa sas .....98

Anexo2: Salidas completas del Software R ..... 120

Anexo3: Tablas de estimaciones .....124

**INDICE DE TABLAS**

Tabla 1.1. Años de medición para las parcelas permanentes de La Tirimbina..8

Tabla 1.2. Años de medición para las parcelas permanentes de Corinto.....8

Tabla 1.3. Caracterización del área de estudio ..... 9

Tabla 1.4. Reducción de área basal en las parcelas de Tirimbina y Corinto.... 9

Tabla 2.1. Expresiones para  $a$ ,  $\psi(\cdot)$  y  $\alpha(\cdot)$  para distribuciones de un solo parámetro ..... 12

Tabla 2.2. Datos de Astringencia de vinos ..... 41

Tabla 2.3. Estimación de los parámetros del modelo 2.34 mediante la Aproximación de Laplace ..... 42

Tabla 2.4. Estimación de los parámetros del modelo mediante el método de cuadratura de Gauss-Hermite (10 nodos) ..... 43

Tabla 2.5. Prueba de Razón de Verosimilitud – (Rating = Contacto + Temperatura vs Rating = Temperatura) ..... 44

---

Tabla 2.6. Prueba de Razón de Verosimilitud – (Rating = Contacto + Temperatura para evaluar el efecto Juez) .....	44
Tabla 2.7: Intervalo de confianza para $\sigma_u$ .....	45
Tabla 2.8. Test de verosimilitud de los modelos analizados .....	48
Tabla 2.9. Estimación de los parámetros del modelo. TAMAÑO ~ ANIO + EDAD + (1 Enc) .....	48
Tabla 3.1. Análisis de efectos Tipo 3. Resumen Bosque Corinto .....	59
Tabla 3.2. Análisis de efectos Tipo 3. Resumen Bosque Corinto .....	60
Tabla 3.3. Test de disparidad proporcional. Bosque Corinto .....	62
Tabla 3.4. Análisis de efectos Tipo 3. Resumen Bosque La Tirimbina .....	63
Tabla 3.5. Análisis de efectos Tipo 3. Resumen Bosque La Tirimbina .....	64
Tabla 3.6. Test de disparidad proporcional. Bosque Tirimbina .....	66
Tabla 3.7. Información del Ajuste del Modelo 1. Bosque Corinto .....	76
Tabla 3.8: Información del Modelo 2. Bosque Corinto .....	76
Tabla 3.9. Estimaciones del Modelo 2. Bosque Corinto .....	77
Tabla 3.10. Estimaciones de los umbrales del Modelo 2. Bosque Corinto .....	77
Tabla 3.11. Test de verosimilitud para el efecto de la parcela. Bosque Corinto.....	78
Tabla 3.15. Información del Ajuste del Modelo 1. Bosque La Tirimbina .....	80
Tabla 3.16. Test de verosimilitud para comparar el Modelo 1 y el Modelo 2. Bosque La Tirimbina .....	81
Tabla 3.17. Test de verosimilitud para comparar el Modelo 2 y el Modelo 3. Bosque La Tirimbina .....	81

Tabla 3.18. Información del Modelo 3. Bosque La Tirimbina .....	82
Tabla 3.19. Estimaciones del Modelo 3 .....	82
Tabla 3.20. Estimaciones de los umbrales del Modelo 3. Bosque La Tirimbina.....	82
Tabla 3.21. Test de verosimilitud para el efecto de la parcela. Bosque La Tirimbina .....	82
Tabla 3.22. Test de verosimilitud para el efecto de la parcela. Bosque Corinto.....	85
Tabla 3.23. Test de disparidad proporcional. Bosque Corinto .....	85
Tabla 3.24. Test de verosimilitud para el efecto de la parcela. Bosque La Tirimbina .....	86
Tabla 3.25. Test de disparidad proporcional. Bosque La Tirimbina .....	87

**INDICE DE FIGURAS**

Figura 1.1. Parcelas permanentes de los Bosques Corinto y La Tirimbina .....	6
Figura 1.2 Mapa de la región de estudio .....	9
Figura 2.1. Representación de distintas Funciones de enlace .....	14
Figura 2.2. Modelo logit acumulado odds – proporcional .....	19
Figura 2.3. Probabilidades de las Categorías en modelo logit acumulativo .....	20
Figura 2.3. Perfiles Sujeto específico y promedio poblacional .....	24

---

Figura 2.4. Relación entre la habilidad de lectura vs la edad .....	26
Figura 2.4. Perfil de verosimilitud de $\sigma_u$ .....	45
Figura 2.5: Intervalos del 95% de confianza basados en la varianza condicional para el efecto Juez .....	46
Figura 2.6. Perfil relativo de verosimilitud .....	49
Figura 2.7. Intervalos de confianza para cada uno de los Relevadores.....	49
Figura 2.8. Probabilidades para el relevador promedio y extremos en distintas condiciones .....	50
Figura 3.1. Proyección de las variables en plano .....	54
Figura 3.2. Proyección de los individuos en plano identificados por los Grupos Funcionales .....	54
Figura 3.3. Proyección de los individuos en plano identificados por el Bosque al que pertenecen .....	55
Figura 3.4. Proyección de los individuos en plano identificados por Año .....	56
Figura 3.5. Proyección de los individuos en plano identificados por Tratamiento.....	56
Figura 3.6. Gráfico de probabilidad de acumuladas. Bosque Corinto .....	61
Figura 3.7. Grafico de probabilidad de acumuladas. Bosque Tirimbina .....	65
Figura 3.8. Gráfico de Mosaicos. Grupos Funcionales – Tratamiento – Clases Diamétricas. Bosque Corinto .....	68

Figura 3.9. Gráfico de Mosaicos. Grupos Funcionales – Parcela - Clases Diamétricas. Bosque Corinto .....	69
Figura 3.10. Gráfico de Mosaicos. Grupos Funcionales – Año – Clases Diamétricas. Bosque Corinto. ....	69
Figura 3.11. Gráfico de Mosaicos. Grupos Funcionales – Año - Tratamiento. Bosque Corinto .....	70
Figura 3.12. Gráfico de Mosaicos. Grupos Funcionales versus Años versus Parcela. Bosque Corinto .....	70
Figura 3.13. Gráfico de Mosaicos. Grupos Funcionales – Tratamiento – Clases Diamétricas. Bosque Tirimbina .....	71
Figura 3.14. Gráfico de Mosaicos. Grupos Funcionales – Parcela – Clases Diamétricas. Bosque Tirimbina .....	72
Figura 3.15. Gráfico de Mosaicos. Grupos Funcionales – Años – Clases Diamétricas. Bosque Tirimbina .....	73
Figura 3.16. Gráfico de Mosaicos. Grupos Funcionales versus Años versus Tratamientos. Bosque Tirimbina .....	74
Figura 3.17. Gráfico de Mosaicos. Grupos Funcionales versus Años versus Parcela. Bosque Tirimbina .....	74
Figura 3.18. Intervalo de confianza para los efectos predichos de cada parcela. Bosque Corinto .....	78
Figura 3.19. Perfil relativo de verosimilitud. Bosque Corinto .....	79
Figura 3.20. Intervalo de confianza para los efectos predichos de cada parcela. Bosque La Tirimbina .....	83
Figura 3.21. Perfil relativo de verosimilitud. Bosque La Tirimbina .....	83

## INTRODUCCION

La investigación en las ciencias agrícolas y forestales comúnmente involucra situaciones donde es difícil utilizar los modelos lineales clásicos del análisis de la varianza y regresión, porque no siempre se cumplen los supuestos de independencia, normalidad, igualdad de varianzas e incluso linealidad para todas las variables intervinientes. Para lograr un análisis completo ya sea de diagnóstico o de inferencia, puede ser necesaria la utilización de modelos lineales generalizados mixtos. La modelación de datos experimentales desde el marco teórico de los modelos lineales generalizados mixtos brinda la posibilidad de analizar datos con estructuras de dependencia, desbalances y falta de normalidad. Estos permiten contemplar la falta de cumplimiento de los supuestos tradicionales y modelar, de manera flexible, complicadas estructuras de datos. Existen muchos beneficios que pueden ser obtenidos con el uso de modelos mixtos. En algunas situaciones se incrementa la precisión de las estimaciones; en otras, se amplía el espacio de inferencia y se comprende mejor la estructura de los datos.

Se han desarrollado distintos métodos de análisis para el tratamiento de variables de respuesta categóricas ya que son muy comunes en las más diversas disciplinas: sociales, médicas, biológicas, genéticas, de salud pública, epidemiológica, agronómica y forestal, entre otras. En estos casos la escala de medida de la variable determina qué método estadístico es apropiado. Las variables categóricas tienen dos tipos de escala: las variables que no tienen una escala de orden natural son llamadas nominales y las variables donde sus categorías tienen un orden son llamadas ordinales (Agresti, 2002). El uso de variables de respuesta ordinales es muy común y sin embargo el estudio de las mismas es escaso cuando se presentan situaciones con datos longitudinales o mediciones repetidas. Numerosos autores han desarrollado los Modelos

Lineales Generalizados Mixtos (GLMMs) (condicional) para variables de respuestas con distribución binomial (por ejemplo, Stiratelli et al., 1984; Aitkin y Anderson, 1985; Gilmour et al., 1985) o con distribución Poisson (por ejemplo, Hinde, 1982; de Breslow, 1984). En cambio ha habido menos para el desarrollo de GLMMs para respuestas multinomiales. Harville y Mee (1984) propusieron un modelo de efectos aleatorios con enlace probit acumulado donde utilizaron la aproximación de la serie de Taylor para las integrales. Jansen (1990) y Ezzet y Whitehead (1991) propusieron modelos probit y logit acumulado con intercepto aleatorio. Ten Have y Uttal (1994) utilizó un modelo logit de efectos aleatorios para el análisis de múltiples tiempos en perfiles de supervivencia. También Coull y Agresti (2000) y Dos Santos y Berridge (2000), Crouchley (1995) elaboraron modelos con función de enlace log-log complementario usando distribuciones de los efectos aleatorios de manera que la distribución marginal tuviese forma cerrada. He y Ten Have (1996) extendieron un modelo binario de efectos aleatorios de Conaway (1990) a datos ordinales, dando una forma cerrada para la función de verosimilitud de los resultados utilizando enlace acumulado complemento log-log y distribución de efectos aleatorios log-gamma. Hartzel J., Agresti A., y Caffo B. (2001) analizan el modelo multinomial con efectos aleatorios y Hartzel J., I-Ming Liub y Agresti A. (2001), aplican el modelo de odds proporcional en tablas de contingencia donde las variables son ordinales analizando los algoritmos de la cuadratura de Gauss –Hermite y Montecarlo para la estimación de los parámetros. Además, respecto a los avances en el software R se han desarrollado distintos paquetes y funciones para Modelos Generalizados Mixtos, Agresti A. (2011) analiza distintos ejemplos mediante el uso de la función “*vglm*” del software R mediante Modelos Generalizados, Christensen R. (2011) analiza la función “*clm*” del software libre R para la estimación de modelos con función de enlace logit acumulado, donde la variable de respuesta es ordinal pero sin considerar el efecto aleatorio sujeto específico, sin embargo en el mismo año también analiza la función “*clmm2*” que considera el efecto del individuo mediante un Modelo Generalizado Mixto también para una variable de respuesta ordinal considerando un modelo de odds proporcional.

Si bien todos estos avances han sido fundamentales para el desarrollo de los modelos generalizados mixtos, en situaciones reales no están muy extendidos

y donde existe una gran falencia es en situaciones donde la variable de respuesta tiene una distribución Multinomial, en particular si dicha variable tiene naturaleza ordinal; es por ello que es de fundamental importancia el análisis de la modelación de variables de respuesta ordinal en el contexto de modelos lineales generalizados mixtos.

Por otro lado, en problemas aplicados es común tener mediciones repetidas secuenciales, también llamadas datos longitudinales, es decir, varias observaciones tomadas secuencialmente sobre una misma unidad (parcela, árbol, sujeto, etc.). La secuencia puede ser temporal (mediciones repetidas en el tiempo) o espacial (datos correlacionados espacialmente).

Las medidas repetidas en el tiempo están correlacionadas dentro de los individuos, pero son independientes entre los sujetos. Las técnicas de regresión ordinaria asumen independencia entre las medidas repetidas y por lo tanto no son apropiadas para analizar datos longitudinales (Verbeke y Molenberghs, 2000). Los Modelos Mixtos permiten inducir la estructura de correlación intra-sujeto, y más aún, permiten la inclusión en el modelo de la variabilidad entre sujetos. Para los datos longitudinales, el componente aleatorio del modelo mixto permite interceptos sujetos específicos y pendientes a través del tiempo. Además, permite la presencia de datos faltantes y variables en el tiempo (Pinheiro y Bates, 2000; Verbeke y Molenberghs, 2000).

El marco teórico de modelos mixtos brinda herramientas poderosas para modelar datos longitudinales contemplando correlaciones espaciales y/o temporales.

El sistema de análisis estadístico SAS posee nuevos procedimientos para el uso de modelos mixtos. Con base en este recurso computacional se realizará el análisis avanzado de los datos en el contexto de Modelos Mixtos. Y para el análisis de Modelos Generalizados Mixtos se utilizarán herramientas gráficas y de análisis estadístico del software libre R, pudiendo comparar las ventajas y limitaciones respecto del uso del software SAS.

Como aplicación de las técnicas abordadas en esta investigación se modelará la variable ordinal *clase diamétrica* de distintos grupos funcionales de plantas expuestas a diferentes intensidades de perturbación, en dos bosques tropicales Mesoamericanos de Costa Rica: Tirimbina y Corinto. (Fuente: Dr. Brian Finegan, Datos de parcelas permanentes de la Cátedra Latinoamericana de



Ecología en el Manejo de Bosques Tropicales, Centro Agronómico Tropical de Investigación y Enseñanza, CATIE, Costa Rica). En particular se modelarán las frecuencias registradas en la variable de respuesta *clase diamétrica* (con seis categorías), en función de tres factores: tiempo, grupos funcionales de plantas y tratamientos de recuperación del bosque.

La variable de respuesta *clase diamétrica* cuenta con seis categorías: C1: [10-20), C2: [20-30), C3: [30-40), C4: [40-50), C5: [50-60), C6: [60,+∞), es decir los datos son multinomiales ordinales. Por lo tanto será necesario recurrir al uso de modelos generalizados. Asimismo, esta variable de respuesta está registrada a lo largo de diferentes periodos en las mismas parcelas, donde identificamos al individuo como cada parcela, por lo que se analizarán dichas observaciones a través del enfoque de los Modelos Mixtos los cuales nos facilitarán el análisis de datos correlacionados espacialmente.

## **CAPITULO I**

### **Planteo del problema:**

Por lo expuesto anteriormente, el problema de investigación que se aborda en este trabajo de tesis consiste en el análisis y estudio de la aplicabilidad de los Modelos Lineales Generalizados Mixtos, para la modelación de una variable de respuesta de naturaleza ordinal.

Para ilustrar el estudio se utilizará un conjunto de datos relevados en la investigación llevada a cabo por el Dr. Brian Finegan titulada “Efecto de diferentes intensidades de perturbación en un bosque primario”. Se utiliza esta base de datos debido a que los mismos permiten ser analizados mediante las distintas técnicas (técnicas multivariadas, MLG, MLGM, técnicas para estudios longitudinales) que se desean estudiar a lo largo de todo el trabajo de investigación.

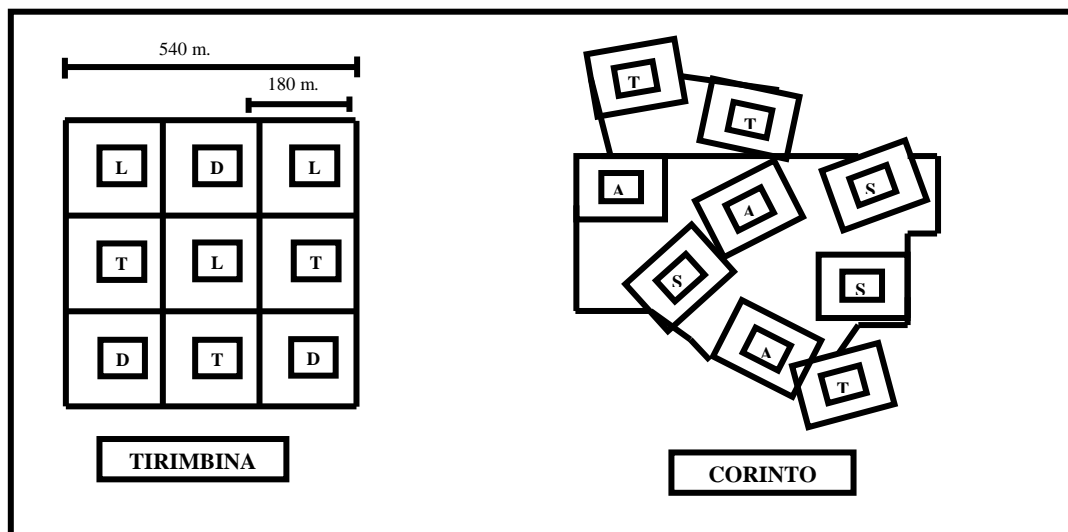
Este conjunto de datos está identificado de la siguiente manera: Datos de parcelas permanentes de la Cátedra Latinoamericana de Ecología en el Manejo de Bosques Tropicales, Centro Agronómico Tropical de Investigación y Enseñanza, CATIE, Costa Rica.

El objetivo de dicha investigación fue estudiar cómo un ecosistema se ve afectado por perturbaciones a sus componentes. Las perturbaciones forestales pueden llegar a ocasionar efectos profundos en la estructura, composición y funciones del bosque. Existen distintos aspectos a tener en cuenta al estudiar las perturbaciones que sufre un Bosque Tropical. Las más importantes son el origen de la perturbación (Natural o Antropogénico), la frecuencia (Aguda o Crónica) y la severidad de la misma (mortalidad parcial o completa). El eje de este trabajo es estudiar las perturbaciones Antropogénicas, es decir aquellas que no sean naturales y por lo tanto son provocadas por el hombre. Se estudiarán las referidas a los tratamientos que se desarrollan en ambos bosques: Aprovechamiento forestal, Aprovechamiento Silvicultural y el que combina a ambos. Donde el Aprovechamiento Forestal se refiere al conjunto de operaciones que consisten en: la preparación parcial, la extracción y el

transporte de la madera y leña que se obtienen de las talas realizadas, con una planificación adecuada, en una masa forestal. El Aprovechamiento Silvicultural es en el que emplea diferentes tratamientos silvícolas en función de lo que quiera obtener, como madera, leña, frutos o calidad ambiental. Por ello, la silvicultura siempre ha estado orientada a la conservación del medio ambiente y de la naturaleza, a la protección de cuencas hidrográficas, al mantenimiento de pastos para el ganado y a la fruición pública de los bosques.

El estudio se llevo a cabo en dos bosques primarios de Costa Rica: La Tirimbina y Corinto. En La Tirimbina se encuentran 9 parcelas, de las cuales 6 fueron aprovechadas a finales de 1989 y principios de 1990, y se les aplicaron tratamientos silviculturales y de aprovechamiento forestal. El área de estudio en Corinto tiene 41 ha de bosque primario donde hay establecidas 9 parcelas permanentes, tales parcelas están distribuidas como se muestra en la Figura 1.1.

**Figura 1.1.** Parcelas permanentes de los Bosques Corinto y La Tirimbina.



**Figura 1.1.** Parcelas permanentes evaluadas en el bosque primario en Corinto: T parcelas sin aprovechamiento forestal; A parcelas con aprovechamiento forestal; S parcelas con aprovechamiento forestal y tratamiento silvicultural. Parcelas permanentes evaluadas en el bosque primario en La Tirimbina: T parcelas con aprovechamiento forestal, L parcelas con aprovechamiento y tratamiento silvicultural, D parcelas con aprovechamiento forestal y dosel protector (no fueron muestreadas). La reducción de área basal generada por los aprovechamientos y los tratamientos silviculturales forma un gradiente denominado para este estudio como grado de perturbación, el cual tiene una reducción máxima del 40%.

Debido a la compleja dinámica existente en los bosques tropicales, a las múltiples respuestas de las especies a la perturbación (natural o antrópica), y a la gran diversidad taxonómica, muchos autores justifican la necesidad de agrupar las especies en grupos funcionales de plantas (GF) de acuerdo a sus rasgos funcionales (Díaz *et al.* 1999; Cornelissen *et al.* 2003; Nygaard y Ejrnaes 2004). Estos grupos pueden ser definidos como conjuntos de especies que tienen comportamientos similares a nivel del organismo, responden de manera similar a un factor determinado, y/o producen un efecto similar en el ecosistema, los cuales se caracterizan por tener atributos biológicos comunes que correlacionan su comportamiento (Lavorel *et al.* 1998, Lavorel y Garnier 2002).

Al comenzar el estudio se identificaron tres grupos funcionales de plantas basados en un conjunto de variables (tipo de dispersión, agente polinizador y sistema sexual, estrato, tasa de crecimiento, etc.) usando técnicas multivariadas de agrupamiento. Es decir, un grupo funcional se definió como un conjunto de especies vegetales que son similares en cuanto a su “rol” en el ecosistema.

Teniendo identificados estos grupos se quiere evaluar el comportamiento de la variable de respuesta *clase diamétrica* (número de individuos en las clases diamétricas definidas anteriormente) en función del tiempo. El tiempo de medición varió de acuerdo al sitio y la parcela, hay años que no se midieron todas las parcelas (Cuadro 1). El grado de perturbación se definió como la reducción del área basal generada por el aprovechamiento y los tratamientos silviculturales (Figura 1.1, Tabla 1.4). En este estudio hubo reducciones entre 0% y 40%.

Se midieron 8 variables respuesta: i) área basal, ii) número de individuos, iii) número de muertos, iv) número de reclutas, v) número de individuos en las clases diamétricas, vi) número de especies, vii) número de especies raras, y viii) número de individuos de especies raras.

En la presente investigación se realizará la modelación de la variable “*número de individuos por Clase Diamétrica*” (CD).

**Tabla 1.1.** Años de medición para las parcelas permanentes de La Tirimbina.  
(SM=sin medición).

Parcela	AÑOS DE MEDICIÓN						
	1992	1993	1994	1996	1998	2002	2003
1	SM	X	X	X	X	X	X
2	X	X	X	X	X	X	X
3	X	X	X	X	X	SM	X
4	X	X	X	X	X	SM	X
5	X	X	X	X	X	X	X
6	SM	X	X	X	X	X	X
7	SM	X	X	X	X	X	X
8	X	X	X	X	X	SM	X
9	X	X	X	X	X	X	X

**Tabla 1.2.** Años de medición para las parcelas permanentes de Corinto.  
(SM=sin medición).

Parcela	AÑOS DE MEDICIÓN							
	1992	1993	1994	1995	1996	1997	1998	2000
1	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X
4	X	X	X	X	X	X	X	X
5	X	X	X	SM	X	X	X	X
6	X	X	X	SM	X	X	X	X
7	X	X	SM	X	X	X	X	X
8	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	X

En la Figura 1.2 y la tabla 1.3 se detalla y caracteriza el área de estudio.

Figura 1.2 Mapa de la región de estudio

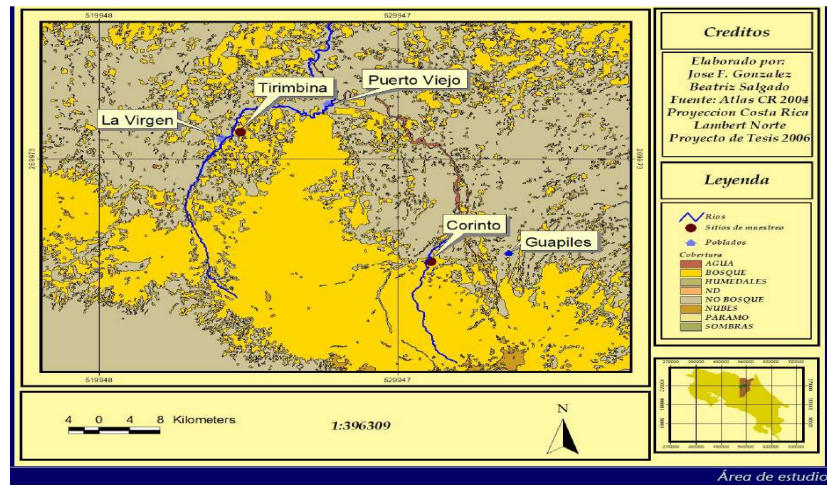


Tabla 1.3. Caracterización del área de estudio

	TIRIMBINA	CORINTO
Precipitación	3864 mm/año	4000 mm
Altitud	180 – 220 metros sobre el nivel del mar	235 – 345 metros sobre el nivel del mar
Temperatura	24.5 C°	23.7 C°
Zona de vida	Bosque muy húmedo tropical (bmh_T)	Bosque muy húmedo tropical (bmh_T)
Paisaje	Colinas bajas de origen volcánico	Topografía plana a accidentada
Área	80 ha	260 ha
Bosque aprovechado	29.16 ha	30 ha
Referencias	Mata 1997; Delgado et al. 1997;	Quirós y Finegan 1994; Carrera et al. 1996; Quirós 1998

Tabla 1.4. Reducción de área basal en las parcelas de Tirimbina y Corinto.

TIRIMBINA		CORINTO	
Parcela	% Reducción área basal	Parcela	% Reducción área basal
2	23.06	1	8.31
3	34.89	2	0
4	20.07	3	21.18
5	28.39	4	5.94
8	28.45	5	3.58
9	39.68	6	29.97
		7	0
		8	0
		9	7.95

**Objetivo general:**

Analizar y evaluar la aplicabilidad de los Modelos Lineales Generalizados Mixtos en la modelación de una variable de respuesta ordinal teniendo en cuenta la naturaleza longitudinal de los datos.

**Objetivos específicos:**

- Aplicar técnicas de análisis multivariado para la caracterización de la variable respuesta ordinal *clase diamétrica* en función de los tratamientos y la perturbación de tres grupos funcionales de plantas.
- Identificar las variables explicativas y las interacciones que más contribuyen en la descripción de un modelo adecuado para el análisis de las observaciones.
- Estudiar la aplicabilidad del *Modelo Logit Acumulado Mixto Generalizado* para la variable de respuesta *clase diamétrica* a través del tiempo.
- Analizar la proporcionalidad en el modelo con enlace Logit Acumulado.
- Aplicar el modelo Generalizado Mixto no proporcional.
- Analizar los aportes de los Modelos Mixtos en el análisis de datos con estructura longitudinal.

## CAPITULO II

### 2. Marco Teórico

#### 2.1. Modelo Lineal Generalizado con efectos fijos

Un Modelo Lineal Generalizado (MLG) es un modelo que vincula la variable de respuesta (dependiente) con las variables explicativas (covariables), mediante tres componentes (McCulloch C. y Searle S., 2001):

- La componente aleatoria (la distribución de probabilidad de las  $Y_i$ ). En general, se supone que las  $Y_i$  son independientes, con una distribución de probabilidad dentro de la familia exponencial lineal (por ejemplo: normal, binomial, poisson, gamma, multinomial, etc.)
- La componente sistemática, que indica la relación entre las variables independientes.  $\eta = x_i' \beta$
- La función de enlace, que es la que vincula los valores esperados de las  $Y_i$  con la componente sistemática. Dicha función debe ser monótona.

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

Es sólo a través de la función de enlace aplicada al predictor lineal que las covariables ejercen efecto sobre la variable respuesta. El enlace canónico (una posibilidad que se deduce desde la formulación dentro de la familia exponencial) proporciona una escala adecuada en la cual el valor esperado se puede describir mediante una función lineal de los parámetros. McCullagh y Nelder (1989) plantean que la elección de la escala es un aspecto importante en la selección del modelo. La escala adecuada para el análisis clásico de regresión lineal, por ejemplo, es aquella que combina varianzas constante, errores independientes distribuidos aproximadamente normal y aditividad en la componente sistemática. Sin embargo, con la introducción de los MLG tanto la normalidad como la varianzas constante ya no son necesarias y la aditividad de



los efectos, aunque importante, puede lograrse en otra escala. Lo que debe conocerse es la manera en que la varianza depende de la media, esto es, la relación entre la media y la varianza establecida por el modelo probabilístico seleccionado, mientras que la linealidad se postula como propiedad de una cierta función de la esperanza de la variable respuesta.

**Componente aleatoria**

Sea Y una variable de respuesta que tiene función de densidad dentro de la Familia exponencial:

$$f(y; \theta, \varphi) = \exp \left\{ \frac{[y\theta - \psi(\theta)]}{a\varphi} + c(y, \varphi) \right\} \quad (2.1)$$

El parámetro  $\theta$  se denomina parámetro “natural” o de localización (no necesariamente la media) y el parámetro  $\varphi$  es el parámetro de dispersión (o de escala), mientras que  $\psi(\cdot)$  y  $c(\cdot)$  son funciones conocidas.

Si la función de densidad tiene un solo parámetro  $\theta$ , la ecuación anterior se puede simplificar de la siguiente manera:

$$f(y; \theta) = \exp \left\{ \frac{[y\theta - \psi(\theta)]}{a} + c(y, \cdot) \right\}$$

Las expresiones para  $a$ ,  $\psi(\cdot)$  y  $c(\cdot)$  para algunas distribuciones se muestran en la tabla 2.1. La media y la varianza de la distribución  $f(y; \theta)$  pueden escribirse en términos de  $a$  y  $\psi(\cdot)$ .

**Tabla 2.1.** Expresiones para  $a$ ,  $\psi(\cdot)$  y  $c(\cdot)$  para distribuciones de un solo parámetro.

Distribución	$a$	$\psi(\cdot)$	$c(\cdot)$
Bernulli	1	$\log(1+\exp(\theta))$	1
Binomial	1/n	$\log(1+\exp(\theta))$	$\log \left[ \frac{n!}{(ny)!(n-by)!} \right]$
Poisson	1	$\exp(\theta)$	$-\log(y!)$

### Componente sistemática:

La componente sistemática de un MLG se define con la siguiente ecuación:

$$g(\mu_i) = g\{E(Y_i)\} = X_i' \beta = \eta \quad (2.3)$$

Donde  $Y_i$  es la variable de respuesta para el  $i$ -ésimo sujeto,  $g(\cdot)$  es una función monótona y diferenciable, llamada función de enlace,  $\beta$  es un vector de parámetros relacionados con las covariables y  $x_i$  las covariables.

### Función de enlace

La función de enlace sirve para restringir los posibles valores de la media, ya que la componente lineal podría tomar cualquier valor real, mientras que la media en muchos casos debe ser positiva, o estar entre 0 y 1, etc. Además es una transformación de las medias en la que la relación con los regresores es lineal. En la representación de la familia exponencial,  $\theta(\mu)$  es la función de enlace canónica.

