



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

COORDINACIÓN DE ESTUDIOS DE POSGRADO

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS
APLICADAS Y EN SISTEMAS

ANÁLISIS DE SECUENCIAS CONDUCTUALES MEDIANTE
MODELOS LINEALES GENERALIZADOS

TESINA PARA OBTENER EL TÍTULO DE
ESPECIALISTA EN ESTADÍSTICA APLICADA

PRESENTA
NATALIA FAVILA VÁZQUEZ

MÉXICO, D.F.

2015



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

A mi familia y a Israel por su apoyo incondicional, una vez más.

A la Dra. Silvia Ruiz-Velasco por aceptar dirigir este trabajo y por ayudarme a desarrollarlo semana con semana.

Al Dr. Gustavo Bachá por aceptar formar parte de este proyecto, y por siempre mantener mi interés por la Psicología vivo.

A mis sinodales M. en C. Antonio Soriano, M. en C. Patricia Romero y M. en C. Leticia Gracia-Medrano por sus comentarios que ayudaron a mejorar este trabajo.

A mis profesores de la especialidad, agradezco su paciencia y esfuerzo al enseñarnos tantas cosas.

Resumen

Eventos biológicamente importantes ejercen un poderoso control sobre la variabilidad de la conducta, modificando directa e indirectamente la frecuencia de más de un comportamiento. El uso de una respuesta discreta como unidad de análisis ha aportado mucha información en el área de Análisis Experimental de la Conducta, no obstante, unidades conductuales más complejas permiten analizar otros procesos que resultan interesantes. En el presente trabajo se estudió la adquisición de patrones conductuales en 16 ratas, en un experimento en el que se les reforzó durante varias sesiones, la ejecución de secuencias conductuales de tres respuestas según su estructura espacio-temporal. Se analizó la conducta de las ratas mediante modelos lineales generalizados, con los cuales se examinó el efecto de varias variables, tanto en las conductas reforzadas como en las no reforzadas. El análisis de supervivencia llevado a cabo dio una primera aproximación a la ejecución de las conductas reforzadas, observándose que la estructura fue una variable importante en su ejecución. Por otro lado, la regresión logística multinivel sugirió que las ratas aprendieron la conducta reforzada, ya que el coeficiente de la variable sesión resultó positivo y significativo. Sin embargo, dado que se permitió variabilidad en los interceptos y pendientes, cada rata tuvo una curva de aprendizaje diferente. Finalmente, el ajuste de un modelo log-lineal permitió probar diferentes distribuciones para los errores cometidos por las ratas. Esto permitió ver que los errores cometidos no fueron completamente aleatorios, ya que una distribución de acuerdo a la hipótesis de contigüidad resultó en un modelo mejor. Parece claro que la estructura espacio-temporal es una dimensión importante de la conducta, con efectos sobre la distribución conductual de los animales.

Índice general

Lista de figuras	4
Lista de tablas	5
1. El estudio de la conducta	6
1.1. Introducción	6
1.2. Condicionamiento operante	8
1.3. Secuencias conductuales	13
1.3.1. Estructura de la conducta	17
2. Modelos lineales generalizados (GLM)	22
2.1. Introducción	22
2.2. Características generales de los GLM	24
2.3. Estimación de los parámetros	26
2.4. Evaluación de los modelos	28
2.4.1. Devianza	28
2.4.2. Análisis de residuales	31
3. Modelos para conteos	33
3.1. Modelos log-lineales	35
3.1.1. Interpretación de los parámetros	39
3.1.2. Evaluación del modelo	41
3.1.3. Sobredispersión	42

4. Modelos multinivel	43
4.1. Características generales	44
4.2. Modelos lineales generalizados mixtos	50
4.2.1. Regresión logística	51
4.2.2. Regresión logística multinivel	56
5. Análisis de supervivencia	58
5.1. Función de supervivencia y de riesgo	59
5.2. Curvas Kaplan Meier	63
5.2.1. Prueba log-rank	65
6. Resultados	67
6.1. Descripción del experimento	67
6.1.1. Sujetos experimentales	67
6.1.2. Material	67
6.1.3. Fases	68
6.1.4. Estructura de los datos	70
6.1.5. Objetivos	72
6.2. Análisis de supervivencia	72
6.3. Ajuste de un modelo log-lineal	84
6.4. Ajuste de una regresión logística multinivel	98
7. Discusión	110
Referencias	119