Es importante notar que no siempre es deseable usar enlace canónico. Por ejemplo, si se consideran datos binarios o proporciones binomiales, la función de enlace canónica es:

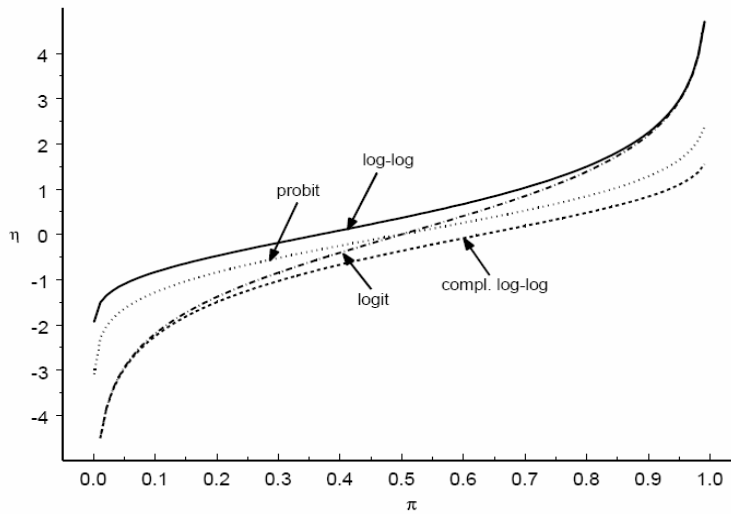
$$g(\mu) = \text{logit}(\mu) = \log \frac{\mu}{1-\mu} = x_i' \beta, \quad g^{-1}(x_i' \beta) = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)} \quad (2.4)$$

Esta función de enlace es la inversa de la función de distribución de una variable logística. Como la inversa de cualquier función de distribución continua (fdc) es monótona, se han propuesto otras funciones de enlace:

- Probit:  $g(\mu) = \text{probit}(\mu) = \varphi^{-1}(\mu) = x_i' \beta$ ,  $g^{-1}(x_i' \beta) = \varphi(x_i' \beta)$  (fdc de una normal).
- Log-log:  $g(\mu) = \log(-\log(\mu)) = x_i' \beta$ ,  $g^{-1}(x_i' \beta) = \exp(-\exp(x_i' \beta))$  (fdc de una distribución de valor extremo de tipo I, o Gumbel).
- Log-log complementario:

$$g(\mu) = \log(-\log(1-\mu)) = x_i' \beta, \quad g^{-1}(x_i' \beta) = 1 - \exp(-\exp(x_i' \beta))$$

**Figura 2.1.** Representación de distintas Funciones de enlace.



## 2.2 Modelos Logit

Una de las aplicaciones más extendidas del modelo lineal clásico tiene lugar cuando el investigador dispone de una variable de respuesta cuantitativa cuya variación desea explicar recurriendo a un conjunto de variables explicativas (experimentales y de clasificación).

Para modelar datos binarios (dos categorías) que provengan de variables aleatorias distribuidas según una distribución binomial, se consideran modelos de regresión logística que relacionen el parámetro  $\pi$  (que es la proporción esperada de “éxitos”) con la(s) variable(s) explicativas. El predictor lineal más simple sería:

$$\pi_i = \alpha + x_i' \beta \quad (2.5)$$

El cual define a la función de enlace como la identidad. Este modelo que iguala la probabilidad de éxito a una función lineal (llamado algunas veces modelo de probabilidad lineal) tiene algunos problemas serios en su aplicabilidad: los valores predichos no están acotados, dado que el miembro derecho tiene imagen en los *Reales*, aunque es obvio que deberían estar siempre entre 0 y 1. Cuando se utiliza como función de enlace el “logit”, modelo de regresión logística, los valores predichos queden siempre entre 0 y 1 (Díaz, 2008):

$$\text{logit}(\pi_i(x)) = \log\left(\frac{\pi_i(x)}{1 - \pi_i(x)}\right) = \alpha + x_i' \beta \quad (2.6)$$

$$\pi_i(x) = \frac{\exp(\alpha + x_i' \beta)}{1 + \exp(\alpha + x_i' \beta)} \quad (2.7)$$

En el ejemplo (ecuaciones 2.6 y 2.7) se propuso como componente sistemática una regresión lineal simple, pero se podría utilizar cualquier otro modelo lineal.

El coeficiente de regresión  $\beta$  se interpreta de la siguiente manera:

- tasa de crecimiento: a medida que  $|\beta|$  crece,  $\pi$  cambia más rápidamente a medida que  $x$  se incrementa,
- la pendiente de la curva en cualquier punto es  $\beta = \pi(1 - \pi)$ , y la pendiente es máxima cuando  $\pi = 0.5$ .
- el valor de  $x$  cuando la pendiente es máxima  $x = -\alpha/\beta$  (también llamado como punto de inflexión o valor efectivo mediano o en estudios de graduación dosis letal 50, denotado comúnmente por LD50), es el valor  $\exp(\beta)$  representa el cociente de chances (odds-ratio) cuando las  $x$  aumentan en una unidad.

La interpretación como cociente de chances es la que hace muy atractiva esta función de enlace (que se prefiere a otras como probit o complemento log-log). La propiedad simétrica del cociente de chances nos permite usarlo tanto en estudios prospectivos como en retrospectivos, (caso-control, etc.), y por lo que se torna muy útil en estudios de Epidemiología y Ecología (Díaz, 2008).

La estimación de los parámetros del modelo se realiza por máxima verosimilitud, de manera que para la inferencia es posible usar todas las propiedades de los estimadores máximos verosímiles. En general, las propiedades son asintóticas, por lo que para que se cumplan aproximadamente bien en una situación particular se necesitan tamaños de muestra grandes. Computacionalmente, en los MLG, los estimadores se obtienen usando el algoritmo de Newton-Raphson o puntuación de Fisher (o algún refinamiento de este algoritmo).

### 2.3 Modelo Logit Multicategórico

La regresión logística clásica se aplica en casos en que la variable de respuesta es dicotómica, es decir puede tomar dos valores con probabilidad  $\pi$  o  $1 - \pi$ , cuyo modelo corresponde a la ecuación (2.6).

Análogamente se puede utilizar para una variable categórica  $Y$  con  $J$  categorías que tomen valores con probabilidades:  $\{\pi_1, \pi_2, \dots, \pi_J\}$  y  $\sum_j \pi_j(x) = 1$ .

Los modelos logits asociados a variables con más de una categoría se conocen como Modelos Logit Multicategóricos (Díaz, 2008).

Para respuestas multicategóricas es posible modelar un MLG multinomial de la siguiente forma:

$$\text{logit}(\pi_j) = \log\left(\frac{\pi_j}{\pi_{j-1}}\right) = \alpha_j + x' \beta_j \quad (2.8)$$

Con:  $J$  categorías y  $j: 1, 2, \dots, J-1$

Donde se pueden definir las componentes de un MLG de la siguiente manera:

$$Y_i \sim M(n, \pi_1, \dots, \pi_{J-1})$$

$$g(\pi_j) = \text{logit}(\pi_j)$$

$$\eta = x' \beta_j$$

Con:  $J$  categorías y  $j: 1, 2, \dots, J-1$

Se puede demostrar que la Distribución Multinomial, cuya única restricción sobre los parámetros es que la suma de probabilidades de todos ellos sea igual a 1, pertenece a una familia exponencial de dimensión a lo sumo  $m = J-1$ .

Las características de un MLG con variable de respuesta univariada en la familia exponencial se extienden para un MLG multivariado con distribución multivariada de la familia exponencial, tal es el caso de la Multinomial.

Cuando la variable de respuesta es multinomial, es necesario identificar la naturaleza de dicha variable. Puede tratarse de una Variable Nominal, donde el orden de las categorías es irrelevante o una Variable Ordinal donde si nos interesa el orden de las categorías. El análisis difiere según esta naturaleza.

### 2.3.1 Modelo Logit para variables Nominales

Sea  $Y$  una variable categórica con  $J$  categorías. Un modelo Logit Multicategórico para respuesta nominal describe simultáneamente de a pares los log-odds para todas las combinaciones de las  $J$  categorías. Dada una cierta elección  $J-1$  de estos pares, el resto son redundantes.

$$\pi_j(x) = P(Y = j / x) \quad \sum_j \pi_j(x) = 1$$

Para las observaciones del conjunto de variables explicativas  $x$ , tratamos a los conteos de las  $J$  categorías de  $Y$  como multinomial con probabilidades  $\{\pi_1(x), \pi_2(x), \dots, \pi_J(x)\}$ .

Si el orden de las categorías no interesa el modelo que se ajusta se denomina Modelo Logit Generalizado o Modelo Logit con Categoría de Referencia (Baseline category) (Agresti, 2002).

#### Modelo Logit de Categoría de Referencia

En la construcción de los logits se aparea cada categoría de respuesta con una categoría de referencia, generalmente la última o la más frecuente.

El modelo:

$$\text{logit}(\pi_j) = \log\left(\frac{\pi_j(x)}{\pi_J(x)}\right) = \alpha_j + x' \beta_j \quad \text{Con } j = 1, \dots, J-1 \quad (2.9)$$

Describe simultáneamente los efectos de  $x$  en los  $J-1$  logits. Los efectos varían de acuerdo a la variable de respuesta teniendo como base una categoría de referencia. Estas  $J-1$  ecuaciones determinan los parámetros para logits de las otras categorías de respuesta:

$$\log\left(\frac{\pi_a(x)}{\pi_b(x)}\right) = \log\left(\frac{\pi_a(x)}{\pi_J(x)}\right) - \log\left(\frac{\pi_b(x)}{\pi_J(x)}\right)$$

Para una variable con  $J=3$  categorías y tomando como referencia a la última, las ecuaciones logits que se obtienen son:

$$\text{logit}(\pi_1) = \log\left(\frac{\pi_1(x)}{\pi_3(x)}\right) = \alpha_1 + x' \beta_1$$

Aquí es fácil  $\text{logit}(\pi_2) = \log\left(\frac{\pi_2(x)}{\pi_3(x)}\right) = \alpha_2 + x' \beta_2$  ue  $J=2$ , se tiene un modelo Logit

convencional:

$$\text{logit}(\pi_1) = \log\left(\frac{\pi_1(x)}{\pi_2(x)}\right) = \log\left(\frac{\pi_1(x)}{1-\pi_1(x)}\right) = \alpha_1 + x'\beta_1$$

### 2.3.2 Modelo Logit para variables Ordinales

Para el modelado de variables de respuesta ordinales, es decir una variable de respuesta que posee  $k$  categorías ordinales, es posible definir  $k-1$  logits acumulados de acuerdo con una estructura de partición sugerida por la naturaleza de la variable de respuesta o de los intereses de la investigación.

Una manera de utilizar la ordinalidad de la variable de respuesta es aplicando el modelo Log-Acumulado o modelo logit de probabilidades acumuladas:

$$P(Y \leq j/x) = \pi_1(x) + \dots + \pi_j(x); \quad j = 1, \dots, J \quad (2.10)$$

Los logits Acumulados son definidos como:

$$\begin{aligned} \text{logit}[P(Y \leq j/x)] &= \log \frac{P(Y \leq j/x)}{1 - P(Y \leq j/x)} \\ &= \log \frac{\pi_1(x) + \dots + \pi_j(x)}{\pi_{j+1}(x) + \dots + \pi_J(x)} \quad j = 1, \dots, J-1 \end{aligned} \quad (2.11)$$

Cada Logit Acumulado utiliza todas las  $J$  categorías de la variable de respuesta.

A través del enfoque de modelos lineales generalizados se pueden definir las componentes de la siguiente manera:

La variable aleatoria de respuesta  $Y_i$ . Variable de naturaleza ordinal con  $i$  - categorías.

$$Y_i \sim M(n, \pi_1, \dots, \pi_{J-1})$$

$$g(\pi_j) = \text{logit}(\pi_j)$$

$$\eta = x'\beta_j$$

**Modelo de Odds Proporcional:**

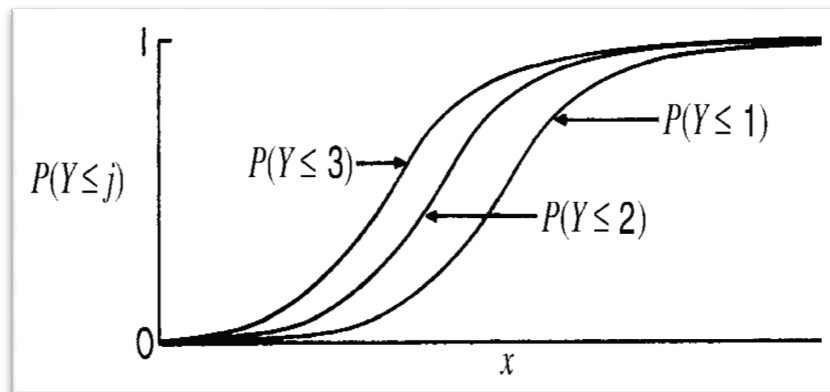
Este modelo que usa simultáneamente todos los logits acumulados, tiene una única pendiente  $\beta$  para todas las categorías:

$$\text{logit}[P(Y \leq j/x)] = \alpha_j + x' \beta, \quad j = 1, \dots, J-1 \quad (2.12)$$

Cada logit acumulado tiene su propio intercepto. Los  $\{\alpha_j\}$  van aumentando en  $j$  dado que las  $[P(Y \leq j/x)]$  aumentan en  $j$  para un valor fijo de  $x$ , es decir que el logit es una función creciente de esas probabilidades (Agresti,2002).

Este modelo tiene la misma pendiente  $\beta$  para todos los efectos para cada logit. Para un predictor continuo  $x$ , la figura 2.2 muestra el modelo cuando se tienen 4 categorías en la variable de respuesta.

**Figura 2.2.** Modelo logit acumulado odds – proporcional.

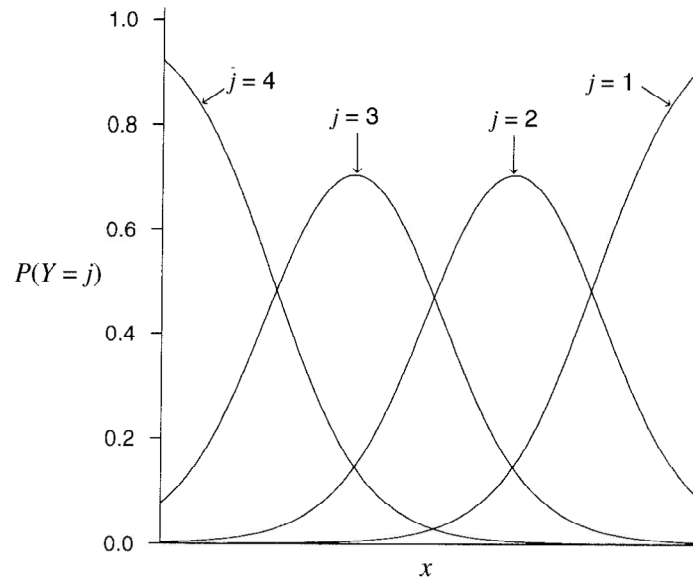


Para un  $j$  fijo, la curva de respuesta es una curva de regresión logística para una respuesta binaria con resultados  $Y \leq j$  e  $Y > j$ . Las curvas de respuesta para  $J= 1, 2, 3$  y  $4$  tienen la misma forma es decir la misma pendiente. Ellas comparten exactamente la misma tasa de aumento o disminución, pero están desplazados horizontalmente entre sí. Para  $j < k$ , la curva para  $P(Y \leq k)$  es la curva para  $P(Y \leq j)$  pero desplazada  $(\alpha_k - \alpha_j)/\beta$  unidades en  $x$ ; es decir:

$$P(Y \leq k|X = x) = P(Y \leq j|X = x + (\alpha_k - \alpha_j)/\beta)$$



**Figura 2.3.** Probabilidades de las Categorías en modelo logit acumulado



El modelo Logit acumulado (2.12) satisface:

$$\text{logit}[P(Y \leq j / x_1)] - \text{logit}[P(Y \leq j / x_2)] = \log \frac{P(Y \leq j / x_1) / P(Y > j / x_1)}{P(Y \leq j / x_2) / P(Y > j / x_2)} = (x_1 - x_2)' \beta$$

A este cociente de chances se lo denomina Odds Ratio Acumulado. El odds ratio para  $Y \leq j$  en  $x = x_1$  es  $\exp(\beta'(x_1 - x_2))$  veces el odds en  $x = x_2$ . El odds ratio del log –acumulado es proporcional a la distancia entre  $x_1$  y  $x_2$ . La misma constante de proporcionalidad se aplica a cada logit. Por esta propiedad la ecuación 2.12 se llama modelo de Odds proporcional (McCulloch, 1997).

### 2.3.3 Estimación

El ajuste por Máxima Verosimilitud de logits multinomiales maximiza la verosimilitud simultáneamente satisfaciendo las  $J-1$  ecuaciones del modelo. Para  $i=1, \dots, n$  sean  $y_i = (y_{i1}, \dots, y_{iJ})$  los ensayos multinomiales para el sujeto  $i$ , donde  $y_{ij} = 1$  si la respuesta corresponde a la categoría  $j$  y es igual a 0 en otro caso. Sean  $x_i = (x_{i1}, \dots, x_{ip})'$  valores de la variable independiente y  $\beta_j = (\beta_{j1}, \dots, \beta_{jp})'$  los parámetros para el  $j$ -ésimo logit.

La contribución a la Verosimilitud del sujeto  $i$  es (Diggle et al., 2002):

$$\log \left[ \prod_{j=1}^J \pi_j(x_i)^{y_{ij}} \right] = \sum_{j=1}^{J-1} y_{ij} \log \pi_j(x_i) + \left( 1 - \sum_{j=1}^{J-1} y_{ij} \right) \log \left[ 1 - \sum_{j=1}^{J-1} \pi_j(x_i) \right]$$

$$= \sum_{j=1}^{J-1} y_{ij} \log \frac{\pi_j(x_i)}{1 - \sum_{j=1}^{J-1} \pi_j(x_i)} + \log \left[ 1 - \sum_{j=1}^{J-1} \pi_j(x_i) \right] \quad (2.13)$$

Asumiendo  $n$  observaciones independientes y sustituyendo en la expresión anterior el predictor lineal para el logit del primer término y a  $\pi_j(x_i)$  por su expresión en el segundo término, el logaritmo de la Verosimilitud es:

$$\begin{aligned} \log \prod_{i=1}^n \left[ \prod_{j=1}^n \pi_j(x_i)^{y_{ij}} \right] &= \sum_{i=1}^n \left\{ \sum_{j=1}^{J-1} y_{ij} (\alpha_j + x'_i \beta_j) - \log \left[ 1 + \sum_{j=1}^{J-1} \exp(\alpha_j + x'_i \beta_j) \right] \right\} = \\ &= \sum_{i=1}^{J-1} \left[ \alpha_j \left( \sum_{i=1}^n y_{ij} \right) + \sum_{k=1}^p \beta_{jk} \left( \sum_{i=1}^n x_{ik} y_{ij} \right) \right] - \sum_{i=1}^n \log \left[ 1 + \sum_{j=1}^{J-1} \exp(\alpha_j + x'_i \beta_j) \right] \end{aligned} \quad (2.14)$$

Las ecuaciones de verosimilitud igualan los estadísticos suficientes a sus valores esperados. El Logaritmo de la Verosimilitud es cóncavo y el método de Newton -Raphson genera estimaciones máximo verosímiles de los parámetros. Los estimadores tienen distribuciones normales para muestras grandes y sus errores estándar asintóticos son la raíz cuadrada de los elementos diagonales de la inversa de la matriz de información.

Las  $J - 1$  ecuaciones determinan los parámetros para logits con otros pares de categorías. Por ejemplo:

$$\begin{aligned} \log \frac{\pi_a(x)}{\pi_b(x)} &= \log \frac{\pi_a(x)}{\pi_j(x)} - \log \frac{\pi_b(x)}{\pi_j(x)} \\ &= (\alpha_a + x' \beta_a) - (\alpha_b + x' \beta_b) \\ &= (\alpha_a - \alpha_b) + x' (\beta_a - \beta_b) \end{aligned}$$

## 2.4 Modelo Lineal Mixto para variables de respuesta con distribución normal

El modelo lineal generalizado se puede extender a situaciones donde se incorporan efectos aleatorios: modelo lineal mixto. En esta sección se presenta un modelo lineal mixto para variables de respuesta con distribución normal.

En primer lugar, se considera el siguiente Modelo Lineal:

$$Y_i = X_i' \beta + \varepsilon_i \quad (2.15)$$

Donde  $\varepsilon_i$  es un vector  $n_i \times 1$  normalmente distribuido con media igual a cero. Se completa el modelo mediante la especificación de una matriz de covariancia  $\Sigma_i$  apropiada para los  $\varepsilon_i$ , que conduce al siguiente modelo multivariado:

$$Y_i \sim N(X_i' \beta, \Sigma_i) \quad (2.16)$$

Cuando  $\Sigma = \sigma^2 I_i$ , este modelo corresponde al modelo de regresión lineal clásico, que asume independencia de los errores. Con esta estructura de covarianza no es posible modelar mediciones repetidas sobre individuos. Para ese caso es posible seleccionar otra estructura, como por ejemplo la correlación uniforme o autorregresiva de primer orden.

El modelo de regresión lineal univariado puede extenderse con el enfoque de efectos aleatorios modelando una regresión con coeficientes sujeto específico. En muchas situaciones, se puede suponer que cada sujeto responde de manera diferente a través del tiempo y, en consecuencia, cada uno debe tener un intercepto y una pendiente. Por lo tanto, se puede enunciar el siguiente modelo (Davidian y Giltinan, 1995).

$$Y_i = X_i' \beta_i + Z_i b_i + e_i \quad i=1, \dots, N \quad (2.17)$$

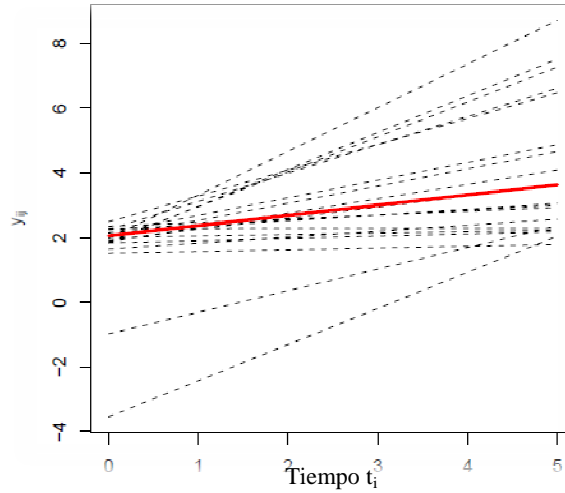
Donde  $y_i$  es un vector ( $n_i \times 1$ ) de respuestas continuas del  $i$  – ésimo sujeto o unidad experimental con distribución normal,  $n_i$  es el número de mediciones

repetidas en el tiempo  $i$  – ésimo sujeto,  $X_i$  es la matriz de diseño de orden  $(n_i \times p)$  que caracteriza la parte sistemática del diseño donde  $p$  es la cantidad de efectos fijos del modelo;  $\beta_i$  es el vector de parámetros de los efectos fijos del modelo que completan la parte sistemática del mismo.  $Z_i$  también es una “matriz de diseño” de orden  $(n_i \times q)$  que caracteriza la parte aleatoria del modelo atribuible a la fuente de variación existente entre los individuos,  $q$  es el número de efectos aleatorios en el modelo;  $b_i$  es un vector  $(q \times 1)$  de efectos aleatorios,  $e_i$  es un vector de errores aleatorios concerniente a la variación “dentro” de cada individuo debida a la forma en que fue medida la variable de respuesta en cada individuo,  $b_i$  y  $e_i$  son las fuentes de variación presentes en el modelo y describen la variación entre y dentro de los individuos respectivamente. La distribución de los efectos aleatorios y los residuales se asumen normales,  $b_i \sim N(0, D)$  y  $e_i \sim N(0, S_i)$ .  $D$  se define como la matriz de covariancias de orden  $(q \times q)$  que describe la variabilidad entre los sujetos y la matriz  $S_i$  de orden  $(n_i \times n_i)$  es la matriz de covariancia que describe la fuente de variabilidad dentro de los sujetos. La matriz  $S_i$  permite distintas estructuras que dependen de los individuos. Además  $b_i$  es independiente de  $e_i$  (Verbeke y Molenberghs, 2000).

Bajo este modelo, es razonable pensar en términos del perfil de promedio poblacional que representa el modelo de regresión lineal para el sujeto promedio.

La siguiente figura representa los perfiles sujeto específico y el perfil promedio poblacional en un ejemplo hipotético de datos longitudinales que pueden describirse mediante un modelo mixto lineal con intercepto y pendiente aleatorias.

**Figura 2.4.** Perfiles Sujeto específico y promedio poblacional. La línea continua representa la evolución de la promedio poblacional y las líneas punteadas indican la evolución sujeto específico.



El modelo presentado en la ecuación (2.17) es un caso especial del *Modelo Lineal Mixto* que asume que el vector  $Y_i$  de medidas repetidas del  $i$ -ésimo sujeto satisface:

$$Y_i | b_i \sim N(x_i' \beta + Z_i b_i, \Sigma_i) \quad b_i \sim N(0, D) \quad (2.18)$$

Los componentes de  $\beta$  son llamados efectos fijos, y los componentes en  $b_i$  se denominan efectos aleatorios.

## **2.5 Medidas Repetidas - Datos Longitudinales**

La característica fundamental de los estudios longitudinales es que siempre se cuenta con un conjunto de observaciones tomadas secuencialmente sobre la misma unidad (persona, animal, parcela, sujeto...). Los datos longitudinales también se denominan “mediciones repetidas” especialmente en el contexto experimental. Por el contrario los estudios trasversales obtienen una sola medición para cada individuo.

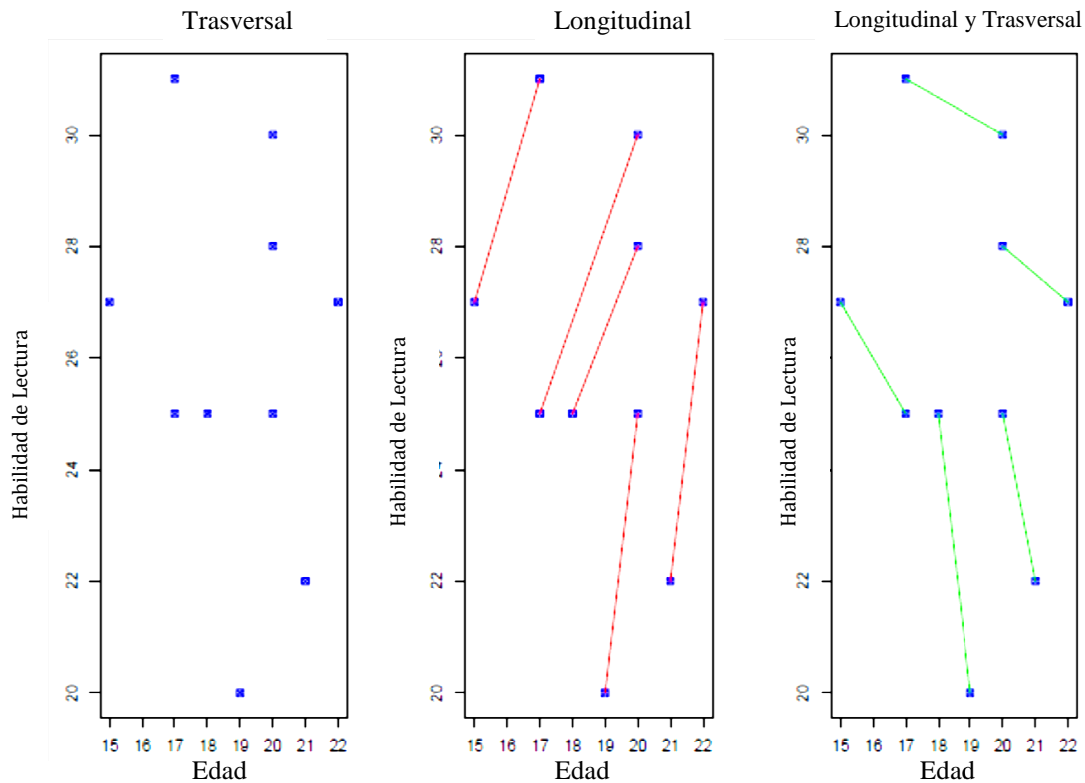
Existen grandes ventajas en el uso de los datos longitudinales, la primera o la principal es la eficacia para el estudio de cambios, es decir, los estudios longitudinales tienen la capacidad de separar, en el contexto de la población estudiada, los efectos dentro de cada individuo de los efectos de cohorte o de diferencias entre individuos; principalmente un investigador puede separar el efecto del tiempo (cambios de los individuos a través del tiempo) de los efectos de cohorte (diferencias entre los individuos al inicio del estudio).

El contexto de Modelos Mixtos brinda herramientas poderosas para modelar datos longitudinales, principalmente a través de la matriz de correlaciones, que permite formular explícitamente las correlaciones existentes (Hartzel et al., 2001).

Para comprender mejor este aspecto, la Figura 2.4 muestra un ejemplo clásico presentado por Diggle et al. (2002). Los gráficos representan la capacidad de lectura versus la edad para un estudio hipotético de niños. En el primer gráfico (estudio trasversal) se puede observar que la capacidad de lectura parece más pobre entre los sujetos de mayor edad. En este gráfico sólo se muestra la relación entre la edad y la capacidad de lectura, sin considerar las medidas repetidas. En el segundo gráfico (longitudinal), se supone que los mismos datos fueron obtenidos en un estudio longitudinal en el que cada individuo se midió en dos ocasiones. Es claro que mientras más jóvenes comienzan a leer alcanzan un nivel superior de lectura, todos mejoran con el tiempo. Sin embargo, el tercer gráfico (transversal y longitudinal), muestra que la capacidad de lectura se deteriora con la edad, es decir que si un individuo comienza a leer a una edad mayor la habilidad de lectura se deteriora. En este gráfico, se han

conectado los pares de puntos para mostrar este resultado inusual. El ejemplo muestra la importancia de los efectos del tiempo en los estudios longitudinales.

**Figura 2.4.** Relación entre la habilidad de lectura vs la edad.



La segunda ventaja es que, en un estudio longitudinal cada sujeto puede ser su propio control. Para la mayoría de los resultados, existe una considerable variabilidad entre los individuos, debido a la influencia de las características no medidas, como la exposición al medio ambiente, los hábitos personales, y otros. Esta tendencia persiste a lo largo del tiempo. Una tercera ventaja es que, para un número determinado de observaciones, los estudios longitudinales son más poderosos que la información transversal. Esto sucede debido a que las reiteradas observaciones del mismo sujeto rara vez son perfectamente correlacionados. Es decir, las mediciones repetidas de un sólo sujeto proporcionan más información que una sola medición obtenida a partir de un sólo sujeto. En cuarto lugar, los datos longitudinales pueden proporcionar información sobre el cambio individual. Las estimaciones de las tendencias individuales se pueden utilizar para comprender mejor la heterogeneidad de la población, los determinantes del crecimiento y el cambio a nivel

individual(Diggle et al., 2002).

Por otra parte, para el análisis de datos longitudinales, es necesario considerar que las observaciones reiteradas sobre un sujeto no son independientes, por lo tanto, este análisis requiere métodos estadísticos especiales. Se debe considerar la correlación existente para poder realizar inferencias válidas, añadiendo complejidad a los modelos.

Hay varios aspectos de los estudios longitudinales que se deben considerar para seleccionar un análisis apropiado (Herdeker, 2005):

- Es importante saber si la respuesta es continua o discreta. El modelo seleccionado dependerá de la naturaleza de la variable de respuesta y de su distribución de probabilidad. Para las respuestas continuas, existen diversos modelos, técnicas de estimación y software de estadística. Sin embargo, para los resultados discretos, los modelos y las técnicas de estimación son más complejos, y aún están siendo desarrolladas en la actualidad. Además, los programas estadísticos para el ajuste de estos modelos son limitados, y las rutinas para la estimación de parámetros pueden ser computacionalmente intensivas, debido a la necesidad de la utilización de métodos numéricos.
- El tamaño de la muestra ( $N$ ), y el número de observaciones por sujeto ( $n_i$ ) son también aspectos importantes en un estudio longitudinal. En modelos complejos, como los modelos lineales generalizados mixtos, si  $N$  y  $n_i$  son pequeñas pueden producirse estimaciones inadecuadas.
- Se debe tener en cuenta la selección de una estructura de covarianza plausible para la  $\text{Var}(Y)$  ya que diferentes especificaciones del modelo pueden conducir a covarianzas homogéneas o heterogéneas de medidas repetidas en el tiempo (Diggle et al., 2002).

A fin de presentar un conjunto de modelos para analizar los datos longitudinales, es necesario entender cuáles son las probables fuentes de variación aleatoria que se producen en la práctica, y cuáles pueden ser importantes para ser incluidas en los modelos (Diggle et al., 2002). Estas



fuentes permiten desarrollar y estudiar cada elemento del modelo. Es posible distinguir las siguientes fuentes de variación aleatoria en un estudio longitudinal:

- Efectos aleatorios: Cuando los sujetos son seleccionados al azar de una población, varios aspectos de su comportamiento pueden mostrar la variación estocástica entre los sujetos.
- Correlación serial: La respuesta del sujeto puede estar influenciada por variaciones en el tiempo de un proceso estocástico dentro de este sujeto. Este tipo de resultados entre pares de mediciones en el mismo sujeto depende del intervalo de tiempo entre las mediciones. Por lo general, la correlación se vuelve más débil a medida que aumenta el tiempo de separación.
- Error de medición: Sobre todo cuando la medición individual involucra algún tipo de muestreo dentro de las unidades, a la medición se le puede añadir una componente de variación de los datos.

Estas características cualitativas deben ser introducidas en los modelos mediante el uso de términos apropiados.

### 2.5.1 Notación en observaciones longitudinales

Sea un estudio longitudinal con  $N$  sujetos, con  $i = 1, \dots, N$ , cada uno medido  $n_i$  ocasiones en el tiempo. Sea  $y_{ij}$  la respuesta para el  $i$  – ésimo sujeto en la  $j$  – ésima medición,  $j=1, \dots, n_i$ .

- $Y'_i = (Y_{i1}, \dots, Y_{in_i})'$  de dimensión  $n_i \times 1$ , vector de respuesta de mediciones repetidas sobre el  $i$ –ésimo sujeto.
- $Y = (Y_1, \dots, Y_N)$  de dimensión  $n \times 1$ , conjunto completo de  $N_T = \sum_{i=1}^N n_i$  mediciones.
- $x'_{ij}$  es un vector de covariables para el  $i$ –ésimo sujeto en la  $j$ –ésima ocasión.
- $X_i$  de dimensión  $n_i \times p$ , matriz de diseño de  $p$  covariables para el  $i$ –ésimo sujeto.

- $X: N_T \times p$  Matriz de diseño que contiene la información de las covariables para todos los sujetos.
- $\beta$ : de dimensión  $p \times 1$ , vector de efectos fijos asociados a las covariables.
- $z_{ij}$  es el vector de covariables de efectos aleatorios.
- $Z_i: n_i \times q$  Matriz de diseño de  $q$  efectos aleatorios para el  $i$ -ésimo sujeto.
- $b_i$ : vector de dimensión  $q$ , de efectos aleatorios para el  $i$ -ésimo sujeto.
- $\varepsilon_i: n_i \times 1$  vector de errores para el  $i$ -ésimo sujeto.
- $D: q \times q$  Matriz de covariancias que describe la fuente de variación entre sujetos.
- $S_i: n_i \times n_i$  Matriz de covariancias que describe la fuente de variación dentro de los sujetos.

### 2.5.2 Modelos de análisis de datos longitudinales

Como se mencionó anteriormente en los Modelos Longitudinales las observaciones están correlacionadas y es necesario tener en cuenta esta correlación. Existen dos formas de hacerlo:

- De forma explícita, es decir modelando directamente la matriz de covarianza marginal.
- De forma indirecta, es decir a través de la introducción de efectos aleatorios.

Los parámetros de la parte fija del modelo son los mismos para ambas formulaciones, ya que la distribución normal multivariada tiene la propiedad de que las distribuciones marginales y condicionales siguen siendo normales, además las interpretaciones de los parámetros en la media no dependen de la estructura de correlación. Pero en Modelos Lineales Generalizados no es lo mismo considerar una distribución marginal de las observaciones que una distribución condicional de los efectos aleatorios. Para analizar estos enfoques,

a continuación se presenta una clasificación de los posibles modelos para analizar datos longitudinales en tres familias generales. En primer lugar, se deben tener en cuenta los siguientes resultados para modelos con una estructura lineal de la media (Diggle et al.,2002).

- Un modelo marginal se caracteriza por una función de la media marginal con la siguiente forma:

$$g\left(E\left(Y_{ij} \mid x_{ij}\right)\right)=x'_{ij} \beta \quad (2.19)$$

Donde  $x_{ij}$  es un vector de covariables para el i-ésimo sujeto en la j-ésima ocasión. Aunque la media está condicionada en términos de las covariables, a esta formulación se la considera marginal, porque no depende de otras variables o de los resultados anteriores. Es decir un modelo marginal se caracteriza por la especificación de la esperanza marginal en función de las covariables y la estructura de correlación marginal.

- Los modelos de efectos aleatorios, se centran en la esperanza condicional sobre un efecto aleatorio

$$g\left(E\left(Y_{ij} \mid b_{ij}, x_{ij}\right)\right)=x'_{ij} \beta+Z'_{ij} b_i \quad (2.20)$$

Evidentemente, la respuesta media depende de las covariables y los efectos aleatorios.

- Modelo de Transición: por último, hay un tercer tipo de modelo que es condicional a un resultado particular que puede ser la respuesta anterior o un subconjunto de las mismas. Un caso específico, con enlace identidad, es el llamado modelo de transición o de primer orden que toma la siguiente forma:

$$E\left(Y_{ij} \mid Y_{ij-1}, \dots, Y_{i1}, x'_{ij}\right)=x'_{ij} \beta+\alpha Y_{ij-1} \quad (2.21)$$

Los modelos anteriores tienen una clara relación cuando la respuesta tiene distribución normal. Sin embargo, no hay una relación estrecha entre ellos cuando las respuestas son discretas.

Con el fin de estudiar otras alternativas al modelado de datos longitudinales,

diferentes a las respuestas normales, se presentan tres familias de modelos que se describen brevemente:

- **Modelo Marginal:** En un modelo marginal, la regresión de la variable de respuesta sobre las variables explicativas es modelada separadamente desde la correlación dentro de cada individuo (dentro-sujeto). En la regresión se modela la esperanza marginal  $E(Y_{ij})$ , como una función de variables explicativas (Diggle et al., 2002). La esperanza marginal se entiende como el promedio de la respuesta en las subpoblaciones que comparten un valor de  $x$  y su modelado es el objetivo de un estudio transversal. Específicamente un modelo marginal tiene las siguientes hipótesis:
  - ✓ La esperanza marginal de respuesta,  $E(Y_{ij}) = \mu_{ij}$ , depende de las variables  $x_{ij}$  donde  $g(\mu_{ij}) = x'_{ij} \beta$  en el cual  $g$  es la función de enlace, como lo es el Logit para respuestas binarias o el log para variables discretas.
  - ✓ La variancia marginal depende de la media marginal de acuerdo con:  $V(Y_{ij}) = V(\mu_{ij}) \cdot \phi$ , donde  $V(\cdot)$  es una función de variancia conocida y  $\phi$  es un parámetro de escala que puede ser necesario estimarlo.
  - ✓ La correlación entre  $Y_{ij}$  y  $Y_{ik}$  es una función de las medias marginales y tal vez de parámetros adicionales  $\alpha$ , esto es la  $corr(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha)$  donde  $\rho(\cdot)$  es una función conocida.

Los coeficientes de las regresiones marginales tienen la misma interpretación que los de un estudio transversal (es decir, promedio a través de todos los individuos de la subpoblación con el mismo valor de  $x$ ). En los MLG los modelos marginales y los transversales son naturalmente análogos tanto si los datos son correlacionados o son datos independientes.

En las respuestas que son modeladas de forma marginal, es decir promediadas sobre todas las demás respuestas o resultados, la estructura de asociación es capturada utilizando un conjunto de

parámetros de asociación, como las correlaciones, los odds ratios, etc. Estos modelos también se los llama modelos de promedio poblacional ya que los parámetros caracterizan la esperanza marginal.

- **Modelo sujeto específico:** Estos modelos se diferencian de los marginales o modelo promedio poblacional, por la inclusión de parámetros específicos para cada sujeto. A diferencia de las respuestas normales, los parámetros sujeto específico y los modelos de promedio poblacional describen diferentes tipos de efectos de las covariables. El término sujeto específico se utiliza de forma equivalente al término de efectos aleatorios, sin embargo, los parámetros del modelo sujeto específico pueden ser tratados como efectos fijos (I), como efectos aleatorios (II), o eliminados a través de un condicionamiento(III). El primer caso es muy simple, pero presenta muchos problemas cuando el número de parámetros es grande. Se hará hincapié en el segundo enfoque, es decir donde pueden ser tratarlos como efectos aleatorios mediante los modelos lineales generalizados mixtos.
- **Modelos de transición:** En estos modelos cualquier respuesta dentro de la secuencia de medidas repetidas se encuentra condicionada a otros resultados previos. Es decir, bajo un modelo de transición la correlación entre  $Y_{i1}, Y_{i2}, \dots, Y_{ini}$ , existe debido a que la respuesta anterior influye explícitamente en la observación actual  $Y_{ij}$ . Las observaciones pasadas son tratadas como variables predictoras adicionales. Este enfoque no es muy utilizado ya que la interpretación de un parámetro de efectos fijos se encuentra condicionado a las demás respuestas para el mismo sujeto, a los resultados de otros sujetos, y al número de medidas repetidas.

## 2.6 Modelo Sujeto específico

Como se mencionó anteriormente los modelos sujeto específico pueden ser tratados mediante tres enfoques: efectos fijos, efectos aleatorios y condicionales. En esta sección, se hace hincapié en el modelo de efectos aleatorios, como una forma de modelación de datos longitudinales. Cuando se mencione el modelo sujeto específico se referirá siempre a los modelos de efectos aleatorios. En contraste con los modelos marginales, donde la correlación entre medidas repetidas sobre el mismo sujeto se modela utilizando una matriz de correlación de trabajo, en los modelos sujeto específico, se induce la correlación entre las observaciones mediante la inclusión de efectos aleatorios de los sujetos. "La correlación entre las observaciones repetidas surge porque no podemos observar la curva de crecimiento subyacente, es decir, los coeficientes de regresión verdaderos" (Diggle et al. 2002). La importancia de la inclusión de efectos aleatorios en los modelos ha sido estudiada por gran cantidad de autores, que consideran que su omisión podría generar una pérdida de eficiencia y, por tanto, un aumento de los errores estándar de las estimaciones de parámetros (Caroll, 2003). Este modelo también es conocido por algunos autores como modelo condicional (Diggle et al., 2002; Lee y Nelder, 2004), modelo multinivel (Fitzmaurice y Lard, 1993) o, modelos de variable latente, dependiendo de la formulación del modelo. El Modelo sujeto específico consta de los siguientes supuestos:

- La esperanza condicional de la variable de respuesta está relacionada con las covariables y los efectos aleatorios por:

$$g(\tilde{\mu}_{ij}) = g\left[E(Y_{ij}|b_i)\right] = x'_{ij}\tilde{\beta} + z'_{ij}b_i \quad (2.22)$$

donde  $g(\cdot)$  es la función de enlace,  $\tilde{\mu}_{ij}$  es la media condicional,  $x'_{ij}$  es el vector de covariables,  $\tilde{\beta}$  es vector de parámetros de efectos fijos,  $z'_{ij}$  es el vector de covariables de efectos aleatorios, y  $b_i$  representa los efectos aleatorios.

- Condicionando sobre  $b_i$ , las respuestas  $y_{i1}, \dots, y_{in_i}$  son mutuamente independientes, siguen un MLG, y tienen una función de densidad en la

familia exponencial  $f(y|\beta, D, \sigma^2)$ .

- Los efectos aleatorios  $b_i$ , son independientes entre sí, con una distribución multivariada común subyacente  $f(b)$ . De manera similar a los modelos marginales, si  $g(\cdot)$  es la función identidad, entonces, el modelo de la ecuación (2.22) se reduce al modelo lineal mixto:  $E(Y_{ij} | b_i) = x'_{ij} \tilde{\beta} + z'_{ij} b_i$

En estos modelos,  $b_i$  es una muestra de variables independientes no observables desde una distribución de efectos aleatorios. Cualitativamente, este supuesto implica lo que se puede conocer acerca de los coeficientes de un individuo mediante la comprensión de la variabilidad de los coeficientes de toda la población. Cuando hay poca variabilidad, se debe confiar en los coeficientes de promedio poblacional para estimar los de un individuo. Cuando la variación es considerable, se debe confiar más en los datos de cada individuo a fin de estimar sus propios coeficientes. Los modelos de efectos aleatorios son más útiles cuando el objetivo es *hacer inferencia sobre los individuos más que sobre la media de la población*. Los coeficientes de regresión,  $\beta$  representan los efectos de las covarianzas sobre un sujeto en particular. Esta es la *interpretación sujeto específico*, a diferencia de los modelos marginales donde  $\beta$  tienen una interpretación de promedio poblacional. La estimación de máxima verosimilitud en la ecuación (2.22) se basa en la función de probabilidad de  $y$ :

$$f(y) = \prod_{i=1}^N \int \prod_{j=1}^{n_i} f(y_{ij} | \beta, D, \sigma) f(b) db \quad (2.23)$$

Se trata de una distribución marginal de  $y$  obtenida mediante la integración de la distribución conjunta de  $(y, b)$  respecto a  $b$ . En general, esta integral no tiene una forma cerrada, se han propuesto aproximaciones muy diferentes, por ejemplo, la cuadratura de Gauss-Hermite, la aproximación lineal de efectos mixtos (MLM), la aproximación de Laplace (Pinheiro y Bates), algoritmo de Monte Carlo (McCulloch, 1997) entre otras.

Existen algunos modelos sujeto específico muy conocidos como: beta – binomial, probit - normal. En particular los modelos lineales generalizados

mixtos (MLGM) serán el centro de interés en esta tesis, debido a su amplio uso en el análisis longitudinal de datos discretos.

## 2.7 Modelos Lineales Generalizados Mixtos (MLGM)

Los MLGM son una extensión del modelo lineal generalizado (MLG), adicionando los efectos aleatorios (Demidenko, 2004). En un MLGM se considera que las observaciones del vector de respuesta,  $y$ , son independientes condicionadas a los efectos aleatorios, cada uno con función de densidad en la familia exponencial (McCulloch y Searle, 2001).

Supongamos que  $\tilde{\mu}_{ij}$ , la media condicional de  $y_{ij}$ , está relacionada con la componente sistemática y los efectos aleatorios por la siguiente ecuación:

$$g(\tilde{\mu}_{ij}) = x'_{ij} \tilde{\beta} + z'_{ij} b_i \quad (2.24)$$

donde  $x'_{ij}$  y  $z'_{ij}$  son los vectores de covariables de los efectos fijos y aleatorios, respectivamente. La relación entre la variancia y la media está dada por:

$$Var(Y|b_i) = \phi a_i v(\tilde{\mu}_{ij}) \quad (2.25)$$

donde  $v(\cdot)$  es la función de variancias que relaciona las medias condicionales y las variancias,  $\phi$  es un factor de escala (es decir, se supone igual a uno para los modelos binomial estándar y Poisson),  $a_i$  es un peso a priori, tal como la inversa de un denominador binomial. Por último,  $b_i$  tiene una función  $f_{b_i}(\mu_i)$  de distribución conocida  $q$ -dimensional, tal como la distribución normal  $N(0, D)$ , que se utiliza en muchos casos.

Debido a las propiedades de la función  $g$ , es posible aplicar  $g^{-1}$  a ambos lados de la ecuación (2.24):

$$E(y_{ij}|b_i) = \tilde{\mu}_{ij} = g^{-1}(x'_{ij} \beta + z'_{ij} b_i)$$

Las consecuencias de la inclusión de efectos aleatorios en los modelos, pueden ser analizadas mediante el estudio de los dos primeros momentos de la distribución marginal de  $Y_{ij}$ . Si se tiene en cuenta que la media marginal de  $Y_{ij}$  inducida por los efectos aleatorios, que se denota por la  $E(Y_{ij})$ , viene dada por:



$$E_b(Y_{ij}) = E_b \left[ g^{-1} \left( x'_{ij} \tilde{\beta} + z'_{ij} b_i \right) \right] \quad (2.26)$$

En general esta expresión no se puede simplificar, debido a la no linealidad de la función  $g^{-1}$ . Un caso particular es el de regresión de Poisson con intercepto aleatorio distribuido normalmente con media cero y varianza  $2\mu$ . En este caso, la media marginal de  $Y_{ij}$  inducida es igual a:

$$E_b(Y_{ij}) = x'_{ij} \beta + \frac{\sigma_b^2}{2}$$

La media marginal obtenida (inducida) por la integración de las heterogeneidades individuales, será un promedio poblacional si y sólo si los sujetos en estudio pueden ser considerados como una muestra aleatoria de una población. Según algunos autores, el concepto de *promedio - poblacional* es válido sólo si estamos analizando una muestra aleatoria (Lee and J. Nelder, 2004).

La variancia marginal de  $Y_{ij}$  inducida por los efectos aleatorios, denotada por  $Var_b(Y_{ij})$ , tiene la siguiente expresión:

$$\begin{aligned} Var_b(Y_{ij}) &= Var(E[Y_{ij}|b_i]) + E(Var[Y_{ij}|b_i]) \\ &= Var(\mu_{ij}) + E[\phi a_i v(\mu_{ij})] \\ &= Var(g^{-1}[x'_{ij} \tilde{\beta} + z'_{ij} b_i]) + E\{\phi a_i v[g^{-1}(x'_{ij} \tilde{\beta} + z'_{ij} b_i)]\} \end{aligned} \quad (2.27)$$

Nuevamente esta expresión en la mayoría de los casos tampoco puede simplificarse. En un modelo lineal mixto, la media marginal inducida en la ecuación 2.26 se reduce a:  $E_b(Y_{ij}) = x'_{ij} \tilde{\beta}$ . Asimismo, la variancia marginal inducida en la ecuación 2.27 se puede reducir a:  $Var_b(Y_{ij}) = Z_i D Z'_i + \Sigma_i$ .

Finalmente, asumiendo independencia condicional de los elementos de  $Y_i$  obtenemos lo siguiente:

$$\begin{aligned} Cov_b(Y_{ij}, Y_{ik}) &= Cov(E[Y_{ij}|b_i], E[Y_{ik}|b_i]) + E[Cov(Y_{ij}, Y_{ik}|b_i)] \\ &= Cov[g^{-1}(x'_{ij} \tilde{\beta} + z'_{ij} b_i), g^{-1}(x'_{ik} \tilde{\beta} + z'_{ik} b_i)] \end{aligned} \quad (2.28)$$

Del mismo modo, en un modelo lineal mixto, la ecuación anterior se reduce a la siguiente expresión:

$$Cov_u(Y_{ij}, Y_{ik}) = D_i z_{ij} z'_{ik} \quad (2.29)$$

Si el modelo lineal mixto tiene un solo intercepto aleatorio se tiene que  $Cov_u(Y_{ij}, Y_{ik}) = \sigma_b^2$  donde  $\sigma_b^2$  es la variancia de los interceptos aleatorios.

**Inferencia en MLGM:**

Para ajustar un MLGM es posible utilizar dos alternativas: el enfoque Bayesiano o la Estimación Máximo Verosímil (EMV). En el enfoque bayesiano, es necesario especificar la función de densidad previa para  $\beta$ ,  $D$ , y  $\phi$ , que se denota por  $f(\beta)$ ,  $f(D)$ , y  $f(\phi)$ , respectivamente. Una vez definidas las funciones a priori, se puede encontrar la distribución a posteriori (Molenberghs y Verbeke, 2005).

Por otra parte, la función completa de verosimilitud para MLGM puede escribirse a partir del producto de las distribuciones conocidas  $y|b$  y  $b$ . La función de densidad conjunta puede escribirse como:

$$L(y, b) = L(y|b)L(b) \tag{2.30}$$

Si asumimos que  $b$  tiene una distribución normal multivariada  $N(0, D)$ , entonces:

$$L(y, b) \propto L(y|b)|D|^{-1/2} \exp\left(-\frac{1}{2}b'D^{-1}b\right) \tag{2.31}$$

Basándose en la ecuación 2.31, la probabilidad marginal para  $y$  se podría obtener integrando sobre los efectos aleatorios  $b$ . Sin embargo, esto no siempre es posible porque la integral resultante de la parte derecha de la ecuación 2.31 no tiene una solución analítica, es decir, debe ser resuelto mediante métodos numéricos.

### 2.7.1 Modelo Logit Mixto para variable de respuesta nominal

Primero se define el modelo cuando las categorías de  $Y_{ij}$  no tienen orden, en este caso cada logit junta o aparea cada categoría de respuesta con una categoría de respuesta arbitraria o de referencia, por ejemplo la categoría  $r$ , fijando un grupo específico de efectos aleatorios  $b_{ir}$  para cada logit. Luego, se define el Modelo logit de efectos aleatorios con categoría de referencia de la siguiente manera:

$$\text{logit} \left( \frac{\pi_{ijr}}{\pi_{ijR}} \right) = \alpha_r + x'_{ij} \beta_r + z'_{ij} b_{ir} \quad r = 1, \dots, R-1 \quad (2.32)$$

Los parámetros  $\alpha_r$ ,  $\beta_r$  y los efectos aleatorios dependen de  $r$ , ya que es la categoría de referencia; en la respuesta nominal no hay ninguna razón para esperar que los efectos sean similares para los diferentes  $r$ . La definición del modelo se completa mediante la especificación de una distribución para  $b'_i = (b'_{i1}, b'_{i2}, \dots, b'_{iR-1})$ . Generalmente se definen los efectos aleatorios  $\{b_i\}$  como variables normales multivariadas independientes, con una matriz de covarianza arbitraria  $D$ .

En particular, es razonable definir diferentes variancias para los efectos aleatorios que se aplican a diferentes logits. Si por el contrario se definiera una variancia común, esta variancia no podría ser la misma que la implícita en un efecto aleatorio para un logit de categoría de referencia,  $\text{logit} \left( \frac{\pi_{ijj}}{\pi_{ijj'}} \right)$ .

Con una estructura de covarianza arbitraria el modelo es estructuralmente el mismo, independientemente de la elección de la categoría (Hartzel et al., 2001).

Por el contrario la estructura general (2.32) supone que los  $\{b_i\}$  son independientes para diferentes  $r$  o perfectamente correlacionados. Dado que los efectos aleatorios de diferentes logit derivan del mismo sujeto, la hipótesis de que son independientes puede ser poco realista. Sin embargo, asumir que son perfectamente correlacionados es demasiado restrictivo (Hartzel et al., 2001).

### 2.7.2 Modelo Logit Acumulado Mixto para respuesta Ordinal

Existen diversos modelos para variables de respuesta ordinales, por ejemplo el modelo logit de categoría adyacente (Agresti, 2002) y los modelos secuenciales (Tuz et al., 1996), los cuales podrían proporcionar un modelo básico.

Una variedad de modelos de regresión ordinal usan el enlace logit. El más popular es el modelo logit acumulado. Por lo tanto, se decide aplicar este modelo, debido a que proporciona interpretaciones más sencillas que los modelos anteriores sin violar el orden en la variable de respuesta.

Sea  $y_{it}$  que denota la respuesta de la observación  $t$  en el grupo  $i$ . Sea  $x_{1it}, x_{2it}, \dots, x_{kit}$  denota los valores de las  $k$  variables explicativas en todas las observaciones. El efecto aleatorio para el grupo  $i$  se identifica como  $b_i$ . El modelo más simple y más utilizado es el que se considera el intercepto aleatorio. Para las categorías  $j=1,2,\dots, c-1$ , el modelo logit acumulado odds proporcional con intercepto aleatorio tiene la siguiente forma:

$$\text{logit} \left( P(Y_{it} \leq j) \right) = b_i + \alpha_j + \beta_1 x_{1it} + \beta_2 x_{2it} + \dots + \beta_k x_{kit} \quad (2.33)$$

Este modelo considera el mismo efecto aleatorio para cada probabilidad acumulada. Usar un intercepto sobre todos los términos de la forma  $b_i + \alpha_j$  es también una vía para modelar subjetivamente los cambios sujeto – específicos en las categorías de la variable de respuesta ordinal, permitiendo que los puntos de corte de la variable de respuesta varíen a través de los sujetos, de acuerdo a la siguiente expresión:

$$\log \left( \frac{\pi_{ij1} + \dots + \pi_{ijr}}{\pi_{ij,r+1} + \dots + \pi_{ijR}} \right) = \alpha_r + x'_{ij} \beta + Z_{ij} b_i \quad r = 1, \dots, R - 1$$

En contraste en el modelo (2.32), el efecto fijo  $\beta$  es el mismo para todos los logit, llamado supuesto de probabilidades proporcionales (odds proporcional). Modelos más generales permiten que el efecto  $\beta_r$  varíe según los distintos logit para, al menos, algunos predictores. En el modelo (2.33), cada logit para cada  $i$ -ésimo grupo también tiene el mismo efecto aleatorio  $b_i$ . El parámetro  $\alpha$ ,

llamado intercepto o umbral, satisfice  $\alpha_1 < \alpha_2 < \dots < \alpha_{R-1}$ .

En un modelo lineal generalizado mixto el cálculo de la integral de la función de las estimaciones de máxima verosimilitud de los parámetros no es cerrada, se requieren técnicas de integración numérica:

$$\int f(y_1, \dots, y_n | b_1, \dots, b_n; \beta) = \int \prod_{i=1}^n \prod_{t=1}^{T_i} f(y_{it} | b_i; \beta)$$

$b_i$  es una variable aleatoria específica de grupo que se distribuye normalmente,  $b_i \sim N(0, \sigma^2)$ .

La más utilizada es la técnica de Laplace y en algunos paquetes estadísticos se ha implementando la técnica de Montecarlo y la de Gauss-Hermite (Tutz y Hennevogl, 1996), donde ésta última técnica suele ser más precisa pero a la vez computacionalmente más exigente.

El supuesto de que los efectos aleatorios se distribuyen normal es necesario en el momento de la estimación con el enfoque de la Cuadratura de Gauss Hermite, pero para la integración a través de Monte Carlo puede dejarse de lado este supuesto (Hartzel J., Agresti A., Caffo B., 2001).

Este modelo es de gran utilidad en todas las ciencias debido a que en la práctica, las muestras a menudo están correlacionadas o vienen en grupos. En particular, si una variable de respuesta se mide repetidamente o las unidades primarias de muestreo son luego submuestreadas, no se puede esperar que las respuestas sean independientes. Hay varias maneras de abordar este problema: existen métodos promedios en la población y métodos de grupo o sujeto específicos. En los métodos de promedio poblacional los coeficientes se interpretan con referencia a la población y no para cualquier individuo, lo que sí se logra con los modelos sujeto específico. Es por eso que son tan ricos tanto para las estimaciones como para las interpretaciones.

En este trabajo de investigación se utilizarán los modelos Log – Acumulado mixtos, por lo tanto para lograr un mejor entendimiento se presentan a continuación dos ejemplos.

**Ejemplo 1.** Evaluación acerca de la astringencia de vino (Tutz y Hennevolg, 1996). La Tabla 2.2 muestra los datos proporcionados por Randall (1989), los mismos son resultado de un experimento factorial donde hay nueve jueces que evalúan el sabor de ocho botellas de vino. Las botellas de vino se caracterizan por los factores contacto y temperatura. El efecto Contacto (C: sí o no) determina si hubo contacto del mosto con la piel de las uvas durante la fermentación. El efecto temperatura consta de dos categorías (T: frío y caliente). En la degustación de los vinos los jueces dieron una evaluación acerca de la astringencia de vino en una escala de cinco puntos que va de menos a más amargo. Dado que cada juez cata 8 botellas de vino, las observaciones no pueden ser consideradas como independientes, ya que cada juez tiene una sensibilidad diferente, resultando una correlación causada por la heterogeneidad de los sujetos.

Luego, la amargura de vinos es juzgada con una escala de cinco puntos que va de menos a más astringente. Las covariables son el contacto (C, sí / no), temperatura (T, frío / caliente) y la botella (B, 1/ 2). Luego hay nueve jueces donde evalúan cada vino a partir de dos botellas, de cada una de las cuatro condiciones de tratamiento, por lo tanto, hay 72 observaciones en total.

**Tabla 2.2:** Datos de Astringencia de vinos.

Juez	Temperatura: Frío				Temperatura: Caliente			
	Contacto: No		Contacto: Si		Contacto: No		Contacto: Si	
	Botella 1	Botella 2	Botella 1	Botella 2	Botella 1	Botella 2	Botella 1	Botella 2
1	2	3	3	4	4	4	5	5
2	1	2	1	3	2	3	5	4
3	2	3	3	2	5	5	4	4
4	3	2	3	2	3	2	5	3
5	2	3	4	3	3	3	3	3
6	3	2	3	2	2	4	5	4
7	1	1	2	2	2	3	2	3
8	2	2	2	3	3	3	3	4
9	1	2	3	2	3	2	4	4

En este ejemplo se espera heterogeneidad entre los individuos, debido a que para un juez el concepto de más amargo (categoría 5) no será el mismo que

para el resto de los jueces. Lo que es más amargo para el juez 1 puede ser de un nivel intermedio para el juez 2. Es por ello que se debe tener en cuenta la distinta sensibilidad de los jueces y esto se puede realizar planteando un modelo sujeto específico, es decir que tenga en cuenta el efecto del juez. La modelización explícita de la heterogeneidad en la forma de efectos aleatorios es una herramienta común en los modelos lineales. Es menos común para los datos categóricos, aunque en la actualidad se han realizado importantes trabajos al respecto. En particular, se han desarrollado los modelos de efectos aleatorios para los datos binarios o de recuento (Agresti, A. 2010).

Se plantea el siguiente modelo:

$$\text{logit}(P(Y_i \leq j)) = \theta_j - \beta_1(\text{temp}_i) - \beta_2(\text{contact}_i) - u(\text{juez}_i) \quad (2.34)$$

$$i = 1, \dots, n, \quad j = 1, \dots, J - 1$$

el cual describe la probabilidad acumulada de la *i*-ésima calificación que cae en la *j*-ésima categoría o por debajo, donde el índice *i* representa todas las observaciones y el índice de las categorías de respuesta que va desde *j* = 1, ..., *J* (*J*=5). Los  $\{\theta_j\}$  son parámetros llamados umbrales o puntos de corte. Se toma el efecto juez como aleatorio y se asume que el efecto juez es independiente e idénticamente distribuido (iid) normal:  $u(\text{juez}_i) \sim N(0, \sigma_u^2)$ .

Para el ajuste de este modelo se utiliza el paquete “ordinal”, la función “clmm2” (Christensen, R. 2013) del software R. Se realizan las primeras estimaciones a través de la Aproximación de Laplace, obteniendo los siguientes resultados.

**Tabla 2.3:** Estimación de los parámetros del modelo 2.34 mediante la Aproximación de Laplace

Efectos aleatorios	Var	Desvío Estándar		
Juez	1.279455	1.13113		
<b>Efectos fijos</b>				
Coeficientes	Temperatura (Caliente)	Contacto (si)		
	3.062993	1.834883		
Umbrales	1 2	2 3	3 4	4 5
	-1.623664	1.51336	4.228525	0.088770
Log-likelihood: -81.56541	AIC: 177.1308			

Las estimaciones de máximos verosímiles de los parámetros se proporcionan

usando la aproximación de Laplace. Una aproximación más exacta es proporcionada por el método de cuadratura de Gauss-Hermite. Se utilizaron 10 nodos de cuadratura la salida se muestra en la tabla 2.4.

**Tabla 2.4:** Estimación de los parámetros del modelo mediante el método de cuadratura de Gauss-Hermite (10 nodos)

Efectos aleatorios	Var	Desvío Estándar		
Juez	1.287741	1.134787		
Efectos fijos				
Coeficientes	Temperatura (caliente)	Contacto (si)		
	3.0619	1.8334		
Umbrales	1 2	2 3	3 4	4 5
	-1.6235	1.5128	4.2271	6.0862
Log-likelihood: -81.53246		AIC: 177.0649		

Los cambios en las estimaciones de los parámetros son pequeños, lo que muestra que la aproximación de Laplace era de hecho bastante precisa en este ejemplo. Los resultados muestran las estimaciones de máxima verosimilitud de los parámetros:

$$\hat{\beta}_1 = 3,06, \hat{\beta}_2 = 1,83, \hat{\sigma}_u^2 = 1,29 = 1,13^2, \{\hat{\theta}_j\} = [-1,62; 1,51; 4,23; 6,09]$$

En la cuadratura de Gauss Hermite el valor del Hessiano se toma como una medida de bondad de ajuste. Valores altos como  $10^4$  o  $10^6$  indican que el modelo está mal definido, incompleto o saturado, lo que significaría que el modelo se puede simplificar, o que posiblemente algunos parámetros no son identificables y luego la optimización del modelo puede ser complicada. En este ejemplo el valor del Hessiano no indica que exista algún problema con el modelo (Christensen, 2013).