Lista de figuras

1.	Caja de Skinner. Tomada de Lieberman (2012).	12
2.	Simetría axial. Tomada de VonEye y Mun (2013).	39
3.	Interceptos con su intervalo de confianza. Tomada de Snijders y Bosker (1999).	46
4.	Modelo de interceptos aleatorios.	47
5.	Modelo de interceptos y pendientes aleatorias.	48
6.	Función logit. Tomada de Snijders y Bosker (1999).	52
7.	Función de supervivencia. Tomada de Kleinbaum y Klein (2012).	61
8.	Función de riesgo. Tomada de Lee y Wang (2003).	62
9.	Caja operante.	68
10.	Secuencias de tres respuestas a dos palancas.	69
11.	Estructura de los datos.	71
12.	Curvas Kaplan-Meier para la estructura reforzada por rata.	74
13.	Medianas por sujeto con intervalo de confianza al 95 %.	77
14.	Curvas Kaplan-Meier para cada conducta por rata del grupo AAA.	78
15.	Curvas Kaplan-Meier para cada conducta por rata del grupo ABB.	80
16.	Curvas Kaplan-Meier para cada conducta por rata del grupo ABA.	81
17.	Curvas Kaplan-Meier para cada conducta por rata del grupo AAB.	83
18.	Tablas de contingencia para las 5 sesiones consideradas.	85
19.	Mosaicplot.	86

20.	Variables para poner a prueba diferentes ejecuciones.	88
21.	Comparación de los valores predichos vs los observados.	91
22.	Valores ajustados y observados. Las líneas punteadas representan lo ajustado y las continuas lo observado.	95
23.	Gráficas del análisis de residuales.	97
24.	Proporción de respuestas correctas por ensayo.	99
25.	Gráfica de oruga.	100
26.	Probabilidades predichas por el modelo 2.	102
27.	Probabilidades predichas por el modelo 3.	104
28.	Valores predichos y observados por rata del modelo 3.	107
29.	Residuales del modelo 3 para los datos agrupados.	108
30.	Simulación de respuestas aleatorias a las dos palancas.	117

Lista de tablas

1.	Ejemplo de tabla de contingencia de $I \times J$	35
2.	Ejemplo de simetría puntual con respecto a c_{33} . Tomada de VonEye y Mun (2013).	39
3.	Secuencias reforzadas.	70
4.	Medianas de las curvas Kaplan-Meier del grupo AAA.	75
5.	Medianas de las curvas Kaplan-Meier del grupo ABB.	75
6.	Medianas de las curvas Kaplan-Meier del grupo ABA.	76
7.	Medianas de las curvas Kaplan-Meier del grupo AAB.	76
8.	Prueba log-rank por grupo.	77
9.	Comparación de modelos.	89
10.	Resultados del modelo final con distribución binomial negativa.	93
11.	Índices de ajuste.	100
12.	Resultados modelo 2.	101
13.	Resultados modelo 3.	103
14.	Comparación de modelos para datos desagregados.	105
15.	Modelos para datos agregados por sesión.	106
16.	Modelo sin los datos de la primera rata.	109

Capítulo 1

El estudio de la conducta

1.1. Introducción

La Psicología es una disciplina que abarca diversos temas desde múltiples aproximaciones, desde el estudio de la mejor manera de dar clases, hasta el análisis de las bases neurobiológicas del aprendizaje. Sin embargo, esto nos deja con la pregunta de cuál es el objeto de estudio de la Psicología. ¿Acaso es la mente, la conducta o ambos? La respuesta a esta pregunta es un tema controversial, y diferentes áreas de la Psicología no suelen estar de acuerdo en una sola respuesta. En el siglo XX surgió una rama en la Psicología que fue llamada “conductismo”, la cual se basó en la idea fundamental de que una ciencia de la conducta es posible. Esta corriente de pensamiento inició bajo el argumento de que estudiar la conciencia directamente no suele dar resultados confiables y que los datos más adecuados para resolver algunos problemas se pueden limitar al estudio de la conducta (Boring, 2006).

En los inicios de este movimiento, el método reinante en la investigación psicológica era el de introspección, en donde los individuos hacen un reporte verbal de sus sensaciones, emociones y motivaciones. No obstante, para varios investigadores de la época se volvió evidente que esto resultaba en mediciones muy subjetivas, ya que no es posible conocer la precisión de una medida si está basada en eventos privados que son inaccesibles a otros. Asimismo, muchas veces las personas no somos capaces de explicar con

exactitud las motivaciones de nuestro comportamiento, y fallamos en describir cómo es que realizamos una conducta (e.g. tratar de explicar a alguien cómo andar en bicicleta es muy complicado). Entonces, parece claro que así como nuestro sentido de la vista no es infalible, la introspección puede dar como resultado datos que no son precisos (Lieberman, 2012).