Los coeficientes Temperatura y Contacto son positivos lo que indica que a mayor temperatura y mayor contacto aumenta la amargura de vino, es decir, que es más probable que el vino obtenga una clasificación en categorías superiores.

La razón de probabilidad (odds ratio) del evento  $Y \geq j$  es  $\exp(\beta_{tratamiento})$ , luego el odds ratio de la amargura es equivalente a la chance de que sea clasificado en la categoría  $j$  o una superior a caliente respecto de fría es:  $e^{3,0619} = 21,3681$ ; que tiene la siguiente interpretación: un vino que se encuentra a mayor



temperatura tiene 21 veces más de chances de estar en la categoría  $j$ -amarga (o menor) que un vino que está a temperatura fría.

Los  $p$ -valores para los coeficientes se basan en el estadístico de Wald. Existen pruebas más precisas como las pruebas de razón de verosimilitud. Por ejemplo, la prueba de razón de verosimilitud para el efecto contacto se realiza comparando el modelo que contiene el efecto contacto y el que no lo contiene:

**Tabla 2.5:** Prueba de Razón de Verosimilitud – (Rating = Contacto + Temperatura vs Rating = Temperatura)

	Grados de libertad	-2logLik	Test	LR stat.	Pr(Chi)
Temp	66	177.4090			
Temp + Contact	65	163.0649	1 vs 2	14.34409	0.0001522572

El  $p$ -valor del efecto contacto es ligeramente menor que en la prueba anterior.

Para evaluar si el efecto aleatorio juez es significativo, se plantea un test de razón de verosimilitud, donde se comparan los dos modelos completos (Rating = Contacto + Temperatura), pero considerando primero el efecto juez como un efecto aleatorio versus el modelo completo pero sin el efecto juez. Donde se obtienen los siguientes resultados:

**Tabla 2.6:** Prueba de Razón de Verosimilitud – (Rating = Contacto + Temperatura para evaluar el efecto Juez).

	Grados de libertad	-2logLik	Test	Grados de libertad	LR stat.	Pr(Chi)
Temp + Contacto	66	172.9838				
Temp + Contacto (con Efecto juez)	65	163.0649	1 vs 2	1	9.918925	0.001635879

Lo que representa que el término juez es significativo.

Dado que esta prueba de  $\sigma_u$  está en el límite del espacio de parámetros (una variancia no puede ser negativa), se suele decir que un  $p$ -valor más correcto se obtiene al dividir por dos el  $p$ -valor producido por la prueba de razón verosimilitud convencional.

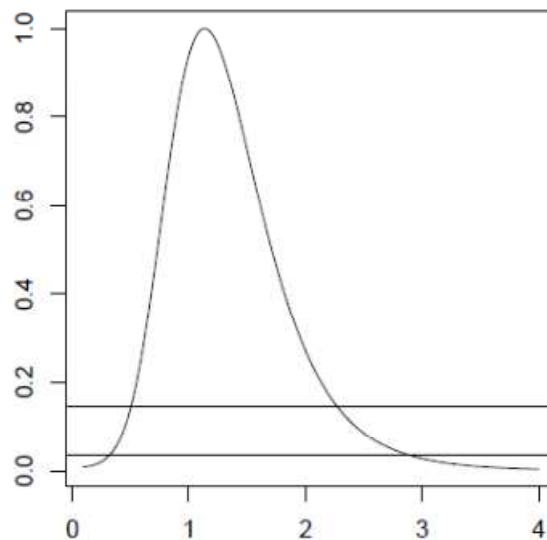
En este caso, reducir a la mitad el  $p$ -valor es de poca relevancia.

Se calcula el intervalo de confianza para el perfil de verosimilitud de  $\sigma_u$ :

**Tabla 2.7:** Intervalo de confianza para  $\sigma_u$ .

	2,5 %	97,5 %
0.5048597	0.5048597	2.272327

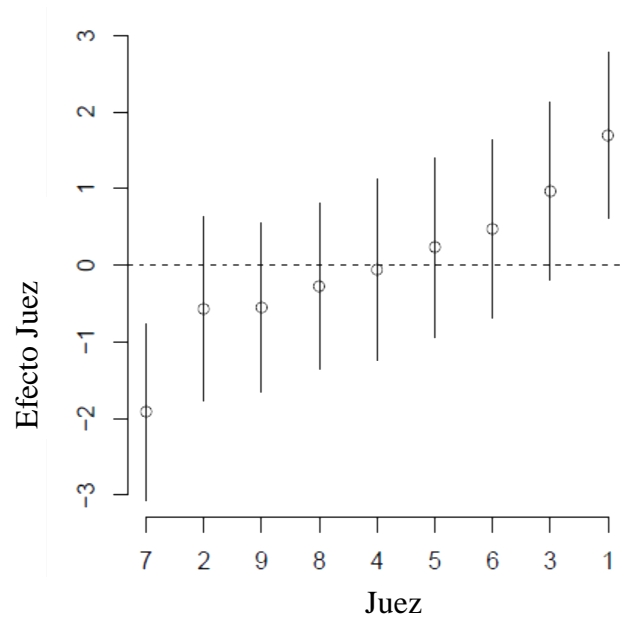
**Figura 2.6.** Perfil de verosimilitud de  $\sigma_u$



La figura 2.4 representa el perfil de verosimilitud, en el cual las líneas horizontales indican el 95% y el 99% límites de confianza.

El efecto juez no es un parámetro, por lo que no puede calcularse en el sentido convencional, sino que una "mejor predicción" es proporcionada por los modos condicionales. Del mismo modo la varianza condicional proporciona una medida de incertidumbre de los modos condicionales. La figura 2.5 ilustra el efecto juez a través de los modos condicionales con intervalos de confianza del 95% basados en la variancia condicionada.

**Figura 2.7.** Intervalos del 95% de confianza basados en la varianza condicional para el efecto Juez.



El séptimo juez dio las calificaciones más bajas de amargura, mientras que el primer juez dio las calificaciones más altas de la amargura. Que el efecto juez sea significativo indica que los jueces percibieron la amargura de los vinos diferente. Como se observa los jueces tienen distintas interpretaciones de amargura o en realidad los jueces perciben la amargura de los vinos diferente. Posiblemente ambas afectan a la calificación.

**Ejemplo 2.** Relevadores de tamaño de frutos peras William's. (Giménez G., et al. 2013).

Los datos analizados surgen del relevamiento de información, procesamiento y elaboración del pronóstico de producción de las principales variedades de fruta de pepita existentes en el Alto Valle. Las observaciones corresponden al recuento de frutos por planta y registro de diámetros en la variedad peras William's. Las mismas fueron realizadas por 20 relevadores los cuales registraron de forma subjetiva y objetiva las mediciones de los frutos. Se entiende que las mediciones subjetivas se realizan clasificando a los frutos en pequeños, medianamente pequeños, medianos, medianamente grandes y grandes, mientras que las medidas objetivas se realizan tomando los registros del diámetro ecuatorial de los frutos en milímetros con un calibre digital. Debido a que la variable de respuesta (Tamaño de Fruto) consta de cinco categorías y es de naturaleza ordinal, se plantea un Modelo Logit Acumulado. En este trabajo se analizan las apreciaciones subjetivas de los frutos por lo tanto es necesario considerar el efecto del relevador como un efecto aleatorio.

- ❑ El total de observaciones obtenidas 835.
- ❑ Temporadas relevadas: 2011 - 2012
- ❑ Variable de respuesta subjetiva Tamaño de Fruto. Codificación: 1: pequeño, 2: medianamente pequeño, 3: mediano, 4: medianamente grande, 5: grande.
- ❑ Efecto Aleatorio: Relevador. Total de relevadores: 20.
- ❑ Efectos fijos: Edad de la plantación cuyos estratos son 10 a 19; 20 a 29; 30 a 39 y mayor de 40 años, Conducción de las plantas (L: libre, E: en espaldera) , Año (2012, 2013)
- ❑ Software libre R, utilizando la librería ordinal y la librería ggplot 2.
- ❑ Se analizó desde un modelo saturado con efectos fijos y todas sus interacciones hasta un modelo parsimonioso.
- ❑ Para comparar los modelos y verificar el ajuste se utilizó el valor del AIC (criterio de Akaike), el valor del Hessiano y la log-verosimilitud.

Se comparó de forma secuencial desde el modelo saturado log-acumulado para tamaño de fruto teniendo en cuenta: edad de la plantación, conducción de las plantas, año y las interacciones triples y dobles.

Se concluyó que el modelo apropiado contiene solamente los efectos principales, de los cuales el efecto Conducción no fue significativo. Los resultados se muestran en la siguiente tabla 2.8.

**Tabla 2.8.** Test de verosimilitud de los modelos analizados

Test 1. Modelo con interacciones triples versus modelo con interacciones dobles				
AIC	logLik	LR.stat	df	P(>Chi)
1375.0	-670.48			
1380.8	-670.40	0.1743	3	0.9816
Test 2. Modelo con interacciones dobles versus modelo con efectos principales				
AIC	logLik	R.stat	df	P(>Chi)
1371.3	-675.63			
1375.0	-670.48	10.289	7	0.1728
Test 3. Modelo con todos los efectos principales versus modelo sin el efecto conducción				
AIC	logLik	LR.stat	df	P(>Chi)
1369.4	-675.71			
1371.3	-675.63	0.1736	1	0.6769

Se estimó la integral mediante el algoritmo de Laplace. Para el efecto año se obtuvo un odds ratio de 1,85 representando casi el doble de frutos más grandes en el año 2012 que en el año 2013.

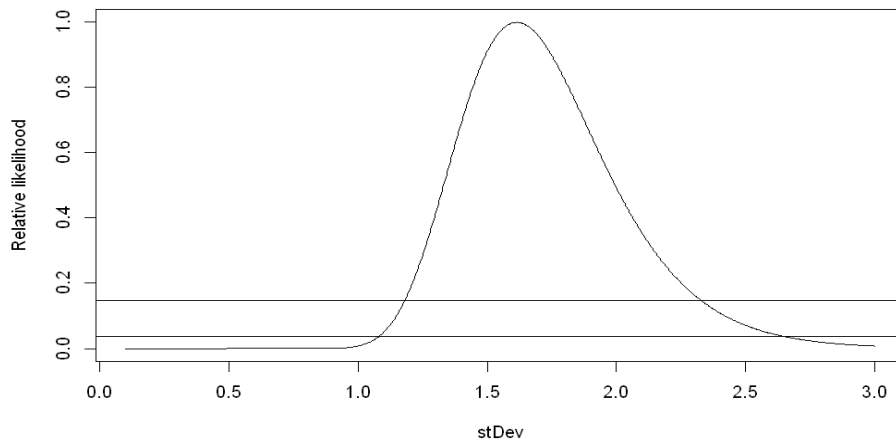
Las edades 20-29 y 30-39, ambas respecto de la categoría más de 40, tienen casi el doble de chance (1,73 y 1,82 respectivamente) de tener frutos más grandes.

**Tabla 2.9:** Estimación de los parámetros del modelo. TAMAÑO ~ ANIO + EDAD + (1|Enc)

Efectos fijos				
Coeficientes	Año 2013	Edad 10-19	Edad 20-29	Edad 30-39
	0.5165	-0.4001	-0.5469	-0.5799

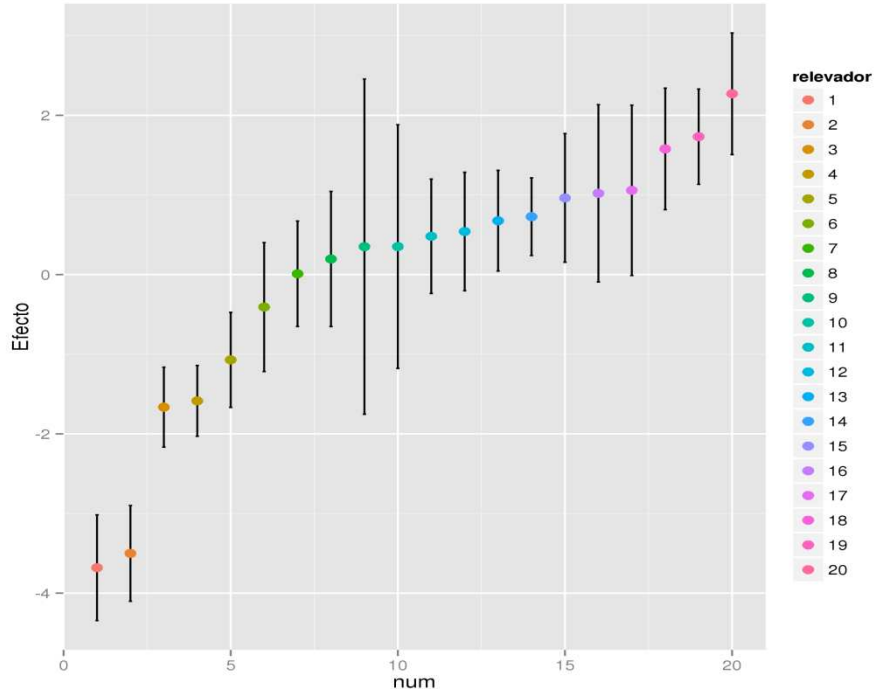
Los valores obtenidos del efecto aleatorio en términos del desvío estándar con un 95% de confianza fueron de 1,18 y 2,32. Debido a que estos valores son distintos de 1 confirman la importancia del efecto aleatorio en la modelación de los datos.

**Figura 2.8.** Perfil relativo de verosimilitud.



Se hallaron intervalos de confianza para el efecto aleatorio de cada uno de los relevadores. Se observa gráficamente que existe efecto Relevador.

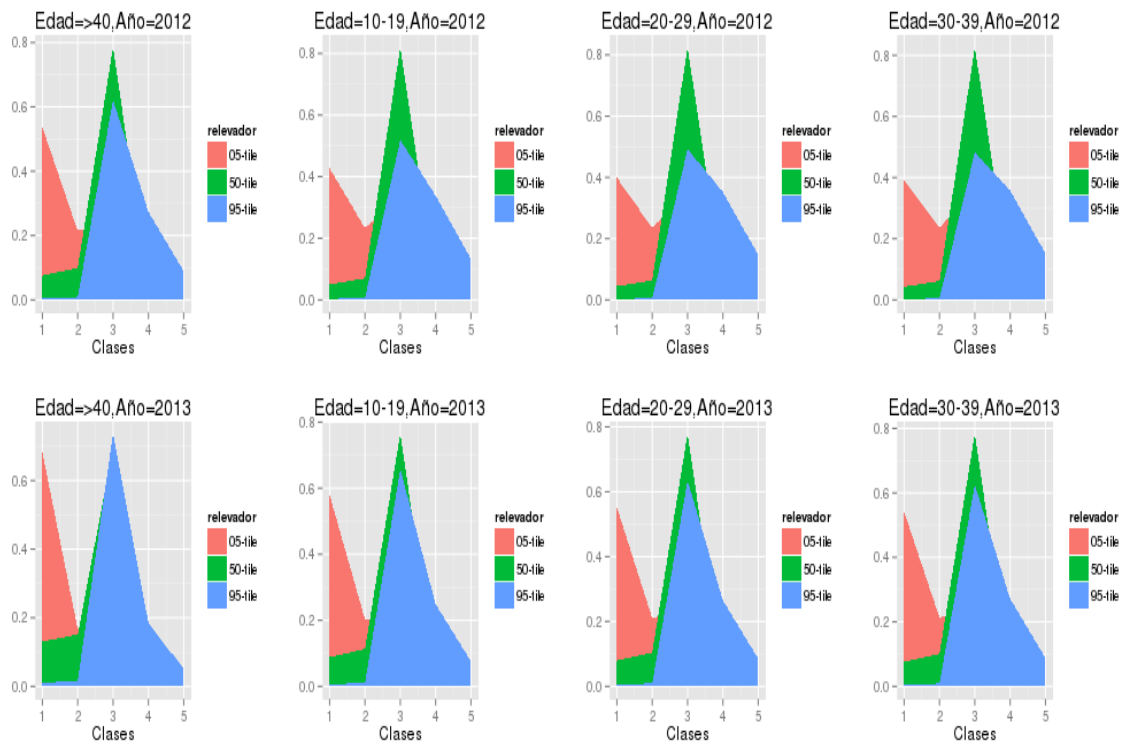
**Figura 2.9.** Intervalos de confianza para cada uno de los Relevadores



Respecto a la edad de la plantación, al comparar las temporadas considerando al relevador medio se observa que la probabilidad de encontrar frutos pequeños y medianamente pequeños (1 y 2) es mayor en la temporada 2013.

Esto se cumple para todos los estratos de edades pero se refleja con mayor claridad para las plantas de mayor edad. Dentro de cada temporada el relevador medio es más probable que observe frutos medianos en todas las edades de las plantaciones.

**Figura 2.10.** Probabilidades para el relevador promedio y extremos en distintas condiciones.



Los valores obtenidos en el ajuste del modelo final fueron los siguientes: valor del AIC 1371, el valor del Hessiano 150, recordando que dicho valor debe ser menor a  $10^4$  para asegurar su ajuste adecuado y la convergencia del modelo, y por último la Log-verosimilitud fue de - 676. Debido a que los pronósticos siempre se relativizan a la temporada anterior, este modelo es muy consistente, permitiendo utilizar las apreciaciones subjetivas respecto del tamaño de los frutos, con ambas mediciones se arriba a similares conclusiones, contemplando la variabilidad existente entre los relevadores.

## **CAPITULO III**

### **Aplicación**

Para el estudio de las distintas aristas de los Moldeos Lineales Generalizados Mixtos se utilizará el conjunto de datos relevados en la investigación llevada a cabo por el Dr. Brian Finegan en la Cátedra Latinoamericana de Ecología en el Manejo de Bosques Tropicales, Centro Agronómico Tropical de Investigación y Enseñanza, CATIE, Costa Rica.

El estudio se llevo a cabo en dos bosques primarios de Costa Rica: La Tirimbina y Corinto. En La Tirimbina se encuentran 6 parcelas a las cuales se les aplicaron distintos tratamientos silviculturales, registrando dos de ellos, el aprovechamiento forestal (A) y el aprovechamiento forestal y tratamiento Silvicultural (S). En el área de estudio del Bosque Corinto hay establecidas 9 parcelas permanentes donde se aplicaron tres tratamientos, los mismos que en el bosque La Tirimbina más un control (N). Se analizó la variable de respuesta Clase Diamétrica, la cual representa la medición del diámetro de distintas plantas que fueron clasificadas en tres Grupos funcionales (GF1, GF2 y GF3). Se identificaron estos tres grupos funcionales de plantas basados en un conjunto de variables (tipo de dispersión, agente polinizador y sistema sexual, estrato, tasa de crecimiento, etc.) usando técnicas multivariadas de agrupamiento. Es decir, un grupo funcional se definió como un conjunto de especies vegetales que son similares en cuanto a su “roll” en el ecosistema.

El tiempo de medición varió de acuerdo al sitio y la parcela, hay años que no se midieron todas las parcelas, el periodo en estudio fue desde el año 1990-2003.

En el siguiente apartado 3.1 se realiza un análisis multivariado como paso previo a la modelación con el fin de identificar cuáles son los efectos que intervienen en la explicación de la variable de respuesta Clase Diamétrica.

En el apartado 3.2 se aplica un modelo log – acumulado que tiene en cuenta la



ordinalidad de la variable de respuesta y se analiza tanto el cumplimiento del supuesto de proporcionalidad de los efectos que intervienen en el modelo como así también cuales son los efectos que son significativos en la explicación de las observaciones.

En el apartado 3.3 se amplía el análisis teniendo en cuenta al individuo (parcela) mediante el uso de un Modelo Generalizado Mixto (MLGM) que tiene en cuenta la correlación existente entre las observaciones que provienen de la misma parcela. En este apartado se presenta el modelo definitivo con las respectivas estimaciones de los coeficientes incluyendo el análisis de proporcionalidad mediante las parcelas promedio.

### **3.1 Análisis descriptivo Multivariado**

Sobre cada unidad experimental se registran los valores de 12 variables, el análisis descriptivo multivariado puede facilitar la comprensión de las interrelaciones entre dichas variables. Los resultados de estos análisis pueden utilizarse como paso previo a la modelación principalmente los referidos a la variable de respuesta clase diamétrica.

En esta etapa principalmente se realizó un análisis exploratorio a través del cual se identificaron grupos de variables que más contribuyen a la explicación del comportamiento de los individuos y además cuáles son los efectos que tendría sentido incorporar posteriormente en la etapa de modelación.

Para comenzar con el desarrollo del análisis se unificó el nombre de los niveles del efecto tratamiento de ambos bosques de la siguiente manera:

A: parcelas con aprovechamiento forestal, las cuales en el bosque Tirimbina estaba identificado por la letra T.

S: parcelas con aprovechamiento forestal y tratamiento Silvicultural, las cuales en el bosque Tirimbina estaba identificado por la letra L.

N: parcelas sin aprovechamiento forestal, las mismas fueron muestreadas solo en el Bosque Corinto y estaban identificadas por la letra T.

Antes de comenzar con el análisis de componentes principales se transformó los recuentos originales en porcentajes, por ejemplo número de árboles en las clases diamétricas a porcentaje de plantas en clase diamétrica 1.

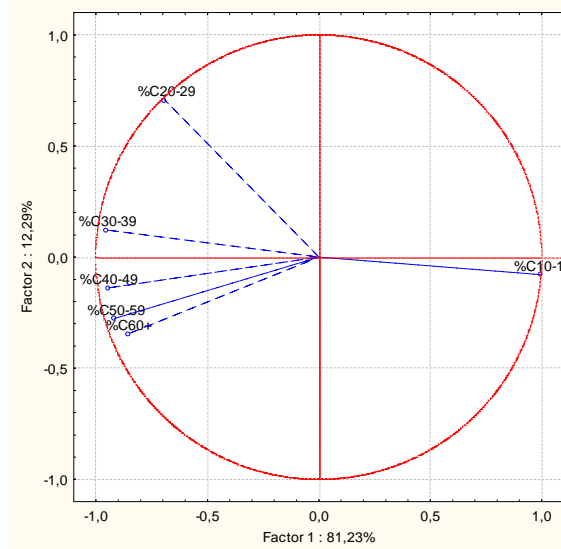
Los identificadores de los individuos, es decir cada parcela, son de la siguiente forma: Bosque – Año – Tratamiento – Grupo Funcional. (Por ejemplo: T90TA2: Tirimbina, año 1990, Tratamiento A, Grupo Funcional 2).

Al realizar un análisis de componentes principales en cada bosque se analizaron los autovectores obteniendo la siguiente información: la primera componente explica un 81,23% de la variabilidad total y la segunda componente explica el 12,29% de la variabilidad total.

A partir de la figura 3.1 se puede identificar un posible agrupamiento de categorías. Pudiendo definir una única variable de respuesta “Clase Diamétrica” (CD), separando la misma en tres categorías, la primera: la clase diamétrica de plantas más pequeñas, Clase diamétrica 1 (CD1): plantas con una clase diamétricas entre 10 -19 cm; plantas medianas Clase diamétrica 2 (CD2): plantas con una clase diamétrica entre 20 -29 cm y por último la clase

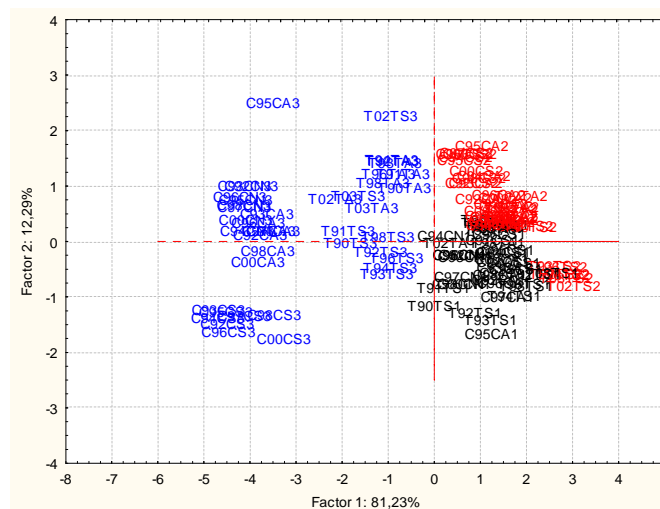
diamétrica que contiene las plantas más grandes, Clase diamétrica 3 (CD3): plantas con una clase diamétrica entre 30 -39 cm, 40 -49 cm, 50 -59 cm y 60 o más.

**Figura 3.1.** Proyección de las variables en plano.



A continuación, las siguientes figuras muestran la proyección de los individuos en el plano principal destacando los distintos factores que puedan intervenir en la posterior modelación.

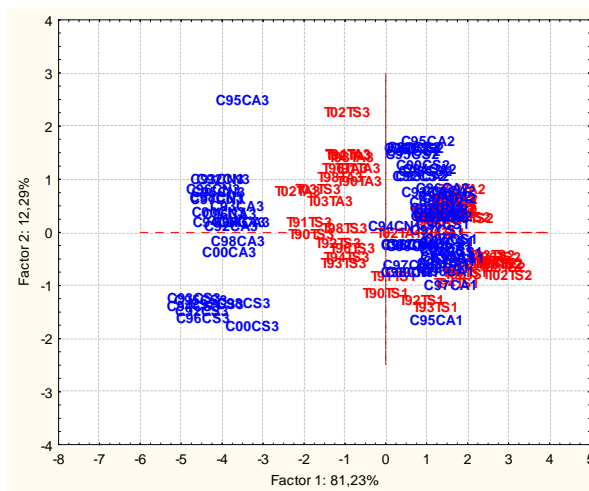
**Figura 3.2.** Proyección de los individuos en plano identificados por los Grupos Funcionales



Los conjuntos representados en la figura 3.2 corresponden a los tres Grupos

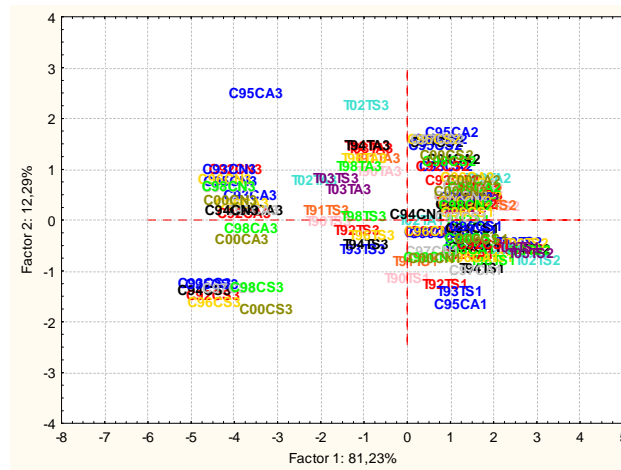
Funcionales de plantas. Se observa en la disposición de la agrupación por grupo funcional de plantas que la configuración promedio de los individuos proyectados en el primer factor (81,23%) del Grupo funcional 3 difiere de la del Grupo funcional 2 y 3. Mientras que las configuraciones medias de los grupos funcionales 1 y 2 son similares, por lo que se deberá incluir en los modelos posteriores el efecto Grupo Funcional.

**Figura 3.3.** Proyección de los individuos en plano identificados por el Bosque al que pertenecen



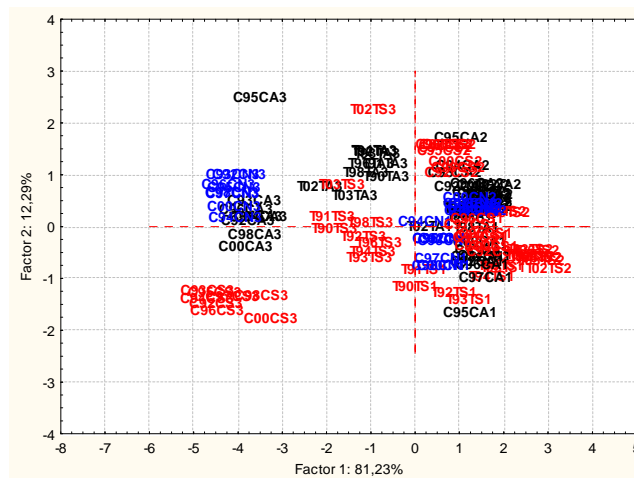
Si se analiza en la figura 3.3 sólo el efecto Bosque comparando las configuraciones promedio de los individuos nuevamente en el factor 1 que es el que más explica, se observa que ambos bosques (Tirimbina y Corinto) difieren, por lo que sería razonable incluir el efecto Bosque en una modelación posterior. Si además se observa la proyección de los individuos medios en el primer factor pero dentro de cada Grupo Funcional nuevamente difieren, por lo que es un indicio una posible interacción entre el bosque y los grupos funcionales.

**Figura 3.4.** Proyección de los individuos en plano identificados por Año



La figura 3.4 muestra cada año con un diferente color, lo cual refleja que en su mayoría la configuración promedio difiere para los distintos años, por lo que el efecto año debería estudiarlo en una modelación posterior.

**Figura 3.5.** Proyección de los individuos en plano identificados por Tratamiento



La figura 3.5 muestra cada tratamiento con un diferente color, lo cual refleja que las proyecciones de las configuraciones medias son similares para los distintos tratamientos, lo que sería un indicio de que no existe efecto tratamiento.

De esta etapa de análisis multivariado se concluye:

- Se define la variable de respuesta Clase Diamétrica con tres categorías CD1, CD2 y CD3, donde la categoría CD1 agrupa las plantas pequeñas,

la categoría CD2 las plantas medianas y la categoría CD3 plantas grandes. Considerando las distintas agrupaciones de la siguiente manera: **CD1**: 10-19, **CD2**: 20-29, **CD3**: 30-39, 40-49, 50-59 y > 60. Se realiza esta agrupación para simplificar las interpretaciones ya que los efectos de los parámetros son invariantes a la elección de las categorías de la variable latente diámetro. Dicha reagrupación se realizó en ambos bosques.

- Debido a que en los dos Bosques no se realizaron los mismos tratamientos todas las modelizaciones y análisis se harán por separado en cada uno de los bosques.
- Se consideran los efectos Tratamiento, Año, Grupo Funcional y sus respectivas interacciones.

### 3.2 Modelo Logit para variables ordinales

Debido a la ordinalidad de la variable de respuesta y como paso previo a la introducción de un Modelo Lineal Generalizado Mixto; se plantea la modelación a través de un modelo Log-Acumulado para explicar el comportamiento de la variable de respuesta "Clase Diamétrica" dentro de los Grupos funcionales de plantas y los tratamientos aplicados. Este análisis se realiza en los años 1992, 1994, 1996 y 1998.

#### Modelado de variables ordinales

La variable de respuesta posee 3 categorías ordinales, por lo tanto es posible definir 2 logits acumulados de acuerdo con la estructura de partición sugerida en la etapa descriptiva multivariada.

Una manera de utilizar la ordinalidad de la variable de respuesta es aplicando logits de probabilidades acumuladas:

$$\begin{aligned}
 P(Y \leq 1 / x) &= \pi_1(x) \\
 P(Y \leq 2 / x) &= \pi_1(x) + \pi_2(x); \\
 P(Y \leq 3 / x) &= \pi_1(x) + \pi_2(x) + \pi_3(x);
 \end{aligned}$$

Los Logits Acumulados son definidos como:

$$\begin{aligned}
 \text{Logit}[P(Y \leq 1 / x)] &= \log \frac{P(Y \leq 1 / x)}{1 - P(Y \leq 1 / x)} \\
 &= \log \frac{\pi_1(x)}{\pi_2(x)} \\
 \\
 \text{Logit}[P(Y \leq 2 / x)] &= \log \frac{P(Y \leq 2 / x)}{1 - P(Y \leq 2 / x)} \\
 &= \log \frac{\pi_1(x) + \pi_2(x)}{\pi_3(x)}
 \end{aligned}$$

Cada Logit Acumulado utiliza todas las J categorías de la variable de respuesta.

### 3.2.1 Bosque Corinto

#### **Modelo Log –Acumulado Odds proporcional:**

La estrategia de modelación será ir desde un modelo saturado donde se incluyan todos los efectos principales y sus respectivas interacciones. Donde en forma simultánea se realizará el test de disparidad proporcional, el cual realiza una prueba de hipótesis donde la hipótesis nula es que se cumpla la proporcionalidad del modelo. Lo cual significaría que el Modelo Odds Proporcional es el adecuado

- I) Se realizó la modelación de la variable de respuesta “Clase Diamétrica” a través de un Modelo Log-Acumulado, teniendo en cuenta como factores de clasificación al Tiempo (T) (92, 94, 96 y 98), a los Grupos Funcionales (GF) de plantas (con tres niveles), a los tratamientos (TR) (Testigo: N – Aprovechamiento forestal: A – Aprovechamiento forestal y Tratamiento Silvicultural: S).

Las tablas 3.1 y 3.2 muestran los análisis de efectos Tipo 3 realizados. Las salidas completas se detallan en el anexo3.

**Tabla 3.1.** Análisis de efectos Tipo 3. Resumen Bosque Corinto

Efectos	DF	Wald Chi-Square	Pr > ChiSq
T	3	2.5899	0.4593
TR	2	10.4897	0.0053
GF	2	2206.2684	<.0001
T*TR	6	3.3524	0.7635
T*GF	6	1.7148	0.9440
TR*GF	4	38.9350	<.0001
T*TR*GF	12	5.6399	0.9331

De la aplicación del modelo generalizado logit acumulado se obtuvo que las interacciones triples entre los Grupos Funcionales de plantas, el efecto Tiempo y los tratamientos no significativos, por lo tanto se analizan las interacciones dobles. La interacción entre los tratamientos y los grupos funcionales de plantas resultó altamente significativa mientras que las interacciones dobles con el tiempo no resultaron significativas. El efecto principal del tiempo, resulta no significativo.



A partir de estos resultados el análisis se continúa con un modelo reducido que contempla los efectos Tratamiento (TR) y Grupo Funcional (GF) de plantas junto con la interacción doble entre ellos.

**Tabla 3.2.** Análisis de efectos Tipo 3. Resumen Bosque Corinto

Efecto	DF	Wald Chi-Square	Pr > ChiSq
TR	2	10.6609	0.0048
GF	2	2260.1477	<.0001
TR*GF	4	38.6387	<.0001

De la aplicación de este último modelo generalizado logit acumulado se obtuvo que la interacción entre los Grupos Funcionales de planta y los distintos tratamientos resultó altamente significativa. Esta conclusión es muy importante debido a que los tratamientos están afectando a los distintos tipos de plantas, es decir que los tratamientos tanto el Silvicultural y el de aprovechamiento forestal afectan a los distintos grupos funcionales de plantas. Pero esta conclusión está sujeta al supuesto de proporcionalidad el cual todavía debe ser analizado.

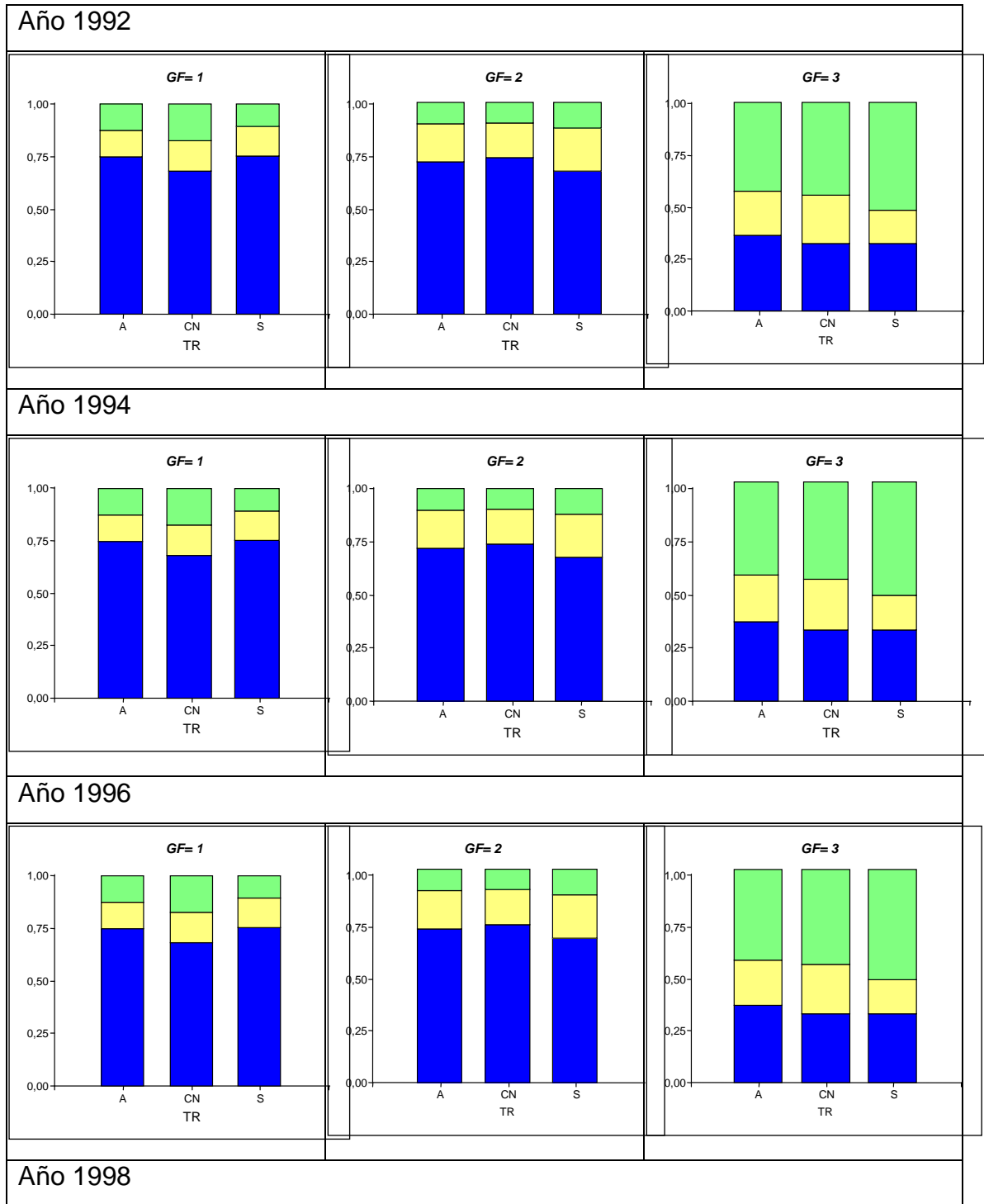
Se analizará a continuación si el supuesto de proporcionalidad se verifica para este modelo simplificado.

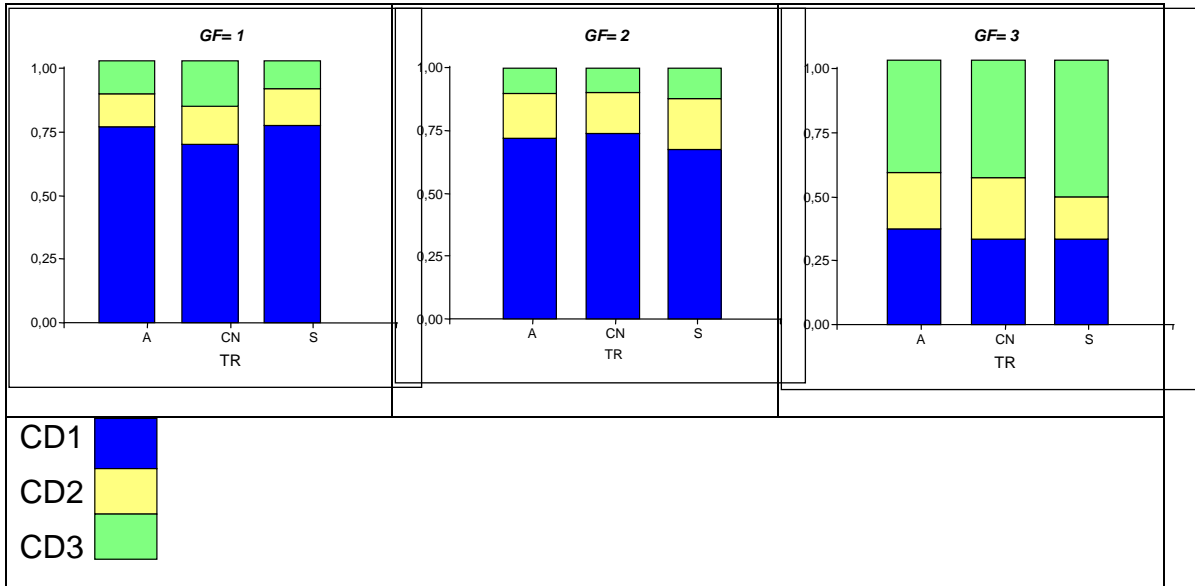
### **Análisis de Proporcionalidad**

Se analiza si se cumple la proporcionalidad tanto gráficamente como analíticamente mediante el Test de Puntuación para la suposición de disparidad proporcional, donde  $\beta$  es la misma para los distintos logits.

Gráficamente se analizó el cumplimiento del supuesto de proporcionalidad de la variable de respuesta dentro de cada una de las variables clasificatorias en los años 1992, 1994, 1996 y 1998. Con las probabilidades acumuladas se realizan los siguientes gráficos:

**Figura 3.6.** Gráfico de probabilidad de acumuladas. Bosque Corinto.





El análisis de la proporcionalidad mediante el uso de estos gráficos, es necesario ya que brindan información muy relevante para la modelización. El análisis de proporcionalidad mediante los software estadísticos está limitado debido a que solo lo implementa el SAS para modelos con efectos fijos. Por lo que es muy importante contar con herramientas gráficas para realizar estos análisis.

Del análisis gráfico surge que la proporcionalidad se mantiene constante a lo largo de los tratamientos con pequeñas diferencias. Y por el contrario a través de los Grupos Funcionales de plantas se observa que dicha proporcionalidad no se cumple.

Para corroborar estas conclusiones graficas, se realiza el test de disparidad proporcional en el modelo que tiene en cuenta la interacción natural existente.

**Tabla 3.3.** Test de disparidad proporcional. Bosque Corinto

Chi-Square	DF	Pr > ChiSq
113.5689	8	<.0001

Las salidas del test coinciden con en el análisis gráfico donde el supuesto de proporcionalidad no se cumple. El modelo de Odds Proporcional no es el adecuado teniendo ambos efectos (GF y TR).

Debido a que no se cumple el Modelo de odds proporcional el predictor lineal debe reescribirse de la siguiente forma:

- El predictor lineal es:  $\eta = \mu + \alpha_i + \beta_i x$   $i=1, 2$

- $\eta = \text{logit}(P(Y \leq j)) \quad j = 1,2$

Luego de arribar a que el modelo de odds proporcional no es adecuado para estas observaciones, se debe continuar con el análisis teniendo en cuenta que las observaciones fueron registradas sobre un mismo individuo (parcela) a lo largo del tiempo. Por lo que estas mediciones deben existir correlaciones propias a cada individuo. Es aquí donde surge la necesidad de tener en cuenta el efecto del individuo, y se presenta como una posible solución la aplicación de los Modelos Lineales Generalizados Mixtos.

### 3.2.2. Bosque La Tirimbina

#### **Modelo Log –Acumulado Odds proporcional:**

La estrategia de modelación es idéntica a la aplicada en el Bosque Corinto, es decir, comenzar el análisis de un modelo saturado donde se incluyan todos los efectos principales y sus respectivas interacciones. Analizando luego la proporcionalidad del modelo reducido.

- II) Se realizó la modelación de la variable de respuesta “Clase Diamétrica” a través de un Modelo Log-Acumulado, teniendo en cuenta como factores de clasificación al Tiempo (T) (92, 94, 96 y 98), a los Grupos Funcionales (GF) de plantas (con tres niveles), a los tratamientos (TR) (Aprovechamiento forestal: A – Aprovechamiento forestal y Tratamiento Silvicultural: S).

La Tabla 3.4 y 3.5 muestran los análisis de efectos Tipo 3 de cada modelización. Las salidas completas se detallan en el anexo1.

**Tabla 3.4.** Análisis de efectos Tipo 3. Resumen Bosque La Tirimbina

Efectos	DF	Wald Chi-Square	Pr > ChiSq
T	3	1.4012	0.7053
TR	1	40.8250	<.0001
GF	2	621.9238	<.0001
T*TR	3	1.1633	0.7618
T*GF	6	4.3298	0.6321
TR*GF	2	27.5017	<.0001
T*TR*GF	6	1.8316	0.9345

De la aplicación de modelo generalizado logit acumulado se obtuvo resultados similares a los obtenidos en el bosque Corinto. Las interacciones triples entre los Grupos Funcionales de plantas, el efecto Tiempo y los tratamientos no son significativas, por lo tanto se analizan las interacciones dobles. La interacción entre los tratamientos y los grupos funcionales de plantas resultó altamente significativa. Mientras que las interacciones dobles con el tiempo no resultaron significativas en ninguno de los bosques. El efecto principal tiempo resultó no significativo.

A partir de estos resultados el análisis se continúa con un modelo reducido que contempla los efectos Tratamiento (TR) y Grupo Funcional (GF) de plantas junto con la interacción doble entre ellos.

**Tabla 3.5:** Análisis de efectos Tipo 3. Resumen Bosque La Tirimbina.

Efecto	DF	Wald Chi-Square	Pr > ChiSq
TR	1	41.3034	<.0001
GF	2	631.4999	<.0001
TR*GF	2	28.3573	<.0001

De la aplicación de este último modelo generalizado logit acumulado se obtuvo, al igual que en el bosque Corinto, que la interacción entre los Grupos Funcionales de planta y los distintos tratamientos resultó altamente significativa.

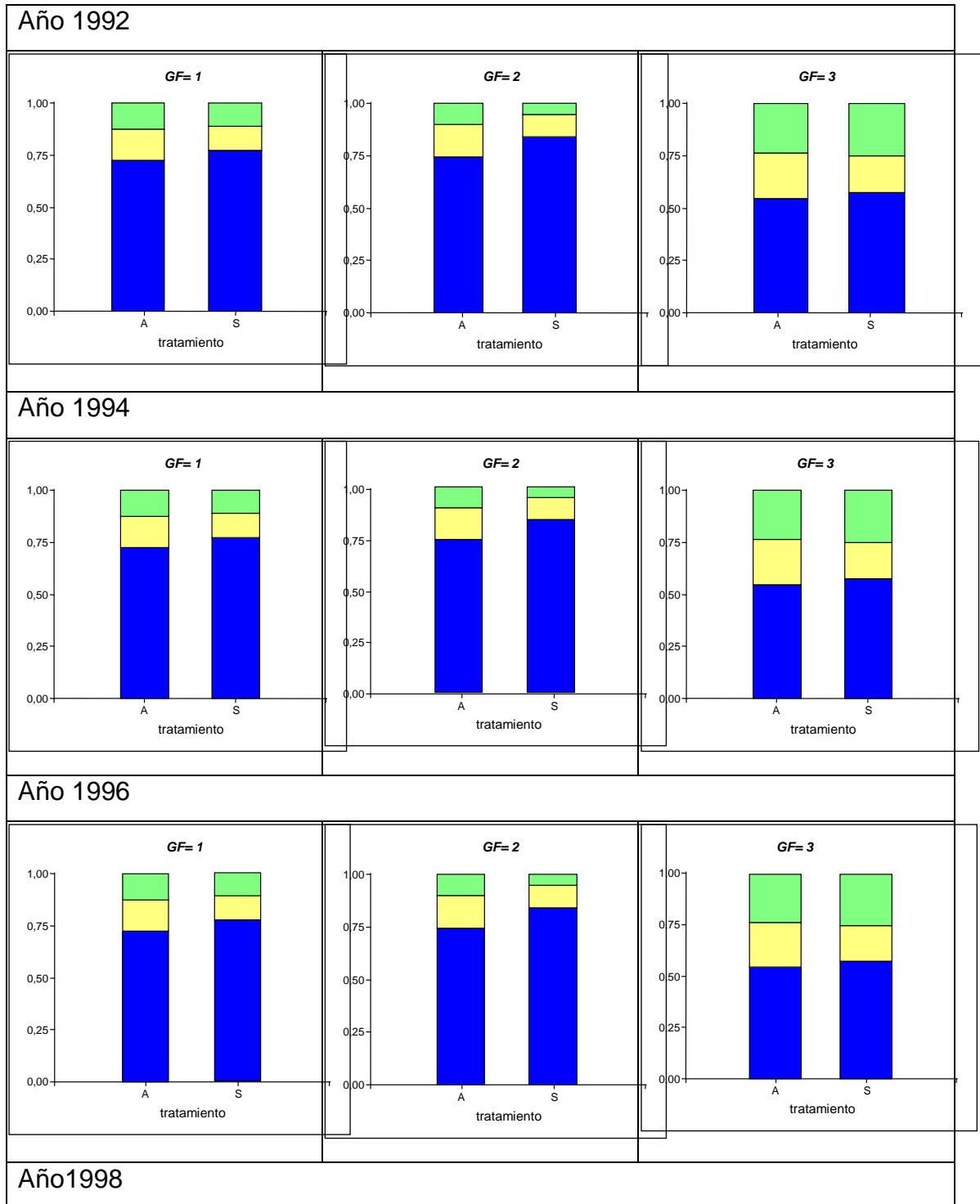
Se analizará a continuación si el supuesto de proporcionalidad se verifica en este modelo simplificado.

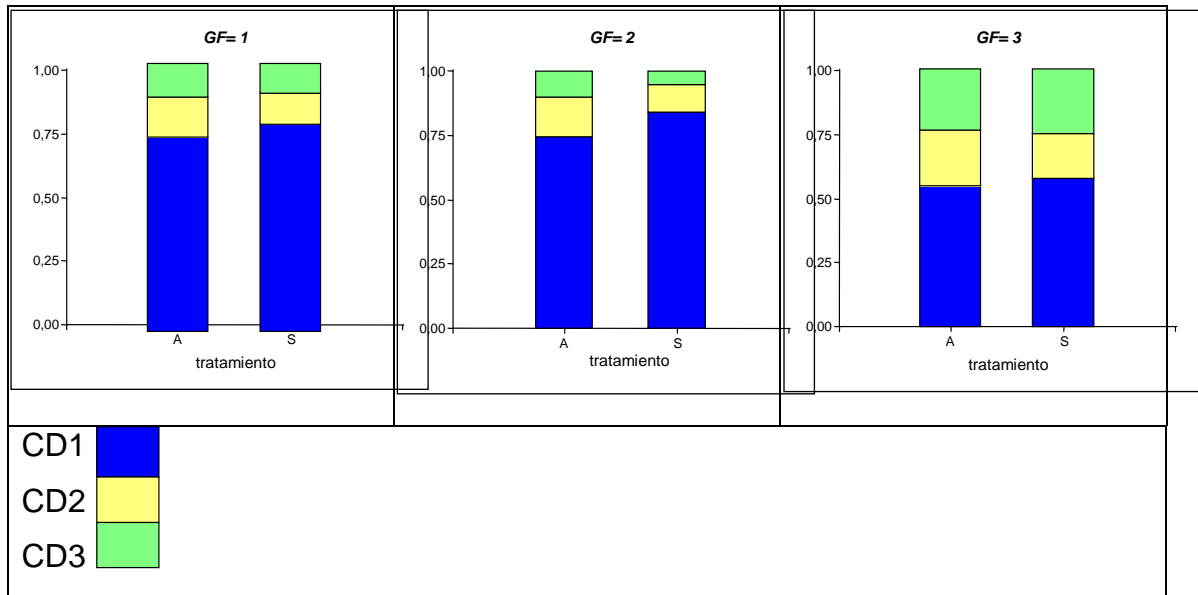
### **Análisis de Proporcionalidad**

Se analiza si se cumple la proporcionalidad tanto gráficamente como analíticamente mediante el Test de Puntuación para la suposición de disparidad proporcional.

Gráficamente se analizó el cumplimiento del supuesto de proporcionalidad de la variable de respuesta dentro de cada una de las variables clasificatorias en los años 1992, 1994, 1996 y 1998. Con las probabilidades acumuladas se realizan los siguientes gráficos:

**Figura 3.7:** Grafico de probabilidad de acumuladas. Bosque Tirimbina





Del análisis grafico surge que la proporcionalidad se mantiene constante a lo largo de los tratamientos con pequeñas diferencias sólo en los Grupos funcionales de plantas 1 y 3. Y por el contrario a través de los Grupos Funcionales de plantas se observa que dicha proporcionalidad no se cumple. Por lo cual se puede suponer que dicha proporcionalidad no se cumple en ninguno de los dos efectos.

Para corroborar estas conclusiones graficas, se realiza el test de disparidad proporcional en el modelo que tiene en cuenta la interacción natural existente.

**Tabla 3.6:** Test de disparidad proporcional. Bosque Tirimbina.

Chi-Square	DF	Pr > ChiSq
62.4737	5	<.0001

Las salidas del test coinciden con en el análisis gráfico donde se observa que el supuesto de proporcionalidad no se cumple en el modelo en estudio. En ambos bosques el modelo de Odds Proporcional no es el adecuado.

Luego de arribar a las mismas conclusiones que en el Bosque Corinto, donde el modelo de odds proporcional no es adecuado, se debe continuar con el análisis teniendo en cuenta que las observaciones fueron registradas sobre un mismo individuo (parcela) a lo largo del tiempo. Entre estas mediciones deben existir correlaciones propias a cada individuo. Es aquí donde surge la necesidad de tener en cuenta el efecto del individuo, y se presenta como una posible solución la aplicación de los Modelos Lineales Generalizados Mixtos.

### **3.3. Modelo Log – Acumulado Mixto para variables ordinales**

La importancia de los Modelos Lineales Generalizados Mixtos se debe que los mismos permiten controlar la variabilidad existente dentro los sujetos ya sea por el diseño del experimento o por la variabilidad propia de los sujetos. En este estudio los sujetos son identificados como las parcelas, en las mismas se identificaron tres Grupos Funcionales de plantas a los que se les realizaron distintos tratamientos los cuales fueron observados a lo largo del tiempo, y las mediciones repetidas en el tiempo están fuertemente correlacionadas dentro de los individuos, pero son independientes entre los sujetos. Hasta ahora solo se aplicaron técnicas de regresión que asumen independencia entre las medidas repetidas, siendo esta la principal razón o motivación para el uso de los Modelos Generalizados Mixtos.

Estos modelos permiten contemplar la falta de cumplimiento de los supuestos tradicionales y modelar de manera flexible complicadas estructuras de datos. Asimismo, los Modelos Mixtos permiten considerar la estructura de correlación intra - sujeto, y más aún, permiten la inclusión en el modelo de la variabilidad entre sujetos.

En este apartado se realiza la modelación teniendo en cuenta el efecto individuo (Parcela) siguiendo la misma metodología utilizada hasta ahora, partiendo desde un modelo saturado hasta arribar a un modelo más parsimonioso.

Para ello se analizará el comportamiento de las clases diamétricas a lo largo del tiempo, en los distintos grupos funcionales de plantas, en las parcelas y, lo más importante, en los tratamientos. Se plantea la modelación a través de un modelo Multinomial teniendo en cuenta las tres categorías de la variable de respuesta, es decir que la variable de respuesta Clase Diamétrica consta de dos categorías: CD1: plantas pequeñas [10-20), CD2: plantas medianas [20-30), plantas grandes [30 – 60 o más) dentro de los Grupos funcionales de plantas y los tratamientos aplicados. Este análisis se realiza en los años 1992, 1994, 1996 y 1998.

En el siguiente apartado (3.3.1) se analiza gráficamente de forma descriptiva el comportamiento de las Clases Diamétricas en los grupos funcionales y en los tratamientos a lo largo del tiempo y se observa la variabilidad de las parcelas,

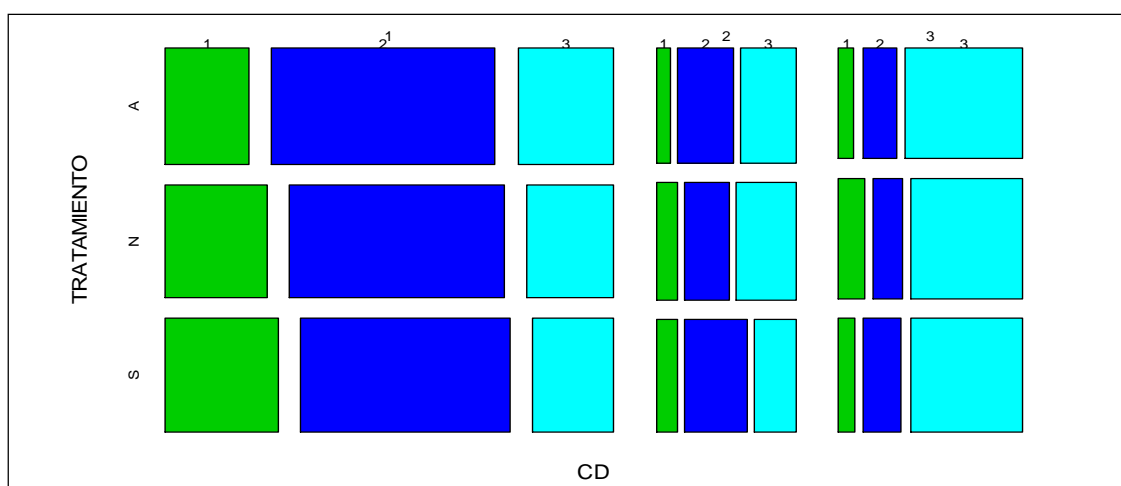


en el apartado (3.3.2) se plantea un Modelo Lineal generalizado Mixto, es decir un modelo Multinomial para la variable de respuesta Clase Diamétrica teniendo en cuenta el efecto aleatorio de la parcela, donde se comparan los modelos con y sin consideración del efecto aleatorio mediante un test de verosimilitud entre el modelo sin efecto aleatorio versus el modelo con el efecto de la parcela. En el apartado (3.3.3) se analiza si los datos se adecuan en un modelo de odds proporcional. Y por último en el apartado (3.3.4) se realizan las estimaciones del modelo definitivo. A partir de esta instancia se trabaja con el Software libre R, ya que el mismo presenta una forma práctica para realizar tanto la modelación a partir de un modelo mixto o como herramientas graficas para completar estos análisis.

### 3.3.1. Análisis descriptivo

Se muestran los gráficos de mosaicos realizados a partir de las frecuencias relativas mediante los cuales se comparó el comportamiento de las distintas clases diamétricas teniendo en cuenta todos los cruces entre los efectos a estudiar: Tratamientos, Grupos Funcionales de plantas, Tiempo e incluyendo el efecto de la Parcela.

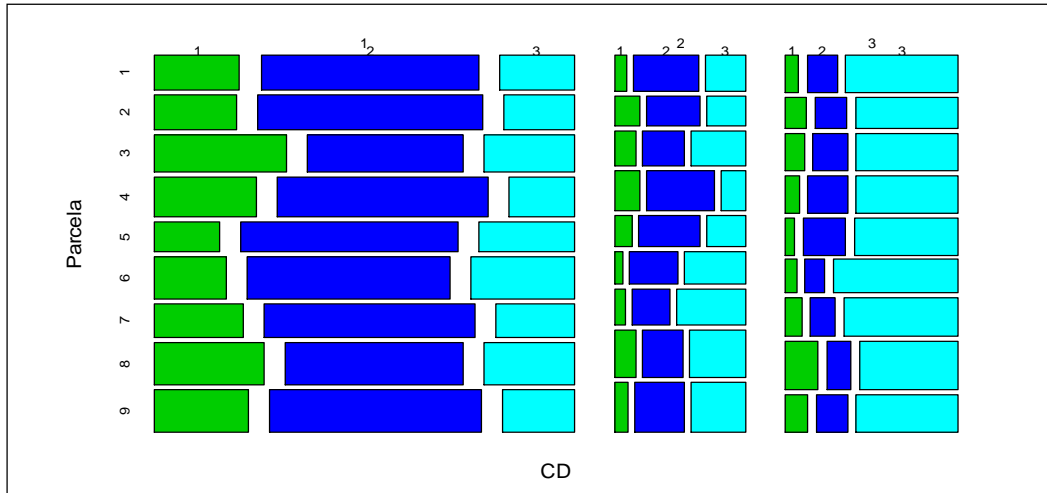
**Figura 3.8.** Gráfico de Mosaicos. Grupos Funcionales – Tratamiento – Clases Diamétricas. Bosque Corinto.



En la figura 3.8 se ve claramente que existen diferencias entre los grupos funcionales en las distintas clases diamétricas pero no en los tratamientos ya que los mosaicos son de dimensiones aproximadamente iguales.

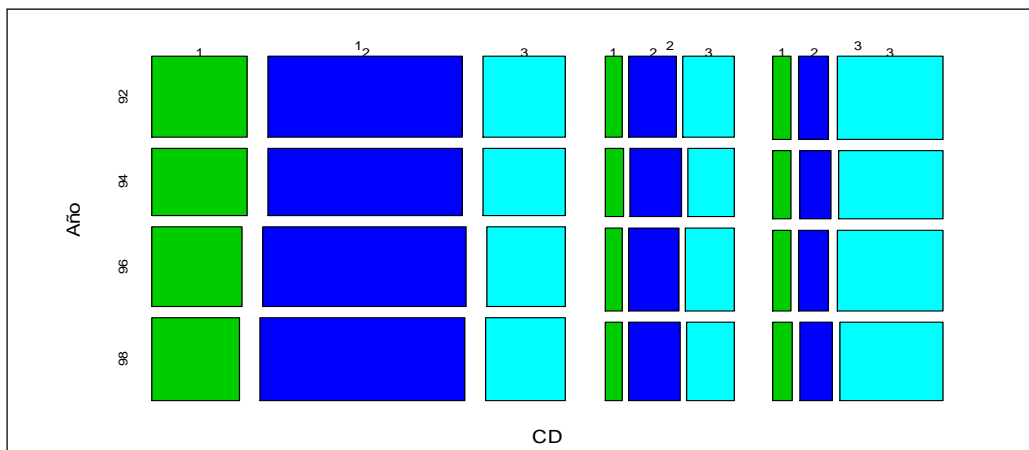
Al comparar el efecto de las parcelas y los Grupos Funcionales en las clases diamétricas (figura 3.9) se obtuvo la siguiente representación:

**Figura 3.9.** Gráfico de Mosaicos. Grupos Funcionales – Parcela - Clases Diamétricas. Bosque Corinto.



En la figura 3.9 donde se compara el efecto Parcela y el efecto Grupo Funcional en el Bosque Corinto, donde se observa que existen pequeñas diferencias en los mosaicos de las parcelas y en los grupos funcionales se acentúa la diferencia en las distintas Clases Diamétricas.