Fue así que el conductismo surgió como una reacción a los métodos de su época, buscando hacer de la Psicología una disciplina objetiva, cuyo método central no fuera la introspección. Watson fue el iniciador de este movimiento en 1913 con su artículo *Psychology as the behaviorist views it*, en el cual propuso que la Psicología tenía que redefinirse como el estudio del comportamiento (Boring, 2006). Para Watson, los eventos observables y objetivos que la Psicología debería de estudiar eran los estímulos ambientales y las respuestas conductuales de los animales, proponiendo así un modelo de caja negra sobre la conducta. En esta búsqueda, los primeros conductistas tomaron una posición extrema, dejando de lado todo análisis relacionado a eventos mentales y restringiéndose únicamente a describir la conducta y los eventos ambientales que podían ser manipulados (Mazur, 1998). En efecto, la posición inicial fue radical, sin embargo, probablemente se necesitaba tomar una posición completamente contraria a la que había en esos tiempos para lograr un cambio en el paradigma científico de la Psicología. Hoy en día, dicha aproximación se ha relajado reconociendo la importancia del estudio de los procesos cognitivos subyacentes, pero siempre buscando una aproximación científica y basada en datos conductuales.

Al hacer este cambio en el método y en el énfasis sobre la conducta como objeto de estudio, el conductismo buscaba conceptualizar a la Psicología como una ciencia natural del comportamiento de todas las especies, dentro de las cuales están los humanos. Entonces los sujetos de estudio ya no sólo eran los humanos, sino también los monos, las aves y hasta las ratas. Esto sólo se pudo dar gracias a la teoría de la evolución de Darwin, ya que ésta permitió establecer la continuidad de la mente en la escala evolutiva.

Fue así que en muchas investigaciones del conductismo se empezaron a usar diversos animales para tratar de entender los mecanismos subyacentes a la conducta. La comparación entre especies no suena como una idea fuera de lugar si pensamos en la existencia de

un continuo evolutivo como lo planteó Darwin, de manera que así como podemos ver los orígenes de características anatómicas en otras especies, también podemos ver las características de nuestros procesos mentales y conductuales en diversas especies. El supuesto de que los procesos cognitivos que se dan en los humanos pueden aparecer en otras especies en diferentes grados, fue fundamental para estudiar la conducta en animales como ratas y palomas (Baum, 2005).

Otra razón, más bien metodológica, por la que comúnmente se usan animales como sujetos experimentales, surge de la facilidad y necesidad de controlar variables ambientales, lo que permite entender la contribución de las variables de interés de manera más clara y sencilla. Asimismo, se supone que los sistemas de aprendizaje de algunos animales son relativamente más sencillos, lo que hace más fácil entender los procesos fundamentales involucrados (Domjan, 2010).

Muchos investigadores en Psicología no concuerdan con las propuestas básicas del conductismo, bajo el argumento de que la Psicología no es una ciencia y que su objeto de estudio no es la conducta. Sin embargo, la suposición de que la Psicología es una ciencia implica varios supuestos importantes para el presente estudio. El supuesto más importante, y que asumiremos como verdadero, es que el comportamiento no se da azarosamente, es decir, existen leyes que lo gobiernan y por lo tanto puede ser predicho (Lieberman, 2012). Ante la discrepancia con las demás perspectivas de la Psicología, los conductistas se han identificado en un área con los aspectos comentados previamente, que ha sido llamada “Análisis Experimental de la Conducta”(AEC) y en la cual nos enmarcaremos para el presente trabajo (Baum, 2005). Uno de los temas más estudiados en el AEC es el aprendizaje, el cual abordaremos a continuación.

1.2. Condicionamiento operante

Muchos comportamientos en el repertorio conductual de los animales son resultado del aprendizaje. El aprendizaje junto con la memoria son las bases con las que los humanos y la mayoría de los seres vivos construimos nuestras vidas. Si viviéramos en un mundo

completamente aleatorio, sin poder establecer relaciones entre nuestra conducta y sus consecuencias, o entre eventos importantes para nuestra supervivencia, entonces el proceso de aprender no tendría sentido y no estaría presente en la gran diversidad de especies en las que lo está.

Iniciemos con una definición general de aprendizaje, considerando que es aquel proceso por el cual se almacena información en el cerebro, la cual modifica la capacidad de responder a situaciones posteriores. El aprendizaje es un cambio duradero en los mecanismos de la conducta que involucra estímulos y/o respuestas específicas y que es resultado de la experiencia previa con dichos estímulos y respuestas o con otros similares. El aprendizaje se ve reflejado en la conducta, ya sea por la ejecución de una nueva respuesta o la supresión de respuestas que ocurrían previamente (Domjan, 2010).