**Figura 3.10.** Gráfico de Mosaicos. Grupos Funcionales – Año – Clases Diamétricas. Bosque Corinto.

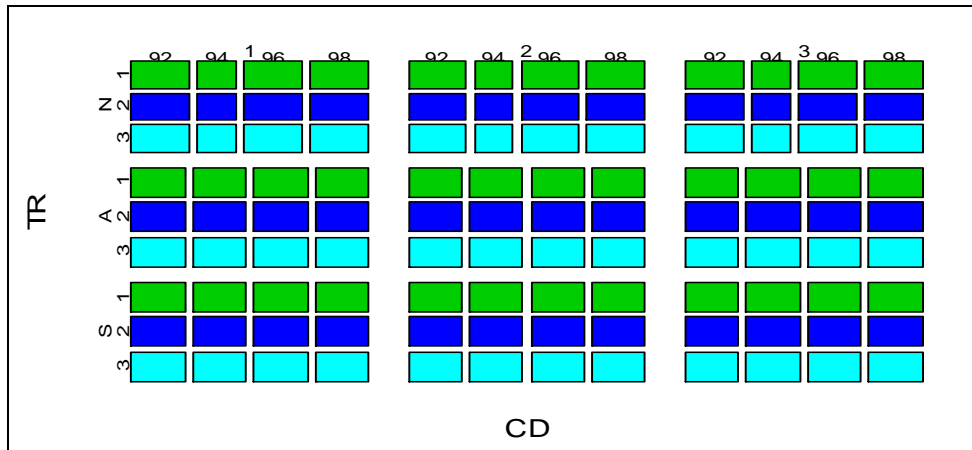


En la figura 3.10 se observa nuevamente que existen diferencias importantes en los grupos funcionales mientras que en los distintos años no existen

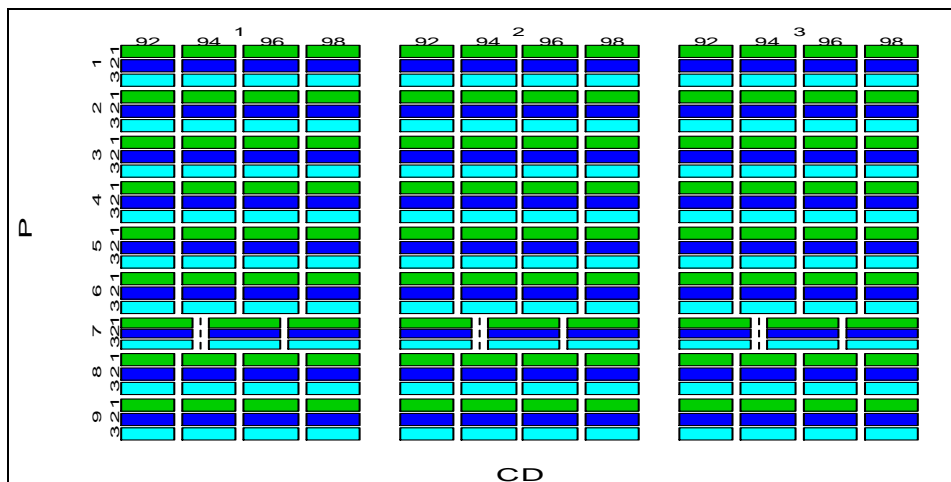
diferencias.

Por último se representaron las frecuencias observadas en gráficos de mosaicos para comparar el efecto de los Tratamientos y los Grupos Funcionales en las clases diamétricas a través de los años (figura 3.11). También se analizó el efecto de las parcelas y los Grupos Funcionales en las clases diamétricas a través de los distintos años (figura 3.12). Se obtuvieron las siguientes representaciones para el Bosque Corinto:

**Figura 3.11.** Gráfico de Mosaicos. Grupos Funcionales – Año - Tratamiento. Bosque Corinto.



**Figura 3.12.** Gráfico de Mosaicos. Grupos Funcionales versus Años versus Parcela. Bosque Corinto.



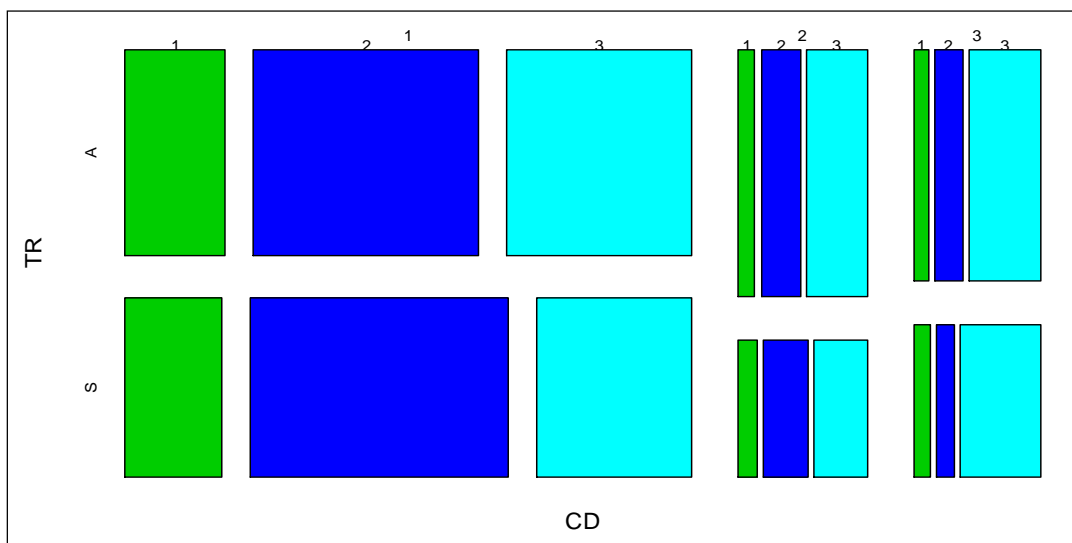
En la figura 3.11 se ve claramente que existen diferencias entre los grupos funcionales en las distintas clases diamétricas pero no en los tratamientos ni en

los años ya que los mosaicos son aproximadamente iguales. En la representación 3.12, donde se compara el efecto Parcela y el efecto Grupo Funcional a través de los distintos años existen pequeñas diferencias en los mosaicos referidos a las parcelas, en los grupos funcionales se acentúa la diferencia y en los distintos años se puede observar que no existen diferencias para ambas clases diamétricas. Se observa claramente que la parcela 7 no fue relevada en alguno de los periodos analizados, dicha parcela no pudo ser muestreadas en el año 1994.

Este análisis descriptivo proporciona información que es de gran utilidad al momento de modelar.

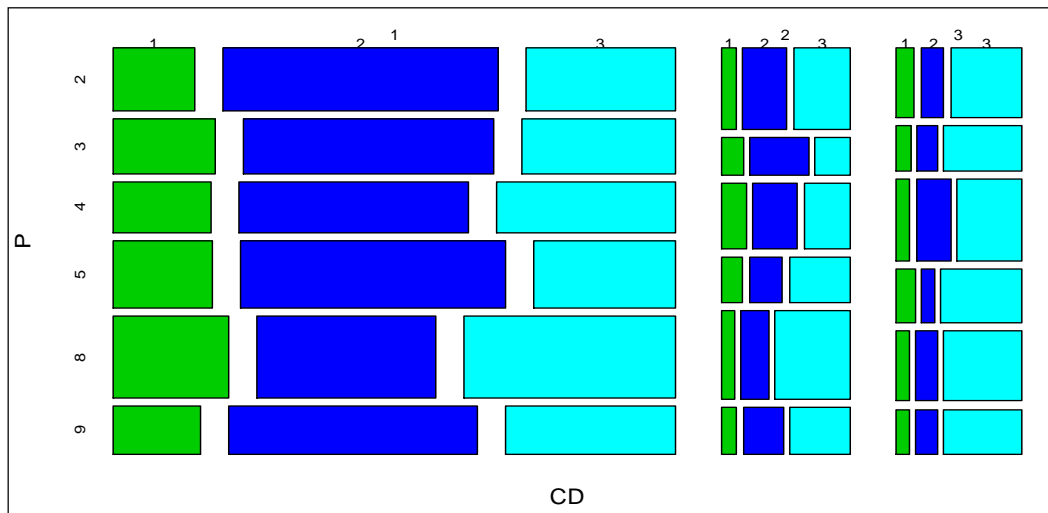
A continuación se realiza un análisis similar en el Bosque La Tirimbina.

**Figura 3.13:** Gráfico de Mosaicos. Grupos Funcionales – Tratamiento – Clases Diamétricas. Bosque Tirimbina



En la figura 3.13 se ve claramente que existen diferencias entre los grupos funcionales en las distintas clases diamétricas y en los tratamientos también. También se puede observar que hay más cantidad de plantas pequeñas respecto de las medianas y grandes.

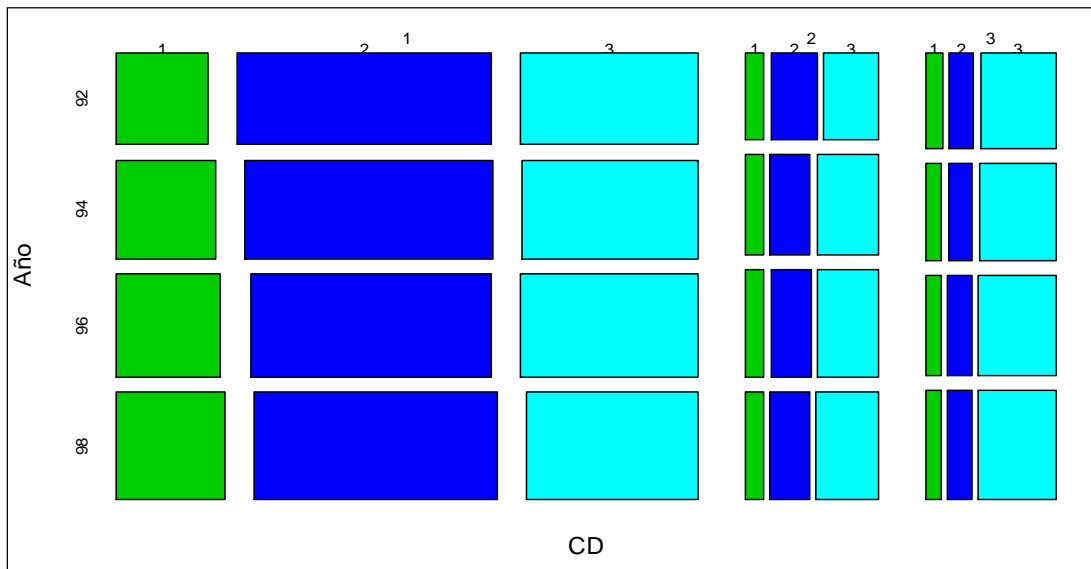
**Figura 3.14:** Gráfico de Mosaicos. Grupos Funcionales – Parcela – Clases Diamétricas. Bosque Tirimbina



En la figura 3.14 donde se compara el efecto Parcela y el efecto Grupo Funcional, se observa que existen pequeñas diferencias en los mosaicos de las parcelas y en los grupos funcionales se acentúa la diferencia en las distintas Clases Diamétricas.

Se representa además el efecto de los Grupos Funcionales a través de los años en las clases diamétricas (Figura 3.15). Obteniendo la siguiente representación:

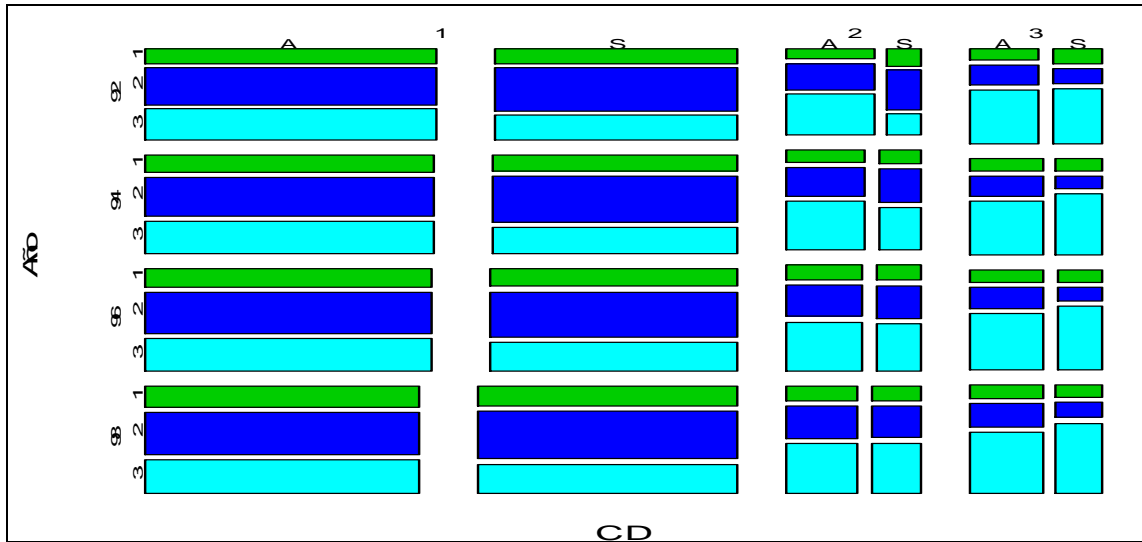
**Figura 3.15:** Gráfico de Mosaicos. Grupos Funcionales – Años – Clases Diamétricas. Bosque Tirimbina



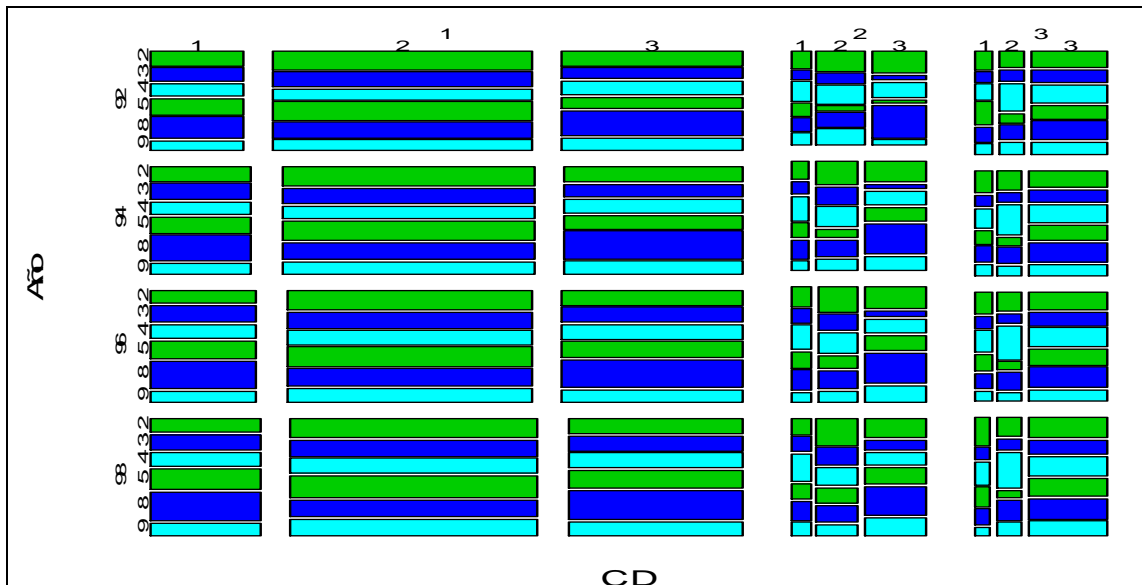
En la figura 3.15 se observa nuevamente que existen diferencias importantes en los grupos funcionales mientras que en los distintos años no existen diferencias.

Por último se representaron las frecuencias observadas en gráficos de mosaicos para comparar el efecto de los Tratamientos y los Grupos Funcionales en las clases diamétricas a través de los años (figura 3.16). También se analizó el efecto de las parcelas y los Grupos Funcionales en las clases diamétricas a través de los distintos años (figura 3.17). Se obtuvieron las siguientes representaciones:

**Figura 3.16:** Gráfico de Mosaicos. Grupos Funcionales - Años - Tratamientos. Bosque Tirimbina



**Figura 3.17:** Gráfico de Mosaicos. Grupos Funcionales - Años - Parcela. Bosque Tirimbina.



En la figura 3.16 se ve claramente que existen diferencias entre los grupos funcionales en las clases diamétricas 2 y 3 pero no en los tratamientos ni en los años ya que los mosaicos son aproximadamente iguales. En la representación 3.17, donde se compara el efecto Parcela y el efecto Grupo Funcional a través de los distintos años existen pequeñas diferencias en los mosaicos referidos a las parcelas, en los grupos funcionales se acentúa la diferencia y en los

distintos años se puede observar que no existen diferencias en las clases diamétricas. Este análisis a pesar de ser sólo descriptivo aporta mucha información que es de gran utilidad para tenerla en cuenta al momento de comenzar la modelación.

### **3.3.2 Modelo Lineal Generalizado Mixto**

Se plantea un Modelo Lineal Generalizado Mixto, es decir un modelo logit acumulado Multinomial para la variable de respuesta ordinal Clase Diamétrica, teniendo en cuenta el efecto aleatorio de la parcela. Las estimaciones se realizan con la aproximación de Laplace.

Se parte de un modelo saturado que tiene en cuenta todos los efectos principales y todas sus interacciones, continuando el análisis hasta arribar a un modelo parsimonioso que mejor ajuste los datos con la menor cantidad de efectos fijos. Todo el análisis se realiza en primero el Bosque Corinto y luego para el Bosque La Tirimbina.

Para comparar los modelos y verificar el ajuste se utilizó el valor del AIC (criterio de Akaike), el valor del Hessiano y la log-verosimilitud.



## Bosque Corinto

Como primer paso del análisis se realiza la modelación de las distintas clases diamétricas teniendo en cuenta los factores: grupo funcional, tiempo (años 1992,1994, 1996 y 1998), tratamiento, las interacciones dobles y triples y el efecto aleatorio del sujeto Parcela.

Modelo especificado en lenguaje R:  $CD \sim \text{Año} + \text{TR} + \text{GF} + \text{Año} * \text{TR} + \text{Año} * \text{GF} + \text{TR} * \text{GF} + \text{Año} * \text{TR} * \text{GF} + (1 | P)$ , del cual se obtienen la siguiente información:

**Tabla 3.7.** Información del Ajuste del Modelo 1. Bosque Corinto

Enlace	threshold	nobs	logLik	AIC	niter	cond.H	Efecto aleatorio	Var
logit	flexible	14760	-13047.37	26170.75	62(15480)	7.2e+03	P	0.01438

Los resultados completos se encuentran en el anexo2 (Tabla A. 2.1 y Tabla A.2.2). De los contrastes realizados se concluye que coinciden con el análisis gráfico debido a que se observa que la interacción triple no es significativa, pero sí la interacción doble entre tratamiento y Grupo funcional. Resultando entonces que la interacción con el efecto Tiempo y la interacción triple no son significativas.

Esto implica que las clases diamétricas no se ven afectadas por el efecto Tiempo en los años analizados. Y a partir de estos resultados se vuelve a plantear el modelo pero ahora teniendo en cuenta la interacción existente entre los tratamientos y los grupos funcionales, y los efectos principales pero sin involucrar el efecto Tiempo.

Modelo 2 (especificado en lenguaje R):  $CD \sim \text{TR} + \text{GF} + \text{TR} * \text{GF} + (1 | P)$ , donde se obtienen los siguientes resultados:

**Tabla 3.8.** Información del Modelo 2. Bosque Corinto

Enlace	threshold	nobs	logLik	AIC	niter	Cond.H	Efecto Aleatorio	Var
Logit	flexible	14760	-13054.56	26131.13	31(2487)	5.5e+02	P	0.01465

Al comparar la tabla 3.7 y la tabla 3.8 se observa que el valor de la log-verosimilitud disminuye de -13047,37 a -13054,56, el valor de AIC también bajo de 26170,75 a 26131,13 y se mejora también tanto el valor del Hessiano como el número de iteraciones. Todo ello refleja una mejora en la modelización.

**Tabla 3.9:** Estimaciones del Modelo 2. Bosque Corinto.

Coeficientes	Estimación	Error Std	z	Pr(> z )
TRA	-0.38474	0.14346	-2.682	0.007322
TRS	-0.42434	0.13775	-3.081	0.002066
GF2	-0.36618	0.08185	-4.474	7.69e-06
GF3	1.43984	0.07846	18.352	< 2e-16
TRA:GF2	0.46388	0.12406	3.739	0.000185
TRS:GF2	0.68823	0.11715	5.875	4.23e-09
TRA:GF3	0.27241	0.12070	2.257	0.024013
TRS:GF3	0.46745	0.11533	4.053	5.05e-05

Al realizar la modelación con la interacción doble y los efectos principales, resulta la interacción doble entre Grupo Funcional y Tratamiento altamente significativa.

Al observar las estimaciones de las pendientes se destaca que discrepan en sus valores de forma significativa, lo cual es un indicio de que no se cumple el concepto de proporcionalidad de un modelo log – acumulado odds proporcional. Por lo tanto debe ser analizado en los siguientes apartados.

**Tabla 3.10:** Estimaciones de los umbrales del Modelo 2. Bosque Corinto.

Umbrales:	Estimación	Error Estándar	z
1 2	0.69609	0.09556	7.285
2 3	1.66765	0.09654	17.274

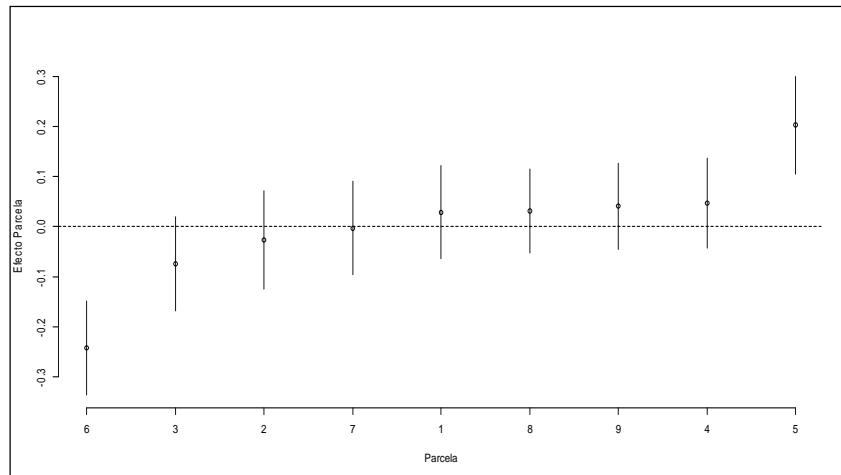
Para analizar el efecto aleatorio de un modelo log-acumulado (Christensen, R. 2013), se compara a través de un test de verosimilitud el modelo con y sin el efecto aleatorio. Por lo que se plantea dicha comparación para el efecto aleatorio de la parcela:

**Tabla 3.11.** Test de verosimilitud para el efecto de la parcela. Bosque Corinto.

corinto1b.ord CD ~ TR + GF + TR * GF					
corinto1.ord CD ~ TR + GF + TR * GF + (1   P)					
	AIC	logLik	LR	df	P(>Chi)
corinto1b.ord	26160	-13070			
corinto1.ord	26131	13055	30.505	1	3.331e-08

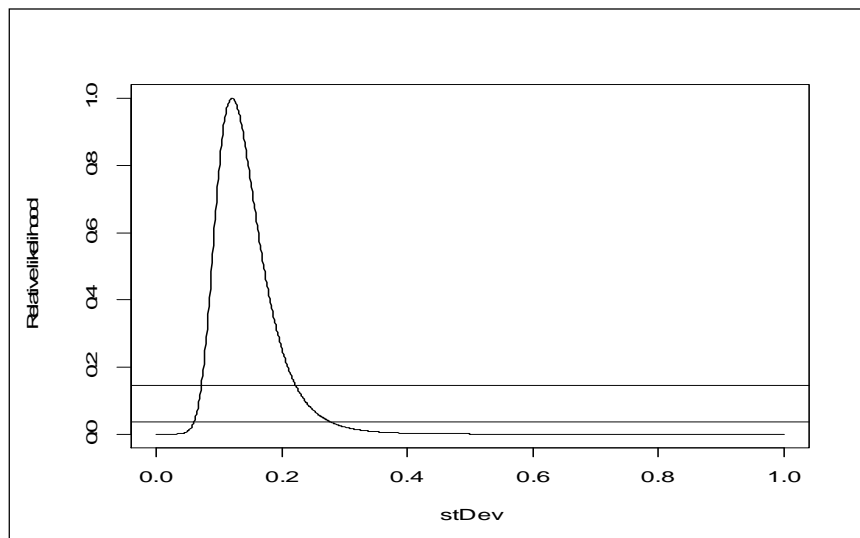
Como resultado del test de verosimilitud se obtuvo que el efecto parcela es altamente significativo, lo cuál también se refleja en el gráfico de los intervalos de confianza para cada una de las parcelas. (Figura 3.18).

**Figura 3.18.** Intervalo de confianza para los efectos predichos de cada parcela. Bosque Corinto.



El gráfico de Intervalos de Confianza (Figura 3.18) representa la variabilidad sujeta a la Parcela luego del ajuste del modelo de Logit Acumulado Mixto, que indica que existe efecto parcela. Se observa que la variabilidad de las parcelas 5 y 6 difiere del resto de las parcelas.

**Figura 3.19.** Perfil relativo de verosimilitud. Bosque Corinto.



Los valores obtenidos del efecto aleatorio en términos del desvío estándar con un 95% de confianza fueron de 0,087 y 0,22.

## Bosque La Tirimbina

Siguiendo el mismo procedimiento realizado para el bosque Corinto, se plantea en el bosque Tirimbina un Modelo lineal generalizado Mixto, es decir un modelo logit acumulado Multinomial para la variable de respuesta ordinal Clase Diamétrica, teniendo en cuenta el efecto aleatorio de la parcela. Dichas estimaciones se realizan con las técnicas de estimación de Laplace, desde un modelo saturado teniendo en cuenta todos los efectos fijos y todas sus interacciones, continuando el análisis hasta arribar a un modelo parsimonioso que mejor ajuste los datos con la menor cantidad de efectos fijos.

Como primer paso del análisis en el Bosque La Tirimbina se aplica un modelo que cuenta con la interacción triple, las interacciones dobles y los efectos principales Tiempo, Tratamiento, Grupo Funcional y el efecto aleatorio de la Parcela.

Modelo 1 especificado en lenguaje R:  $CD \sim \text{Año} + \text{TR} + \text{GF} + \text{Año} * \text{TR} + \text{Año} * \text{GF} + \text{TR} * \text{GF} + \text{Año} * \text{TR} * \text{GF} + (1|P)$ , del cual se obtienen la siguiente información:

**Tabla 3.15.** Información del Ajuste del Modelo 1. Bosque La Tirimbina.

Enlace	threshold	nobs	logLik	AIC	niter	cond.H	Efecto Aleatorio	Var
logit	flexible	11029	-8831.62	17715.23	45(7953)	3.9e+03	P	0.02754

Nuevamente los resultados de los contrastes realizados en el apartado 3.3.1 coinciden con el análisis gráfico de los mosaicos. Las interacciones triples no son significativas pero al observar las interacciones dobles ninguna es significativa salvo la interacción Grupo Funcional 2 y tratamiento Silvicultural (S) donde el p-valor es igual a 0,06 lo cual debe analizar más detenidamente. Se plantea un test de verosimilitud entre el modelo 1 y el modelo 2 que no considera la interacción triple. La salida completa se presenta en el Anexo 2, en la tabla A. 2. 3.

Modelo 2 especificado en lenguaje R:  $CD \sim \text{Año} + \text{TR} + \text{GF} + \text{Año} * \text{TR} + \text{Año} * \text{GF} + \text{TR} * \text{GF} + (1 | P)$

**Tabla 3.16.** Test de verosimilitud para comparar el Modelo 1 y el Modelo 2. Bosque La Tirimbina.

	AIC	logLik	LR	df	P(>Chi)
Tirimbina2.ord	17705	-8832.6			
Tirimbina1.ord	17715	-8831.6	1.8884	6	0.9297

Del test de verosimilitud donde se comparan ambos modelos se desprende que el modelo 1 no es significativo, es decir que la interacción triple entre el efecto Año, los tratamientos y los distintos Grupos Funcionales de plantas no es significativa. Por lo que corresponde analizar el modelo 2, cuyas estimaciones se encuentran en el Anexo 2 tabla 2.5. Del análisis de las estimaciones se concluye que solo la interacción entre el grupo funcional y tratamiento es significativa. Y luego al observar el efecto principal Tiempo se concluye que no es significativo. Esto implica que las clases diamétricas no se ven afectadas por el efecto año, es decir por el paso del tiempo en los años analizados. Por lo tanto se compara nuevamente con un test de verosimilitud el modelo con todas las interacciones dobles (modelo2) y el modelo 3 que cuenta con la interacción doble entre Tratamiento y Grupo Funcional junto a todos los efectos principales.

Modelo 3 especificado en lenguaje R:  $CD \sim \text{TR} + \text{GF} + \text{TR} * \text{GF} + (1 | P)$ .

**Tabla 3.17:** Test de verosimilitud para comparar el Modelo 2 y el Modelo 3. Bosque La Tirimbina.

	AIC	logLik	LR	df	P(>Chi)
Tirimbina3.ord	17685	-8834.7			
Tirimbina2.ord	17715	-8832.6	4.3401	12	0.9765

El modelo 3 resulta altamente significativo, es decir el modelo que tiene en cuenta la interacción entre tratamiento y Grupo funcional, junto con los efectos principales es el adecuado. En la tabla 3.17 se muestra la información del modelo 3.

**Tabla 3.18.** Información del Modelo 3. Bosque La Tirimbina.

Enlace	threshold	nobs	logLik	AIC	niter	Cond.H	Efecto Aleatorio	Var
logit	flexible	11029	-8834.73	17685.46	24(1482)	4.2e+02	P	0.02753

Al comparar la tabla 3.15 y la tabla 3.18 se observa que el valor de la log-verosimilitud disminuye de -8831.62 a -8834.73, el valor de AIC también bajo de 17715,23 a 17685,46 y mejorando también tanto el valor del Hessiano de 3900 a 420, donde el mismo debe ser inferior a  $10^4$  para asegurar su adecuado ajuste y la convergencia del modelo.

**Tabla 3.19.** Estimaciones del Modelo 3

Coeficientes	Estimación	Std. Error	z value	Pr(> z )
TRS	-0.24669	0.17078	-1.445	0.14859
GF2	-0.14878	0.08162	-1.823	0.06833
GF3	0.80848	0.07657	10.559	<2.00E-16
TRS:GF2	-0.33263	0.12849	-2.589	0.00963
TRS:GF3	0.11269	0.11995	0.939	0.34751

**Tabla 3.20.** Estimaciones de los umbrales del Modelo 3. Bosque La Tirimbina.

Umbrales	Estimaciones	Std. Error	z-value
1 2	0.9479	0.1167	8.124
2 3	19.071	0.1181	16.154

Como se mencionó anteriormente para analizar el efecto aleatorio de un modelo log-acumulado, se compara a través de un test de verosimilitud el modelo con y sin el efecto aleatorio. Por lo tanto se plantea dicha comparación para el efecto aleatorio de la parcela.

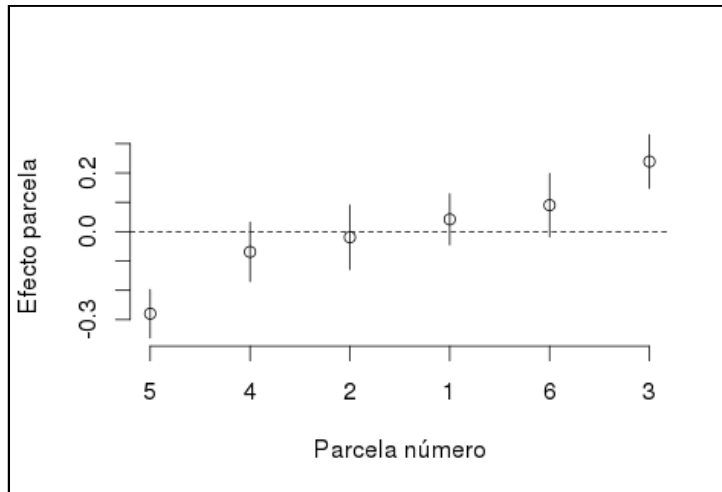
**Tabla 3.21.** Test de verosimilitud para el efecto de la parcela. Bosque La Tirimbina.

tirimbina31.ord CD ~ TR + GF + TR * GF					
tirimbina3.ord CD ~ TR + GF + TR * GF + (1   P)					
	AIC	logLik	LR	df	P(>Chi)
tirimbina31.ord	17743	-8864.6			
tirimbina3.ord	17685	-8834.7	59.823	1	1.038e-14

Como resultado del test de verosimilitud se obtuvo que el efecto parcela es

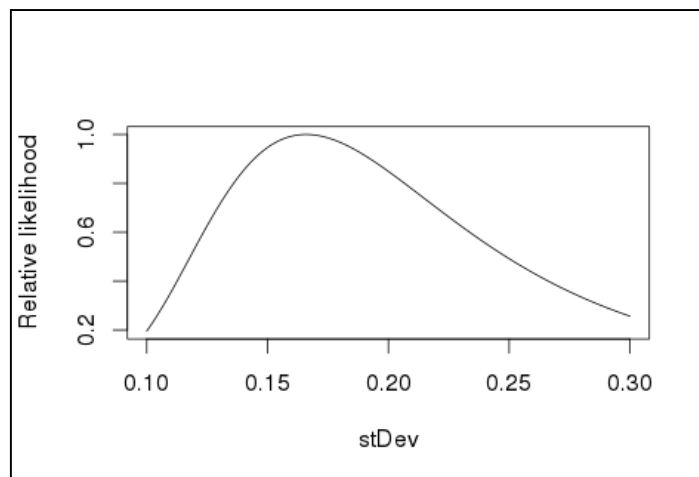
altamente significativo, lo que también se refleja en el gráfico de los intervalos de confianza para cada una de las parcelas. (Figura 3.20)

**Figura 3.20.** Intervalo de confianza para los efectos predichos de cada parcela. Bosque La Tirimbina.



El gráfico de Intervalos de Confianza (figura 3.20) representa la variabilidad sujeta a la Parcela luego del ajuste del modelo Logit Acumulado Mixto. Esto indica que existe efecto parcela, debido a la variabilidad de las parcelas 4 y 8 ya que difiere del resto de las parcelas.

**Figura 3.21.** Perfil relativo de verosimilitud. Bosque La Tirimbina.



Los valores obtenidos del efecto aleatorio en términos del desvío estándar con un 95% de confianza fueron de 0,1 y 0,3.



### **3.3.3. Análisis de Proporcionalidad**

#### **Bosque Corinto**

En este bloque se analiza si corresponde el análisis de los datos desde un modelo odds proporcional que considere una única pendiente para todos los efectos fijos. A diferencia del análisis de proporcionalidad ya realizado para estas observaciones en este apartado se tiene en cuenta el efecto de la parcela, por lo que se cuenta con un Modelo Lineal Generalizado Mixto. Por otra parte, este análisis es necesario ya que el hecho de que se cumpla o no el supuesto de proporcionalidad en el modelo de efectos fijos no implica la misma conclusión en el modelo mixto.

Se cuenta con la limitación de que tanto el software R como el SAS no poseen ningún algoritmo para testear dicho supuesto cuando se trabaja con modelos generalizados mixtos.

Por este motivo se propone estudiar la proporcionalidad en el modelo considerando el efecto aleatorio igual cero y observar dicho supuesto para una parcela típica.

Al observar la figura 3.18 se visualiza que las parcelas 1, 2, 3, 4, 7, 8 y 9 pueden considerarse como parcelas típicas. Para verificarlo se realiza el test de verosimilitud para comparar el modelo con y sin efecto aleatorio de la parcela, pero ahora considerando el modelo que solo cuenta con las parcelas promedio. Obteniendo los siguientes resultados:

**Tabla 3.22.** Test de verosimilitud para el efecto de la parcela. Bosque Corinto.

corinto1c.ord CD ~ TR + GF + TR * GF					
corinto1b.ord CD ~ TR + GF + TR * GF + (1   P)					
	AIC	logLik	LR	df	P(>Chi)
corinto1c.ord	20844	-10412			
corinto1b.ord	20846	-10412	0	1	1

Como resultado del test de verosimilitud se concluye que el efecto parcela no es significativo.

Ahora se analiza la proporcionalidad mediante el software SAS con el test de disparidad proporcional.

**Tabla 3.23.** Test de disparidad proporcional. Bosque Corinto

Chi-Square	DF	Pr > ChiSq
112.3369	8	<.0001

Las salidas del test confirman que el Modelo de Odds Proporcionales no es adecuado para las observaciones.

## Análisis de Proporcionalidad.

### Bosque La Tirimbina.

En este apartado se analiza si corresponde el análisis de los datos desde un modelo odds proporcional tal como se hizo para el Bosque Corinto.

En este sentido se propone estudiar la proporcionalidad en el modelo que considera el efecto aleatorio igual cero y observar dicho supuesto para una parcela típica.

Al observar la figura 3.20 se visualiza que las parcelas 2, 3, 5 y 9 pueden considerarse parcelas típicas. Para verificarlo se realiza el test de verosimilitud para comparar el modelo con y sin efecto aleatorio de la parcela, pero ahora considerando el modelo que solo cuenta con las parcelas promedio.

**Tabla 3.24.** Test de verosimilitud para el efecto de la parcela. Bosque La Tirimbina

tirimbina2.ord CD ~ TR + GF + TR * GF					
tirimbina1.ord CD ~ TR + GF + TR * GF + (1   P) logit flexible					
	AIC	logLik	LR	df	P(>Chi)
tirimbina2.ord	10508	-5246.9			
tirimbina1.ord	10510	-5246.8	0.0408	1	0.8399

Como resultado del test de verosimilitud se obtuvo que el efecto parcela no es significativo.

Ahora se analiza la proporcionalidad mediante el software SAS con el test de disparidad proporcional.

**Tabla 3.25.** Test de disparidad proporcional. Bosque La Tirimbina

Chi-Square	DF	Pr > ChiSq
69.8342	5	<.0001

Las salidas del test afirman que el Modelo de Odds Proporcional no es adecuado para las observaciones.

### 3.3.4. Estimación del modelo definitivo

#### Bosque Corinto

Por lo tanto el **modelo definitivo** para las observaciones correspondientes al Bosque Corinto es el que considera una pendiente diferente para cada efecto fijo. Las estimaciones del modelo son las detalladas en la Tabla 3. 9 y 3.10.

- La variable aleatoria es **Y: clase diamétrica**. Es una variable de naturaleza ordinal con tres categorías.  $Y \sim \text{Multinomial}(n_i; \pi_1, \pi_2, \pi_3)$

- El predictor lineal es:  $\eta = \mu + \alpha_i + \beta_j x \quad i=1, 2 \quad j = 1, \dots, 8$

$$\eta = \mu + \alpha_i + \beta_1 TRA + \beta_2 TRS + \beta_3 GF2 + \beta_4 GF2 + \beta_5 GF3 + \beta_6 TRA.GF2 + \beta_7 TRS.GF2 + \beta_8 TRA.GF3 + \beta_8 TRS.GF3$$

Los contrastes son realizados tomando de referencia dentro de los tratamientos el control (N) y dentro de los grupos funcionales las plantas más pequeñas (GF1).

$$\text{Logit}(Y) = 0.69609 - 0.38474 TRA - 0.42434 TRS$$

- $-0.36618 GF2 + 1.43984 GF3 + 0.46388 TRA.GF2 + 0.68823 TRS.GF2 + 0.27241 TRA.GF3 + 0.46745 TRS.GF3$

- $CD \sim TR + GF + TR * GF + (1 | P),$

- $\eta = \text{logit}(P(Y \leq j)) \quad j = 1, 2$

$$\text{Logit}(Y) = 1.66765 - 0.38474 TRA - 0.42434 TRS - 0.36618 GF2 + 1.43984 GF3 + 0.46388 TRA.GF2 + 0.68823 TRS.GF2 + 0.27241 TRA.GF3 + 0.46745 TRS.GF3$$

El modelo definitivo indica que en el bosque Corinto las clases diamétricas se

ven afectadas tanto como por los distintos grupos funcionales de plantas como con los distintos tratamientos. Es decir que los tamaños de los árboles se ven afectados al realizarles los tratamientos silviculturales y los pastoriles pero dependiendo de qué tipo de árbol sea.

## Bosque La Tirimbina

El modelo definitivo para las observaciones correspondientes al Bosque Tirimbina es el que considera una pendiente diferente para cada efecto fijo. Las estimaciones del modelo se encuentran en la Tabla 3.19 y 3.20.

- La variable aleatoria es **Y: clase diamétrica**. Es una variable de naturaleza ordinal con tres categorías.  $Y \sim \text{Multinomial}(n_i; \pi_1, \pi_2)$

- El predictor lineal es:  $\eta = \mu + \alpha_i + \beta_j x \quad i=1, 2 \quad j = 1, \dots, 8$

$$\eta = \mu + \alpha_i + \beta_1 TRS + \beta_2 GF2 + \beta_3 GF3 + \beta_4 TRS.GF2 + \beta_5 TRS.GF3$$

Los contrastes son realizados tomando de referencia dentro de los tratamientos el control (N) y dentro de los grupos funcionales las plantas más pequeñas (GF1).

- $CD \sim TR + GF + TR * GF + (1 | P),$

- $\eta = \text{logit}(P(Y \leq j)) \quad j = 1, 2$

$$\begin{aligned} \text{Logit}(Y) = & 0.9479 - 0.24669 TRS - 0.14878 GF2 + 0.80848 GF3 \\ & - 0.33263 TRS.GF2 + 0.11269 TRS.GF3 \end{aligned}$$

$$\begin{aligned} \text{Logit}(Y) = & 19.071 - 0.24669 TRS - 0.14878 GF2 + 0.80848 GF3 \\ & - 0.33263 TRS.GF2 + 0.11269 TRS.GF3 \end{aligned}$$

El modelo definitivo indica que en el bosque Tirimbina las clases diamétricas se ven afectadas por los tratamientos realizados a los distintos grupos funcionales de plantas. Es decir que los tamaños de los árboles se ven afectados al realizarles los tratamientos silviculturales y los pastoriles dependiendo de qué tipo de árbol sea. Lo cual afecta al equilibrio ecológico del Bosque. Después de los tratamientos se detectan cambios significativos en los tamaños de las diferentes plantas en estudio.



## **CAPITULO IV**

### **Conclusiones:**

- El análisis descriptivo realizado mediante herramientas multivariadas sirvió principalmente para comprender la estructura de los datos y extraer la información que contienen, antes de realizar modelizaciones más complejas. En esta aplicación hubo dos conclusiones importantes que surgieron del análisis multivariado. La primera la que tiene relación con la variable de respuesta Clase Diamétrica, ya que al realizar un análisis de componentes principales se logra identificar una posible partición o agrupación de sus categorías, separando la misma en tres categorías, la primera: la clase diamétrica de plantas más pequeñas, Clase diamétrica 1 (CD1): plantas con una clase diamétricas entre 10 - 19 cm; plantas medianas Clase diamétrica 2 (CD2): plantas con una clase diamétrica entre 20 -29 cm y por último la clase diamétrica que contiene las plantas más grandes, Clase diamétrica 3 (CD3): plantas con una clase diamétrica entre 30 -39 cm, 40 -49 cm, 50 -59 cm y 60 o más. Esta nueva partición favorece una modelización posterior. Y la segunda conclusión que surge del análisis multivariado es la que se refiere a cuales de los efectos e interacciones se identifican como significativas. Se observó que los efectos Bosque, Grupo Funcional y Año deberían ser incluidos en el modelo de análisis de los datos, mientras que el tratamiento realizado a ambos bosque no parece ser significativo.
- Al momento de la modelización existen muchos aspectos que deben ser tenidos en cuenta: Estudio longitudinal, la naturaleza ordinal de la variable de respuesta y el análisis de la proporcionalidad. Al cruzarse



todos estos puntos se complejiza el estudio, tanto en las estimaciones como en las interpretaciones finales. Como solución a cada una de las aristas de estudios longitudinales de variables de respuesta ordinales es que se analizo y se propone lo siguiente:

- ✓ Para el análisis de los estudios longitudinales o medidas repetidas, el uso de modelos mixtos que consideren al individuo como efecto aleatorio. Para estudiar la importancia del efecto aleatorio o si el mismo es significativo, se puede comenzar con un análisis gráfico de los intervalos de confianza para cada individuo para visualizar el comportamiento de cada uno de ellos y lograr también identificar aquellos individuos promedios. A continuación analizar la variancia estimada para el efecto aleatorio y ver si la misma es mayor a cero. Y por último realizar un test de verosimilitud que compare dos modelos uno con el efecto aleatorio y el otro sin él.
- ✓ Para tener en cuenta la ordinalidad de la variable se plantea un modelo Log –acumulado, el cual no viola la naturaleza ordinal. Donde mediante la estimación de los umbrales podemos ver si las divisiones de las categorías son significativas.
- ✓ El análisis de la proporcionalidad en modelos mixtos es bastante más complejo, ya que no existe ningún software que tenga implementada alguna rutina que testee dicha proporcionalidad para modelos generalizados mixtos. Pero se puede analizar de una manera sencilla, que consiste en identificar los individuos promedios, realizar un test de disparidad proporcional para poner a prueba la proporcionalidad de un modelo con efectos fijos que contenga solo a los individuos promedios.

Por lo tanto al momento de utilizar un modelo Log- Acumulado de Odds proporcional es importante analizar las ventajas y desventajas que presenta este modelo. A continuación se presentan las más destacables:

- La estimación de un modelo Log-Acumulado de odds proporcional es mucho más sencillo ya que se reducen la cantidad de parámetros a estimar, por lo que los algoritmos de estimación

suelen ser más rápidos en el proceso de estimación y por supuesto que al ser más sencillo el modelo, la interpretación de la variable de respuesta en función de los distintos efectos también suele ser más sencilla. Pero a su vez, si el modelo presenta interacciones de orden 2, 3 o más, lograr interpretar una chance puede ser muy dificultoso.

- Otro problema es el computacional, ya que el software SAS posee un test para chequear la proporcionalidad pero solo en las librerías que consideran modelos de efectos fijos. Por lo que en modelos mixtos computacionalmente no se puede realizar. Y en el software R directamente no existe hasta el momento forma de realizar este test.
  - La gran desventaja es que en muy pocas ocasiones se cumple dicho supuesto. Por lo contrario en su gran mayoría los problemas reales no son proporcionales.
  - La ventaja es que los modelos de odds no proporcionales se encuentran en la librería ordinal del software R por default, por lo que su programación es sencilla y suelen converger a pesar de contar con modelos saturados.
- ✓ Respecto a la aplicación en ambos bosques se llegó a conclusiones similares a nivel ecológico. Ambos bosques se ven afectados por los distintos tratamientos realizados, ya sean Silviculturales como de Aprovechamiento Forestal siempre considerando los distintos Grupos Funcionales de Plantas. El efecto tiempo no es significativo en los años que fueron estudiados. Lo cual puede atribuirse a que las diferencias en el tamaño de los árboles no fue significativa debido a que los mismos son muy grandes o que el crecimiento es lento o muy gradual. Pero si se ve afectado el tamaño en los mismos por los distintos tratamientos realizados a los diferentes tipos de plantas.
- ✓ La sensibilidad de los árboles en ambos bosques son similares.
- ✓ La perturbación es una parte natural de los bosques tropicales, también lo es la regeneración natural y los procesos de sucesión. Tanto en el

Bosque Corinto como en La Tirimibina se logra identificar una Regeneración Natural, es decir que existen procesos naturales mediante los cuales el ecosistema forestal se recupera después de experimentar una perturbación. Lo cual puede considerarse beneficioso ya que se produce una sucesión, decir que después de una perturbación, las comunidades vegetales se colonizan, cambian a lo largo de un periodo de tiempo y modifican al sitio perturbado.

- ✓ Se deberá definir o recurrir a las estrategias de reproducción y de crecimiento de las especies arbóreas. Logrando un enriquecimiento tanto en la variedad como en la cantidad y en lograr mantener el crecimiento de las distintas especies arbóreas.
  
- ✓ Para continuar con el estudio de estos modelos se propone estudiar la aplicabilidad de los Modelos Log-Acumulados Mixtos pero que tengan en cuenta, en el caso de existir, la variable subyacente. En muchos estudios la variable de respuesta es continua pero para facilitar el análisis se categoriza en una escala ordinal y se aplica un modelo log-acumulado sin tener en cuenta que en realidad la variable original es cuantitativa continua.

## REFERENCIAS

Agresti, A. (2002). *Categorical Data Analysis* (second edition). John Wiley & Sons, Inc.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data* (second edition). John Wiley & Sons, Inc.

Carroll, R. (2003). Variances are not always nuisance parameters. *Biometrics*.

Christensen, R. (2013). Tutorial on fitting Cumulative Link Mixed Models with `clmm2` from the `ordinal` Package, R package version 2013.9-30, <http://www.cran.r-project.org/package=ordinal/>.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for repeated measurement data*. Chapman & Hall, London.

Demidenko, E. (2004). *Mixed Models. Theory and Applications* (first edition). John Wiley & Sons.

Dra. María del Pilar Díaz (2008). *Modelos Lineales Generalizados Aplicados a las Ciencias Biológicas*. 2008.

Fitzmaurice, G. and Lard, N. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, (Vol 80 (1), pp 141–51).

Giménez G. et al. (2013). Modelo Lineal Generalizado Mixto para construir pronóstico de producción de variable de respuesta ordinal subjetiva en el alto valle de Río Negro y Neuquén. IV Encuentro Iberoamericano de Biometría y XVIII Reunión Científica del GAB. Mar del Plata, Argentina.

Hartzel, J., Agresti, A. and Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling* (Vol 1 pp. 81 – 102).

Hartzel, J., Liu, I-Ming and Agresti, A. (2001). Describing heterogeneous effects in stratified ordinal contingency tables, with application to multi-center clinical

trials. *Computational Statistics & Data Analysis* (Vol 35, pp 429 - 449).

Hedeker, D. (2005). *Longitudinal data analysis*. Reading Material for Course of Longitudinal Data Analysis.

Hedeker, D., Gibbons, R. and Flay, B. (1994). Random – Effects Regression Models for Clustered Data with an example form smoking prevention research. *Journal of Consulting and Clinical Psychology* (Vol. 62 (4), pp. 757 – 765).

Lammertyn, J., De Ketelaere, B., Marquenie, D., Molenberghs, G. and Nicolai B.M. (2003) Mixed models for multicategorical repeated response: modelling the time effect of physical treatments on strawberry sepal quality. *Postharvest Biology and Technology* (Vol. 30, pp. 195\_/207).

Lee, Y. and Nelder, J. (2004). Conditional and marginal models: Another view. *Statistical Science*. (Vol 19 (2), pp. 219–238).

Liang, K., Diggle, P., Heagerty, P. and Zeger, S. (2002). *Analysis of Longitudinal Data* (second edition). Oxford University Press.

McCulloch, C. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*. (Vol 92(437), pp. 162–170).

McCulloch, C. and Searle, S. (2001). *Generalized, Linear and Mixed Models* (first edition). John Wiley & Sons.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data* (first edition). Springer.

Pinheiro, J. and Bates, D. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*. (Vol 4(1), pp. 12–35).

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Versión obtenida 06/03/2014. URL <http://www.R-project.org/>.

Tutz, G. and Hennevogl, W. (1996). Random effects in ordinal regression models, *Computational Statistics & Data Analysis* (Vol 22, pp.537-557).

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Model for longitudinal data*. Springer Series in Statistics. Verlag, New York.

Yu, H. and Cumberland. W. (2002). *Logistic regression models for analysis of multistage survey data*. Technical report, University of California.

**Anexo1: Salidas completas del programa sas.**

**Salida1: Bosque Coritno**

Sistema SAS  
Procedimiento LOGISTIC

Información del modelo

Conjunto de datos           WORK.CORINTO  
Variable de respuesta       cd  
Número de niveles de respuesta   3  
Variable de frecuencia       freq  
Modelo                    logit acumulado  
Técnica de optimización       Puntuación de Fisher

Number of Observations Read    315  
Number of Observations Used    315  
Sum of Frequencies Read       14760  
Sum of Frequencies Used       14760

Perfil de respuesta

Valor ordenado	cd	Frecuencia total
1	1	8572
2	2	2671
3	3	3517

Las probabilidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Valor	Diseño de variables							
tiempo	92	1	0	0					
	94	0	1	0					
	96	0	0	1					
	98	-1	-1	-1					
PA	1	1	0	0	0	0	0	0	0
	2	0	1	0	0	0	0	0	0
	3	0	0	1	0	0	0	0	0
	4	0	0	0	1	0	0	0	0
	5	0	0	0	0	1	0	0	0
	6	0	0	0	0	0	1	0	0
	7	0	0	0	0	0	0	1	0
	8	0	0	0	0	0	0	0	1
	9	-1	-1	-1	-1	-1	-1	-1	-1
trat	A	1	0						
	N	0	1						
	S	-1	-1						
gf	1	1	0						

Sistema SAS  
Procedimiento LOGISTIC

Información de nivel de clase

Clase   Valor   Diseño de variables

2	0	1
3	-1	-1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
107.0724	13	<.0001

Estadístico de ajuste del modelo

Criterio	Términos independientes y	
	Sólo términos independientes	Variables adicionales
AIC	28541.319	26145.533
SC	28556.519	26259.528
-2 LOG L	28537.319	26115.533

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim Puntuación	2421.7862	13	<.0001
Wald	2371.1043	13	<.0001
	2301.0463	13	<.0001

Tipo 3 Análisis de efectos

Efecto	Chi-cuadrado		
	DF	de Wald	Pr > ChiSq
tiempo	3	3.2109	0.3602
trat	2	54.6456	<.0001
PA	6	59.5721	<.0001
gf	2	2264.6878	<.0001

NOTA: The following parameters have been set to 0, since the variables are a linear combination of

Sistema SAS  
Procedimiento LOGISTIC

other variables as shown.

tratA = -PA1 - PA3 - PA4 + PA5 + PA6  
tratN = -PA1 + PA2 - PA3 - PA4 + PA7 + PA8

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Chi-cuadrado			
		Estimador	estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.3924	0.0193	413.7387	<.0001



Intercept	2	1	1.3631	0.0224	3713.6114	<.0001
tiempo	92	1	0.00549	0.0289	0.0360	0.8494
tiempo	94	1	-0.0271	0.0313	0.7516	0.3860
tiempo	96	1	-0.0259	0.0290	0.7984	0.3716
PA	1	1	-0.1022	0.0486	4.4176	0.0356
PA	2	1	0.0260	0.0512	0.2581	0.6115
PA	3	1	0.0479	0.0488	0.9641	0.3262
PA	4	1	-0.1118	0.0467	5.7167	0.0168
PA	5	1	-0.1717	0.0507	11.4623	0.0007
PA	6	1	0.3633	0.0493	54.2385	<.0001
PA	7	1	-0.0150	0.0496	0.0919	0.7618
PA	8	1	-0.0652	0.0441	2.1906	0.1389
trat	A	0	0	.	.	.
trat	N	0	0	.	.	.
gf	1	1	0.5637	0.0301	350.7057	<.0001
gf	2	1	0.5503	0.0245	503.8608	<.0001

Estimadores de cocientes de disparidad;

Efecto	Estimador del punto	95% Wald Límites de confianza	
tiempo 92 vs 98	0.959	0.875	1.051
tiempo 94 vs 98	0.928	0.842	1.023
tiempo 96 vs 98	0.929	0.847	1.019
PA 1 vs 9	0.877	0.764	1.007
PA 2 vs 9	0.997	0.865	1.150
PA 3 vs 9	1.019	0.888	1.170
PA 4 vs 9	0.869	0.759	0.994
PA 5 vs 9	0.818	0.710	0.943
PA 6 vs 9	1.397	1.216	1.606
PA 7 vs 9	0.957	0.832	1.101
PA 8 vs 9	0.910	0.799	1.037
gf 1 vs 3	5.354	4.865	5.891
gf 2 vs 3	5.282	4.900	5.694

Sistema SAS  
Procedimiento LOGISTIC

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	67.4	D de Somers	0.373
Discordancia de porcentaje	30.2	Gamma	0.382
Porcentaje ligado	2.4	Tau-a	0.214
Pares	62437443	c	0.686

**Salida2: Tirimbina**

Sistema SAS  
Procedimiento LOGISTIC

Información del modelo

Conjunto de datos	WORK.TIRIMBINA
Variable de respuesta	cd
Número de niveles de respuesta	3
Variable de frecuencia	freq
Modelo	logit acumulado
Técnica de optimización	Puntuación de Fisher

Number of Observations Read	216
Number of Observations Used	216
Sum of Frequencies Read	11029
Sum of Frequencies Used	11029

Perfil de respuesta

Valor ordenado	cd	Frecuencia total
1	1	7592
2	2	1732
3	3	1705

Las probabilidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Valor	Diseño de variables				
tiempo	92	1	0	0		
	94	0	1	0		
	96	0	0	1		
	98	-1	-1	-1		
trat	A	1				
	S	-1				
PA	2	1	0	0	0	0
	3	0	1	0	0	0
	4	0	0	1	0	0
	5	0	0	0	1	0
	8	0	0	0	0	1
	9	-1	-1	-1	-1	-1
gf	1	1	0			
	2	0	1			
	3	-1	-1			

Sistema SAS  
Procedimiento LOGISTIC

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
76.0235	10	<.0001

Estadístico de ajuste del modelo

Criterio	Términos independientes y Variables adicionales	
	Sólo términos independientes	
AIC	18453.287	17693.915
SC	18467.904	17781.614
-2 LOG L	18449.287	17669.915

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	779.3723	10	<.0001
Puntuación	772.6042	10	<.0001
Wald	745.1756	10	<.0001

Tipo 3 Análisis de efectos

Efecto	DF	Chi-cuadrado	
		de Wald	Pr > ChiSq
tiempo	3	0.3913	0.9420
PA	5	132.7895	<.0001
trat	0	.	.
gf	2	638.8567	<.0001

NOTA: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

$$\text{tratA} = \text{PA2} - \text{PA3} + \text{PA4} - \text{PA5} + \text{PA8}$$

Sistema SAS  
Procedimiento LOGISTIC

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Estimador	Chi-cuadrado		
			estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.8854	0.0238	1381.2154	<.0001
Intercept 2	1	1.8447	0.0295	3910.8848	<.0001
tiempo 92	1	-0.00676	0.0370	0.0333	0.8551
tiempo 94	1	0.0100	0.0361	0.0770	0.7815
tiempo 96	1	0.0141	0.0356	0.1577	0.6913
PA 2	1	-0.2040	0.0438	21.6639	<.0001
PA 3	1	0.1791	0.0521	11.8299	0.0006
PA 4	1	-0.4142	0.0449	84.9154	<.0001
PA 5	1	0.2267	0.0485	21.8311	<.0001
PA 8	1	0.1680	0.0422	15.8496	<.0001
trat A	0	0	.	.	.
gf 1	1	0.1886	0.0376	25.2217	<.0001
gf 2	1	0.4753	0.0315	227.5126	<.0001

Estimadores de cocientes de disparidad;

Efecto	Estimador 95% Wald		
	del punto	Límites de confianza	
tiempo 92 vs 98	1.011	0.900	1.134
tiempo 94 vs 98	1.028	0.918	1.151
tiempo 96 vs 98	1.032	0.923	1.154
PA 2 vs 9	0.780	0.674	0.903
PA 3 vs 9	1.144	0.974	1.344
PA 4 vs 9	0.632	0.545	0.733
PA 5 vs 9	1.200	1.028	1.400
PA 8 vs 9	1.132	0.981	1.305
gf 1 vs 3	2.346	2.090	2.633
gf 2 vs 3	3.124	2.848	3.426

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	63.7	D de Somers	0.298
Discordancia de porcentaje	33.9	Gamma	0.305
Porcentaje ligado	2.4	Tau-a	0.142
Pares	29046764	c	0.649

Salida3: Bosque Corinto: Verificación del supuesto de proporcionalidad

Año 1992: Sistema SAS  
 Procedimiento LOGISTIC

Información del modelo

Conjunto de datos	WORK.CORINTO
Variable de respuesta	CD
Número de niveles de respuesta	3
Variable de frecuencia	freq
Modelo	logit acumulado
Técnica de optimización	Puntuación de Fisher

Number of Observations Read	81
Number of Observations Used	81
Sum of Frequencies Read	3865
Sum of Frequencies Used	3865

Perfil de respuesta

Valor ordenado	CD	Frecuencia total
1	1	2233
2	2	692
3	3	940

Las probabiidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Valor	Diseño de variables	
TR	A	1	0
	S	0	1
	Z	-1	-1
GF	1	-1	-1
	2	1	0
	3	0	1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
15.2626	4	0.0042

Estadístico de ajuste del modelo

Criterio	Términos independientes y Variables adicionales	
	Sólo términos independientes	
AIC	7492.793	6859.319

SC 7505.312 6896.878  
 -2 LOG L 7488.793 6847.319

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	641.4733	4	<.0001
Puntuación	630.1910	4	<.0001
Wald	610.2322	4	<.0001

Tipo 3 Análisis de efectos

Efecto	DF	Chi-cuadrado de Wald	Pr > ChiSq
TR	2	1.5107	0.4699
GF	2	610.0559	<.0001

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Estimador	Chi-cuadrado estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.4085	0.0374	119.4526	<.0001
Intercept 2	1	1.3643	0.0433	990.8434	<.0001
TR A	1	0.0357	0.0473	0.5700	0.4502
TR S	1	-0.0568	0.0466	1.4854	0.2229
GF 2	1	0.5361	0.0480	124.5264	<.0001
GF 3	1	-1.1268	0.0469	578.1992	<.0001

Procedimiento LOGISTIC

Efecto	Estimador del punto	Estimadores de cocientes de disparidad; 95% Wald Límites de confianza
TR A vs Z	1.015	0.865 1.190
TR S vs Z	0.925	0.790 1.083
GF 2 vs 1	0.947	0.784 1.144
GF 3 vs 1	0.180	0.149 0.216

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	62.2	D de Somers	0.353
Discordancia de porcentaje	26.9	Gamma	0.396
Porcentaje ligado	10.9	Tau-a	0.203

Pares	4294736	c	0.677
-------	---------	---	-------

**Año 1994:**Sistema SAS  
 Procedimiento LOGISTIC

Información del modelo

Conjunto de datos	WORK.CORINTO
Variable de respuesta	CD
Número de niveles de respuesta	3
Variable de frecuencia	freq
Modelo	logit acumulado
Técnica de optimización	Puntuación de Fisher

Number of Observations Read 72  
 Number of Observations Used 72  
 Sum of Frequencies Read 3186  
 Sum of Frequencies Used 3186

Perfil de respuesta

Valor ordenado	Frecuencia CD	Frecuencia total
1	1	1827
2	2	591
3	3	768

Las probabiidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Diseño de variables		
	Valor		
TR	A	1	0
	S	0	1
	Z	-1	-1
GF	1	-1	-1
	2	1	0
	3	0	1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
22.9681	4	0.0001

Estadístico de ajuste del modelo

Criterio	Términos independientes y Variables adicionales	
	Sólo términos independientes	
AIC	6212.594	5731.226
SC	6224.727	5767.625
-2 LOG L	6208.594	5719.226

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim Puntuación	489.3680	4	<.0001
Wald	479.5754	4	<.0001
	470.1638	4	<.0001

Tipo 3 Análisis de efectos

Efecto	DF	Chi-cuadrado de Wald	Pr > ChiSq
TR	2	4.5882	0.1009
GF	2	469.5781	<.0001

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Estimador	Chi-cuadrado estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.3616	0.0410	77.8916	<.0001
Intercept 2	1	1.3400	0.0475	796.5638	<.0001
TR A	1	0.0944	0.0507	3.4673	0.0626
TR S	1	-0.0865	0.0500	2.9964	0.0835
GF 2	1	0.5073	0.0518	95.8529	<.0001
GF 3	1	-1.0858	0.0514	446.2749	<.0001

Procedimiento LOGISTIC

Efecto	Estimador del punto	Estimadores de cocientes de disparidad; 95% Wald	Límites de confianza
TR A vs Z	1.108	0.924	1.328
TR S vs Z	0.924	0.772	1.106
GF 2 vs 1	0.931	0.760	1.142
GF 3 vs 1	0.189	0.155	0.232

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	61.8	D de Somers	0.348
Discordancia de porcentaje	27.0	Gamma	0.392
Porcentaje ligado	11.2	Tau-a	0.201
Pares	2936781	c	0.674

**Año 1996:** Sistema SAS

Procedimiento LOGISTIC

Información del modelo

Conjunto de datos	WORK.CORINTO
Variable de respuesta	CD
Número de niveles de respuesta	3
Variable de frecuencia	freq
Modelo	logit acumulado
Técnica de optimización	Puntuación de Fisher

Number of Observations Read	81
Number of Observations Used	81
Sum of Frequencies Read	3828
Sum of Frequencies Used	3828

Perfil de respuesta

Valor ordenado	Frecuencia CD	Frecuencia total
1	1	2197
2	2	710
3	3	921

Las probabiidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Diseño de variables		
	Valor		
TR	A	1	0
	S	0	1
	Z	-1	-1
GF	1	-1	-1
	2	1	0
	3	0	1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
22.7522	4	0.0001

Estadístico de ajuste del modelo

Criterio	Términos independientes y Variables adicionales	
	Sólo términos independientes	
AIC	7460.412	6819.687
SC	7472.913	6857.188
-2 LOG L	7456.412	6807.687

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim Puntuación	648.7249	4	<.0001
Wald	635.3023	4	<.0001
	617.8272	4	<.0001

Tipo 3 Análisis de efectos

Efecto	Chi-cuadrado de Wald		
	DF		Pr > ChiSq
TR	2	5.9184	0.0519
GF	2	615.8939	<.0001



Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Estimador	Chi-cuadrado estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.3619	0.0376	92.4340	<.0001
Intercept 2	1	1.3558	0.0437	961.8821	<.0001
TR A	1	0.1108	0.0477	5.3906	0.0202
TR S	1	-0.0883	0.0475	3.4529	0.0631
GF 2	1	0.5651	0.0477	140.1539	<.0001
GF 3	1	-1.1328	0.0473	573.5054	<.0001

Procedimiento LOGISTIC

Efecto	Estimador del punto	Estimadores de cocientes de disparidad; 95% Wald Límites de confianza
TR A vs Z	1.143	0.976 1.338
TR S vs Z	0.936	0.800 1.096
GF 2 vs 1	0.997	0.826 1.205
GF 3 vs 1	0.183	0.151 0.220

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	60.6	D de Somers	0.366
Discordancia de porcentaje	24.0	Gamma	0.432
Porcentaje ligado	15.4	Tau-a	0.212

Pares 4237217 c 0.683

**Año 1998:** Sistema SAS

Procedimiento LOGISTIC

Información del modelo

Conjunto de datos WORK.CORINTO  
 Variable de respuesta CD  
 Número de niveles de respuesta 3  
 Variable de frecuencia freq  
 Modelo logit acumulado  
 Técnica de optimización Puntuación de Fisher

Number of Observations Read 81  
 Number of Observations Used 81  
 Sum of Frequencies Read 3881  
 Sum of Frequencies Used 3881

Perfil de respuesta

Valor ordenado	Frecuencia CD	Frecuencia total
1	1	2315
2	2	678
3	3	888

Las probabiidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Diseño de		
	Valor	variables	
TR	A	1	0
	S	0	1
	Z	-1	-1
GF	1	-1	-1
	2	1	0
	3	0	1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
24.5020	4	<.0001

Estadístico de ajuste del modelo

Criterio	Términos independientes y Variables adicionales	
	Sólo términos independientes	
AIC	7381.438	6805.199
SC	7393.966	6842.782
-2 LOG L	7377.438	6793.199

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	584.2389	4	<.0001
Puntuación	579.1165	4	<.0001
Wald	561.2796	4	<.0001

Tipo 3 Análisis de efectos

Efecto	Chi-cuadrado		
	DF	de Wald	Pr > ChiSq
TR	2	4.0953	0.1290
GF	2	556.5290	<.0001

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Chi-cuadrado			
		Estimador	estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.4426	0.0375	139.1833	<.0001
Intercept 2	1	1.3883	0.0437	1009.7266	<.0001
TR A	1	0.0781	0.0474	2.7155	0.0994
TR S	1	0.00451	0.0483	0.0087	0.9257
GF 2	1	0.5557	0.0477	135.8497	<.0001
GF 3	1	-1.0692	0.0473	511.5277	<.0001

Procedimiento LOGISTIC

Efecto	Estimadores de cocientes de disparidad; Estimador 95% Wald Límites de confianza		
	del punto		
TR A vs Z	1.174	1.005	1.373
TR S vs Z	1.091	0.930	1.280
GF 2 vs 1	1.043	0.864	1.260
GF 3 vs 1	0.205	0.170	0.248

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	61.9	D de Somers	0.348
Discordancia de porcentaje	27.1	Gamma	0.391
Porcentaje ligado	11.1	Tau-a	0.195
Pares	4227354	c	0.674

**Salida 4: Bosque Tirimbina: Verificación del supuesto de proporcionalidad**

**Año 1992:** Sistema SAS

Procedimiento LOGISTIC

Información del modelo

Conjunto de datos	WORK.TIRIMBINA
Variable de respuesta	CD
Número de niveles de respuesta	3
Variable de frecuencia	freq
Modelo	logit acumulado
Técnica de optimización	Puntuación de Fisher

Number of Observations Read	54
Number of Observations Used	54
Sum of Frequencies Read	2509
Sum of Frequencies Used	2509

Perfil de respuesta

Valor ordenado	CD	Frecuencia total
1	1	1731
2	2	372
3	3	406

Las probabiidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Valor	Diseño de variables	
TR	A	1	
	S	-1	
GF	1	-1	-1
	2	1	0

3 0 1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
49.1218	3	<.0001

Estadístico de ajuste del modelo

Criterio	Términos independientes	
	Sólo términos independientes	y Variables adicionales
AIC	4188.034	4060.842
SC	4199.689	4089.980
-2 LOG L	4184.034	4050.842

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	133.1918	3	<.0001
Puntuación	133.4578	3	<.0001
Wald	127.2233	3	<.0001

Tipo 3 Análisis de efectos

Efecto	Chi-cuadrado		
	DF	de Wald	Pr > ChiSq
TR	1	7.3941	0.0065
GF	2	114.7332	<.0001

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Chi-cuadrado			
		Estimador	estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.8704	0.0494	310.2262	<.0001
Intercept 2	1	1.7523	0.0601	849.8232	<.0001
TR A	1	-0.1209	0.0445	7.3941	0.0065
GF	2	0.4685	0.0647	52.4152	<.0001
GF	3	-0.5542	0.0598	85.9751	<.0001

Procedimiento LOGISTIC

Efecto	Estimadores de cocientes de disparidad;		
	Estimador del punto	95% Wald	Límites de confianza
TR A vs S	0.785	0.660	0.935
GF 2 vs 1	1.466	1.136	1.893

GF 3 vs 1      0.527      0.414      0.671

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	53.3	D de Somers	0.254
Discordancia de porcentaje	27.8	Gamma	0.313
Porcentaje ligado	18.9	Tau-a	0.121
Pares	1497750	c	0.627

**Año 1994:** Sistema SAS

Procedimiento LOGISTIC

Información del modelo

Conjunto de datos	WORK.TIRIMBINA
Variable de respuesta	CD
Número de niveles de respuesta	3
Variable de frecuencia	freq
Modelo	logit acumulado
Técnica de optimización	Puntuación de Fisher

Number of Observations Read	54
Number of Observations Used	54
Sum of Frequencies Read	2713
Sum of Frequencies Used	2713

Perfil de respuesta

Valor ordenado	CD	Frecuencia total
1	1	1868
2	2	434
3	3	411

Las probabiidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Diseño de variables		
	Valor		
TR	A	1	
	S	-1	
GF	1	-1	-1
	2	1	0
	3	0	1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
--------------	----	------------

9.9221 3 0.0192

Estadístico de ajuste del modelo

Criterio	Términos independientes y Variables adicionales	
	Sólo términos independientes	
AIC	4540.359	4372.191
SC	4552.170	4401.720
-2 LOG L	4536.359	4362.191

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim Puntuación	174.1677	3	<.0001
Wald	174.2386	3	<.0001
	166.8375	3	<.0001

Tipo 3 Análisis de efectos

Efecto	Chi-cuadrado de Wald Pr > ChiSq		
	DF		
TR	1	9.1856	0.0024
GF	2	154.5413	<.0001

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Chi-cuadrado de Wald Pr > ChiSq			
		Estimador	estándar		
Intercept 1	1	0.9072	0.0480	357.0305	<.0001
Intercept 2	1	1.8852	0.0599	989.6059	<.0001
TR A	1	-0.1291	0.0426	9.1856	0.0024
GF 2	1	0.4635	0.0634	53.5157	<.0001
GF 3	1	-0.6550	0.0576	129.3332	<.0001

Procedimiento LOGISTIC

Efecto	Estimadores de cocientes de disparidad; Estimador 95% Wald Límites de confianza		
	del punto		
TR A vs S	0.772	0.654	0.913
GF 2 vs 1	1.312	1.022	1.685
GF 3 vs 1	0.429	0.340	0.542

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	54.4	D de Somers	0.273
Discordancia de porcentaje	27.1	Gamma	0.335
Porcentaje ligado	18.6	Tau-a	0.130
Pares	1756834	c	0.636

Información del modelo

Conjunto de datos           WORK.TIRIMBINA  
 Variable de respuesta       CD  
 Número de niveles de respuesta   3  
 Variable de frecuencia       freq  
 Modelo                    logit acumulado  
 Técnica de optimización       Puntuación de Fisher

Number of Observations Read    54  
 Number of Observations Used   54  
 Sum of Frequencies Read       2846  
 Sum of Frequencies Used       2846

Perfil de respuesta

Valor ordenado	CD	Frecuencia total
1	1	1961
2	2	460
3	3	425

Las probabiidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Diseño de variables		
	Valor		
TR	A	1	
	S	-1	
GF	1	-1	-1
	2	1	0
	3	0	1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
7.3397	3	0.0618

Estadístico de ajuste del modelo

Criterio	Términos independientes y Variables adicionales	
	Sólo términos independientes	
AIC	4757.779	4585.666
SC	4769.687	4615.435
-2 LOG L	4753.779	4575.666

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	178.1129	3	<.0001
Puntuación	177.9259	3	<.0001
Wald	171.1980	3	<.0001

Tipo 3 Análisis de efectos

Efecto	DF	Chi-cuadrado de Wald	Pr > ChiSq
TR	1	13.0025	0.0003
GF	2	156.9798	<.0001

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Estimador	Chi-cuadrado estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.9144	0.0465	386.1427	<.0001
Intercept 2	1	1.9064	0.0585	1061.2393	<.0001
TR A	1	-0.1495	0.0415	13.0025	0.0003
GF 2	1	0.4417	0.0617	51.2063	<.0001
GF 3	1	-0.6514	0.0560	135.3508	<.0001

Procedimiento LOGISTIC

Efecto	Estimador del cociente de disparidad	Estimador del punto	95% Wald Límites de confianza
TR A vs S	0.742	0.630	0.872
GF 2 vs 1	1.261	0.990	1.607
GF 3 vs 1	0.423	0.337	0.530

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	54.5	D de Somers	0.274
Discordancia de porcentaje	27.1	Gamma	0.336
Porcentaje ligado	18.4	Tau-a	0.131
Pares	1930985	c	0.637

**Año 1998:** Sistema SAS

Procedimiento LOGISTIC

Información del modelo

Conjunto de datos WORK.TIRIMBINA  
 Variable de respuesta CD  
 Número de niveles de respuesta 3  
 Variable de frecuencia freq  
 Modelo logit acumulado  
 Técnica de optimización Puntuación de Fisher

Number of Observations Read 54  
 Number of Observations Used 54  
 Sum of Frequencies Read 2961  
 Sum of Frequencies Used 2961



Perfil de respuesta

Valor ordenado	Frecuencia CD	Frecuencia total
1	1	2032
2	2	466
3	3	463

Las probabilidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Diseño de variables		
	Valor		
TR	A	1	
	S	-1	
GF	1	-1	-1
	2	1	0
	3	0	1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
4.6179	3	0.2020

Estadístico de ajuste del modelo

Criterio	Términos independientes y Variables adicionales	
	Sólo términos independientes	
AIC	4975.725	4770.878
SC	4987.711	4800.844
-2 LOG L	4971.725	4760.878

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim Puntuación	210.8471	3	<.0001
Wald	210.6777	3	<.0001
	202.5645	3	<.0001

Tipo 3 Análisis de efectos

Efecto	Chi-cuadrado de Wald	DF	Pr > ChiSq
--------	----------------------	----	------------

TR	1	13.7719	0.0002
GF	2	187.7777	<.0001

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Estimador	Chi-cuadrado estándar	de Wald	Pr > ChiSq
Intercept 1	1	0.9094	0.0456	398.4995	<.0001
Intercept 2	1	1.8654	0.0566	1086.3308	<.0001
TR A	1	-0.1499	0.0404	13.7719	0.0002
GF 2	1	0.4481	0.0607	54.5311	<.0001
GF 3	1	-0.7092	0.0550	166.1728	<.0001

Procedimiento LOGISTIC

Efecto	Estimadores de cocientes de disparidad;		
	Estimador del punto	95% Wald	Límites de confianza
TR A vs S	0.741	0.633	0.868
GF 2 vs 1	1.206	0.950	1.530
GF 3 vs 1	0.379	0.303	0.473

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	55.6	D de Somers	0.291
Discordancia de porcentaje	26.5	Gamma	0.355
Porcentaje ligado	18.0	Tau-a	0.140
Pares	2103486	c	0.645

**Salida 5: Corinto.**

Sistema SAS  
Procedimiento LOGISTIC

Información del modelo

Conjunto de datos	WORK.CORINTO
Variable de respuesta	CD
Número de niveles de respuesta	3
Variable de frecuencia	freq
Modelo	logit acumulado
Técnica de optimización	Puntuación de Fisher

Number of Observations Read	315
Number of Observations Used	315
Sum of Frequencies Read	14760
Sum of Frequencies Used	14760

Perfil de respuesta

Valor ordenado	Frecuencia CD	Frecuencia total
1	1	8572
2	2	2671
3	3	3517

Las probabiidades modeladas están acumuladas sobre los valores ordenados inferiores.

Información de nivel de clase

Clase	Diseño de		
	Valor	variables	
TR	A	1	0
	S	0	1
	Z	-1	-1
GF	1	-1	-1
	2	1	0
	3	0	1

Estado de convergencia del modelo

Convergence criterion (GCONV=1E-8) satisfied.

Sistema SAS  
Procedimiento LOGISTIC

Test de puntuación para la suposición de disparidad proporcional

Chi-cuadrado	DF	Pr > ChiSq
84.3151	4	<.0001

Estadístico de ajuste del modelo

Criterio	Términos independientes	
	Sólo términos independientes	y Variables adicionales
AIC	28541.319	26190.436
SC	28556.519	26236.034
-2 LOG L	28537.319	26178.436

Probar hipótesis nula global: BETA=0

Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	2358.8830	4	<.0001
Puntuación	2320.4257	4	<.0001
Wald	2256.9114	4	<.0001

Tipo 3 Análisis de efectos

Efecto	Chi-cuadrado		
	DF	de Wald	Pr > ChiSq
TR	2	11.8492	0.0027
GF	2	2254.4961	<.0001

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Error Chi-cuadrado			
		Estimador estándar	de Wald	Pr > ChiSq	
Intercept 1	1	0.3942	0.0191	426.7348	<.0001
Intercept 2	1	1.3616	0.0222	3771.2409	<.0001
TR A	1	0.0802	0.0240	11.1303	0.0008
TR S	1	-0.0580	0.0239	5.8705	0.0154

---

GF	2	1	0.5427	0.0244	496.4983	<.0001
GF	3	1	-1.1051	0.0240	2112.7317	<.0001

Procedimiento LOGISTIC

Efecto	Estimadores de cocientes de disparidad;		
	Estimador del punto	95% Wald Límites de confianza	
TR A vs Z	1.108	1.021	1.202
TR S vs Z	0.965	0.890	1.046
GF 2 vs 1	0.980	0.891	1.079
GF 3 vs 1	0.189	0.172	0.208

Asociación de probabilidades predichas y respuestas observadas

Concordancia de porcentaje	62.2	D de Somers	0.353
Discordancia de porcentaje	26.9	Gamma	0.396
Porcentaje ligado	10.9	Tau-a	0.202
Pares	62437443	c	0.676

---

**Anexo2: Salidas completas del Software R.**

Salida 2.1: Tablas de Contingencia – Bosque Corinto – Clases diamétricas + Tratamientos + Grupos Funcionales.

TRATAMIENTO = A				TRATAMIENTO = N				TRATAMIENTO = S			
GF	1	2	3	GF	1	2	3	GF	1	2	3
CD				CD				CD			
1	605	1609	685	1	712	1513	599	1	805	1471	573
2	100	398	393	2	151	327	428	2	150	437	287
3	102	229	793	3	192	208	827	3	113	262	791

Salida 2.2: Tablas de Contingencia – Bosque Corinto – Clases diamétricas + Parcelas + Grupos Funcionales.

Parcela = 1				Parcela = 2				Parcela = 3			
GF	1	2	3	GF	1	2	3	GF	1	2	3
CD				CD				CD			
1	201	509	173	1	192	527	164	1	326	386	222
2	31	156	95	2	53	108	80	2	50	99	127
3	32	76	282	3	44	66	214	3	46	86	251

Parcela = 4				Parcela = 5				Parcela = 6			
GF	1	2	3	GF	1	2	3	GF	1	2	3
CD				CD				CD			
1	278	576	178	1	131	438	190	1	203	569	289
2	69	182	65	2	36	129	79	2	19	104	131
3	35	100	258	3	22	105	256	3	24	43	277

Parcela = 7				Parcela = 8				Parcela = 9			
GF	1	2	3	GF	1	2	3	GF	1	2	3
CD				CD				CD			
1	203	476	178	1	317	510	257	1	271	602	206
2	28	92	169	2	70	127	179	2	45	165	183
3	42	65	296	3	106	77	317	3	56	81	260

Salida 2.3: Tablas de Contingencia – Bosque Corinto – Clases diamétricas + Años + Grupos Funcionales.

Año = 92				Año = 94				Año = 96				Año = 98			
GF	1	2	3	GF	1	2	3	GF	1	2	3	GF	1	2	3
CD				CD				CD				CD			
1	574	1167	492	1	468	957	402	1	535	1197	465	1	545	1272	498
2	104	284	304	2	92	264	235	2	106	311	293	2	99	303	276
3	105	184	651	3	90	154	524	3	104	179	638	3	108	182	598

Salida 2.4: Tablas de Contingencia – Bosque Corinto – Clases diamétricas + Años + Grupos Funcionales

+ Tratamientos.

CD = 1, TR A

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>156</b>	<b>147</b>	<b>150</b>	<b>152</b>
<b>2</b>	<b>382</b>	<b>356</b>	<b>417</b>	<b>454</b>
<b>3</b>	<b>176</b>	<b>169</b>	<b>162</b>	<b>178</b>

CD = 2, TR A

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>26</b>	<b>24</b>	<b>24</b>	<b>26</b>
<b>2</b>	<b>98</b>	<b>98</b>	<b>101</b>	<b>101</b>
<b>3</b>	<b>105</b>	<b>93</b>	<b>101</b>	<b>94</b>

CD = 3, TR A

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>27</b>	<b>24</b>	<b>25</b>	<b>26</b>
<b>2</b>	<b>62</b>	<b>54</b>	<b>55</b>	<b>58</b>
<b>3</b>	<b>210</b>	<b>190</b>	<b>196</b>	<b>197</b>

CD = 1, TR N

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>199</b>	<b>126</b>	<b>194</b>	<b>193</b>
<b>2</b>	<b>384</b>	<b>258</b>	<b>431</b>	<b>440</b>
<b>3</b>	<b>168</b>	<b>102</b>	<b>161</b>	<b>168</b>

CD = 2, TR N

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>42</b>	<b>31</b>	<b>42</b>	<b>36</b>
<b>2</b>	<b>72</b>	<b>59</b>	<b>98</b>	<b>98</b>
<b>3</b>	<b>121</b>	<b>67</b>	<b>120</b>	<b>120</b>

CD = 3, TR N

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>49</b>	<b>38</b>	<b>51</b>	<b>54</b>
<b>2</b>	<b>53</b>	<b>36</b>	<b>59</b>	<b>60</b>
<b>3</b>	<b>222</b>	<b>130</b>	<b>237</b>	<b>238</b>

CD = 1, TR S

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>219</b>	<b>195</b>	<b>191</b>	<b>200</b>
<b>2</b>	<b>401</b>	<b>343</b>	<b>349</b>	<b>378</b>
<b>3</b>	<b>148</b>	<b>131</b>	<b>142</b>	<b>152</b>

CD = 2, TR S

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>36</b>	<b>37</b>	<b>40</b>	<b>37</b>
<b>2</b>	<b>114</b>	<b>107</b>	<b>112</b>	<b>104</b>
<b>3</b>	<b>78</b>	<b>75</b>	<b>72</b>	<b>62</b>

CD = 3, TR S

	Año			
<b>GF</b>	<b>92</b>	<b>94</b>	<b>96</b>	<b>98</b>
<b>1</b>	<b>29</b>	<b>28</b>	<b>28</b>	<b>28</b>
<b>2</b>	<b>69</b>	<b>64</b>	<b>65</b>	<b>64</b>
<b>3</b>	<b>219</b>	<b>204</b>	<b>205</b>	<b>163</b>

Salida 2.5: Tablas de Contingencia – Bosque Tirimbina – Clases diamétricas + Tratamientos + Grupos Funcionales.

TR = A

CD	GF		
	1	2	3
<b>1</b>	<b>802</b>	<b>1779</b>	<b>1473</b>
<b>2</b>	<b>163</b>	<b>367</b>	<b>582</b>
<b>3</b>	<b>140</b>	<b>243</b>	<b>643</b>

TR = S

CD	GF		
	1	2	3
<b>1</b>	<b>677</b>	<b>1785</b>	<b>1076</b>
<b>2</b>	<b>104</b>	<b>231</b>	<b>285</b>
<b>3</b>	<b>97</b>	<b>108</b>	<b>474</b>

Salida 2.6: Tablas de Contingencia – Bosque Tirimbina – Clases diamétricas + Parcelas + Grupos Funcionales.

P = 2

CD	GF		
	1	2	3
<b>1</b>	<b>210</b>	<b>702</b>	<b>385</b>
<b>2</b>	<b>49</b>	<b>148</b>	<b>188</b>
<b>3</b>	<b>54</b>	<b>66</b>	<b>203</b>

P = 3

CD	GF		
	1	2	3
<b>1</b>	<b>232</b>	<b>567</b>	<b>348</b>
<b>2</b>	<b>34</b>	<b>92</b>	<b>54</b>
<b>3</b>	<b>28</b>	<b>38</b>	<b>148</b>

P = 4

CD	GF		
	1	2	3
<b>1</b>	<b>204</b>	<b>474</b>	<b>372</b>
<b>2</b>	<b>66</b>	<b>118</b>	<b>123</b>
<b>3</b>	<b>48</b>	<b>115</b>	<b>215</b>

P = 5

	GF		
CD	1	2	3
1	271	727	391
2	40	62	114
3	44	29	181

P = 8

	GF		
CD	1	2	3
1	388	603	716
2	48	101	271
3	38	62	225

P = 9

	GF		
CD	1	2	3
1	174	491	337
2	30	77	117
3	25	41	145

Salida 2.7: Tablas de Contingencia – Bosque Tirimbina – Clases diamétricas + Años + Grupos Funcionales.

Año = 92

	GF		
CD	1	2	3
1	306	835	590
2	58	144	170
3	61	84	261

Año = 94

	GF		
CD	1	2	3
1	358	882	628
2	64	147	223
3	57	84	270

Año = 96

	GF		
CD	1	2	3
1	394	903	664
2	73	154	233
3	58	87	280

Año = 98

	GF		
CD	1	2	3
1	421	944	667
2	72	153	241
3	61	96	306

Salida 2.8: Tablas de Contingencia – Bosque Tirimbina – Clases diamétricas + Años + Grupos Funcionales + Tratamientos.

CD = 1, TR = A

	Año			
GF	92	94	96	98
1	169	197	213	223
2	421	443	460	455
3	355	372	377	369

CD = 2, TR = A

	Año			
GF	92	94	96	98
1	35	40	45	43
2	91	91	94	91
3	142	150	148	142

CD = 3, TR = A

	Año			
GF	92	94	96	98
1	32	35	35	38
2	53	59	63	68
3	150	155	164	174

CD = 1, TR = S

CD = 2, TR = S

CD = 3, TR = S

	Año				
GF	92	94	96	98	
1	137	161	181	198	
2	414	439	443	489	
3	235	256	287	298	

	Año				
GF	92	94	96	98	
1	23	24	28	29	
2	53	56	60	62	
3	28	73	85	99	

	Año				
GF	92	94	96	98	
1	29	22	23	23	
2	31	25	24	28	
3	111	115	116	132	

Salida 2.9: Tablas de Contingencia – Bosque Tirimbina – Clases diamétricas + Años + Grupos Funcionales + Parcelas.

GF = 1, CD = 1

	P					
Año	2	3	4	5	8	9
92	50	46	45	56	74	35
94	55	60	47	61	95	40
96	51	64	56	70	106	47
98	54	62	56	84	113	52

GF = 1, CD = 2

GF = 2, CD = 1

	P					
Año	2	3	4	5	8	9
92	176	140	102	171	143	103
94	178	141	116	183	149	115
96	179	140	125	176	156	127
98	169	146	131	197	155	146

GF = 2, CD = 2

GF = 3, CD = 1

	P					
Año	2	3	4	5	8	9
92	95	75	91	77	169	83
94	100	83	92	89	180	84
96	98	96	96	108	183	83
98	92	94	93	117	184	87

GF = 3, CD = 2

	P					
Año	2	3	4	5	8	9
92	12	6	14	9	9	8
94	12	8	16	10	12	6
96	14	10	17	11	14	7
98	11	10	19	10	13	9

	P					
Año	2	3	4	5	8	9
92	35	17	31	9	25	27
94	36	26	30	13	25	17
96	38	24	30	19	26	17
98	39	25	27	21	25	16

	P					
Año	2	3	4	5	8	9
92	45	10	31	8	66	10
94	48	10	33	32	69	31
96	50	13	30	36	68	36
98	45	21	29	38	68	40

GF = 1, CD = 3

	P					
Año	2	3	4	5	8	9
92	12	7	10	15	10	7
94	13	7	12	8	10	7
96	13	7	13	10	9	6
98	16	7	13	11	9	5

GF = 2, CD = 3

	P					
Año	2	3	4	5	8	9
92	15	10	25	9	13	12
94	17	9	27	7	15	9
96	17	9	30	7	16	8
98	17	10	33	6	18	12

GF = 3, CD = 3

	P					
Año	2	3	4	5	8	9
92	46	36	51	38	53	37
94	46	37	52	45	57	33
96	51	37	55	46	58	33
98	60	38	57	52	57	42



**Anexo3: Tablas de estimaciones.**

Tabla 3.1: Estimaciones del Modelo 1 – Corinto

Coefficientes	Estimaciones	Error Std	z value	Pr(> z )
Año94	0.18070	0.19367	0.933	0.35081
Año96	0.05159	0.17602	0.293	0.76944
Año98	0.04778	0.17745	0.269	0.78774
TRA	-0.30323	0.22385	-1.355	0.17554
TRS	-0.45114	0.21155	-2.133	0.03296 *
GF2	-0.37323	0.16105	-2.318	0.02047 *
GF3	1.46384	0.14958	9.787	< 2e-16 ***
Año94:TRA	-0.22169	0.29961	-0.740	0.45934
Año96:TRA	-0.09252	0.28761	-0.322	0.74770
Año98:TRA	-0.04891	0.28621	-0.171	0.86431
Año94:TRS	-0.07426	0.27793	-0.267	0.78933
Año96:TRS	0.11293	0.26472	0.427	0.66968
Año98:TRS	0.03657	0.26668	0.137	0.89093
Año94:GF2	-0.08000	0.24800	-0.323	0.74703
Año96:GF2	0.04094	0.22321	0.183	0.85447
Año98:GF2	0.03280	0.22417	0.146	0.88366
Año94:GF3	-0.20983	0.23578	-0.890	0.37351
Año96:GF3	0.03750	0.21088	0.178	0.85886
Año98:GF3	0.01308	0.21184	0.062	0.95078
TRA:GF2	0.51816	0.24392	2.124	0.03364 *
TRS:GF2	0.74662	0.23079	3.235	0.00122 **
TRA:GF3	0.25384	0.23352	1.087	0.27703
TRS:GF3	0.59561	0.22413	2.657	0.00787 **
Año94:TRA:GF2	0.11688	0.36248	0.322	0.74712
Año96:TRA:GF2	-0.10452	0.34452	-0.303	0.76160
Año98:TRA:GF2	-0.18234	0.34289	-0.532	0.59489
Año94:TRS:GF2	0.05303	0.34231	0.155	0.87687
Año96:TRS:GF2	-0.11139	0.32342	-0.344	0.73053
Año98:TRS:GF2	-0.13982	0.32520	-0.430	0.66724
Año94:TRA:GF3	0.19634	0.35053	0.560	0.57540
Año96:TRA:GF3	0.01390	0.33319	0.042	0.96673
Año98:TRA:GF3	-0.07289	0.33174	-0.220	0.82610
Año94:TRS:GF3	0.13873	0.33538	0.414	0.67912
Año96:TRS:GF3	-0.22495	0.31711	-0.709	0.47810
Año98:TRS:GF3	-0.39196	0.32004	-1.225	0.22068

Tabla 3.2: Estimaciones de los umbrales del Modelo 1 – Corinto.

Umbrales	Estimación	Error Std	z value
1 2	0.7570	0.1425	5.313
2 3	1.7295	0.1432	12.078

Tabla 3.3: Estimaciones del Modelo 1 - Tirimbina.

Coefficientes	Estimaciones	Std.Error	z-value	Pr(> z )
A94	-0.02787	0.19599	-0.142	0.8869
A96	-0.05153	0.19287	-0.267	0.7893
A98	-0.07157	0.19211	-0.373	0.7095

TRS	-0.01547	0.25443	-0.061	0.9515
GF2	-0.19845	0.17172	-1.156	0.2478
GF3	0.74775	0.16161	4.627	3.71e-06***
A94:TRS	-0.28247	0.30354	-0.931	0.3521
A96:TRS	-0.28059	0.29684	-0.945	0.3445
A98:TRS	-0.33244	0.29498	-1.127	0.2597
A94:GF2	0.02368	0.23724	0.100	0.9205
A96:GF2	0.05710	0.23388	0.244	0.8071
A98:GF2	0.10759	0.23315	0.461	0.6445
A94:GF3	0.02331	0.22276	0.105	0.9166
A96:GF3	0.06820	0.21983	0.310	0.7564
A98:GF3	0.14151	0.21920	0.646	0.5186
TRS:GF2	-0.48110	0.26419	-1.821	0.0686
TRS:GF3	-0.16919	0.25247	-0.670	0.5028
A94:TRS:GF2	0.18559	0.37240	0.498	0.6182
A96:TRS:GF2	0.19428	0.36581	0.531	0.5954
A98:TRS:GF2	0.19079	0.36289	0.526	0.5991
A94:TRS:GF3	0.38847	0.35161	1.105	0.2692
A96:TRS:GF3	0.29766	0.34462	0.864	0.3877
A98:TRS:GF3	0.38918	0.34180	1.139	0.2549

Tabla 3.4: Estimaciones de los umbrales del Modelo 1 - Tirimbina

Umbrales	Estimación	Error Std	z-value
1 2	0.9078	0.1716	5.292
2 3	18.675	0.1725	10.827

Tabla 3.5: Estimaciones del Modelo 2 - Tirimbina

Coficientes	Estimaciones	Std.Error	z-value	Pr(> z )
Año94	-0.13437	0.15810	-0.850	0.39537
Año96	-0.14033	0.15515	-0.904	0.36575
Año98	-0.18002	0.15433	-1.166	0.24345
TRS	-0.19916	0.18809	-1.059	0.28967
GF2	-0.26254	0.14063	-1.867	0.06192
GF3	0.63735	0.13277	4.800	1.58e-06
Año94:TRS	-0.02915	0.12476	-0.234	0.81524
Año96:TRS	-0.07193	0.12317	-0.584	0.55926
Año98:TRS	-0.07670	0.12167	-0.630	0.52843
Año94:GF2	0.10415	0.18272	0.570	0.56868
Año96:GF2	0.13993	0.17977	0.778	0.43633

---

Año98:GF2	0.19068	0.17857	1.068	0.28562
Año94:GF3	0.17710	0.17211	1.029	0.30348
Año96:GF3	0.18647	0.16910	1.103	0.27015
Año98:GF3	0.29833	0.16798	1.776	0.07573
TRS:GF2	-0.33475	0.12855	-2.604	0.00921
TRS:GF3	0.10903	0.12010	0.908	0.36398