Primero debemos hacer una diferencia importante basada en la probabilidad de que una conducta sea seguida de un evento. Existen ocasiones en las que un estímulo siempre va seguido de la misma reacción. Por ejemplo, pensemos en el reflejo de retirar la mano cuando se toca fuego. No es algo enseñado, más bien es una reacción que ya tenemos en nuestro repertorio conductual. En estos casos cuando un estímulo siempre genera la misma respuesta, hablamos de reflejos, conductas adaptativas que a lo largo de las generaciones fueron seleccionadas y establecidas en nuestro repertorio conductual. Tiene sentido que conductas como éstas se den de la misma manera en todos los sujetos de nuestra especie, ya que el resultado es 100 % seguro que será el mismo, si no quitamos la mano, nos quemaremos. Sin embargo, existen otros eventos que no podemos predecir con tal seguridad. Tomemos el ejemplo de una fuente de comida que se puede agotar, es necesario aprender a encontrar nuevas fuentes. Para esta tarea se requiere tener la capacidad para predecir los sucesos importantes del entorno, tales como el lugar y momento en que la comida estará disponible. En este caso, resulta mejor un sistema más flexible, que no siempre produzca la misma respuesta y que permita captar las regularidades particulares de cada ambiente y adaptarse a ellas. Dicho mecanismo es el aprendizaje.

Es posible hacer una analogía del aprendizaje con la selección natural. Esta última permite la adaptación de las especies a lo largo de muchas generaciones, mientras que

el aprendizaje es el mecanismo que permite la adaptación del individuo a lo largo de su vida (Domjan, 2010). Consideremos los componentes de la selección natural: 1) se requiere variabilidad en las características de los individuos, 2) las variantes deben de reproducirse, de manera que los hijos deberían de parecerse a sus padres de generación en generación y 3) entre las variantes, algunas deben de ser más exitosas que otras. Estos mismos elementos los podemos ver en el proceso de aprendizaje: 1) debe haber variabilidad en las conductas, 2) para que una conducta sea aprendida debe aparecer varias veces en el tiempo y 3) diferentes conductas llevan a diferentes resultados con diversos grados de éxito (Baum, 2005). Entonces así como en la selección natural, de una diversidad de conductas, una de ellas lleva a un resultado deseado y es seleccionada mediante el proceso de aprendizaje.

La psicología del aprendizaje es un área muy grande, que se ocupa de todos los tipos de aprendizaje en todos los tipos de criaturas. Ante tal diversidad de situaciones en las que el proceso de aprender ocurre, la investigación sobre aprendizaje ha buscado desarrollar principios generales que sean aplicables a un intervalo amplio de especies y situaciones. El tipo de aprendizaje en el que nos centraremos en este texto es el aprendizaje asociativo, en donde se aprende la relación entre dos eventos. Un caso particular es cuando el primero de estos eventos es una conducta que va seguida de un evento que es biológica o socialmente importante, lo cual suele resultar en un cambio en la probabilidad de la conducta. A esto se le llama condicionamiento operante y puede ser de dos tipos: reforzamiento, cuando el evento que sigue a la conducta es “deseable” y por lo tanto aumenta la probabilidad de ocurrencia de la respuesta; y castigo, que se refiere a cuando el evento que sigue a la conducta es “indeseable” y por lo tanto disminuye la probabilidad de la respuesta. El término condicionamiento operante hace referencia al hecho de que los sujetos operan sobre su ambiente mediante alguna conducta, la cual va seguida de una consecuencia, ya sea deseable o no (Domjan, 2010). En resumen, condicionamiento operante se refiere a los cambios en la conducta como resultado de sus consecuencias.

En el conductismo, mucha de la investigación sobre condicionamiento operante se ha centrado en el tipo de aprendizaje que no requiere de la conciencia como término explica-

tivo. Evidentemente, esto puede resultar en una limitación en cuanto a muchos aspectos interesantes de la conducta humana, sin embargo, se ha encontrado que muchas de nuestras conductas ocurren sin que tengamos conciencia de ello. En un estudio muy interesante Gosling *et al.* (1998) buscaban estudiar qué tan precisas son las personas al reportar su propia conducta. Encontraron que la correlación entre el auto-reporte de conductas y el reporte de observadores externos tenía una correlación media de 0.4, dejando ver que muchas veces no somos conscientes de nuestros propios actos.

El análisis del condicionamiento operante o instrumental en el conductismo inició con los trabajos del psicólogo Thorndike, quien estableció la ley del efecto. Después de la publicación de la teoría de la evolución de Darwin, hubo mucho interés por estudiar la inteligencia de los animales, sin embargo, los datos recolectados hasta ese momento eran mas bien anecdóticos. Thorndike fue de los primeros en buscar una medida objetiva del aprendizaje en animales (Lieberman, 2012). A partir de sus estudios con gatos y otras especies, propuso la ley del efecto, la cual expresa que si en presencia de un estímulo, una respuesta es seguida por un evento “satisfactorio”, la asociación entre el estímulo (E) y la respuesta (R) se fortalecerá. Por otro lado si la respuesta es seguida por un evento “insatisfactorio” la asociación E-R disminuirá, es decir, lo que los animales aprenden es la relación E-R (Domjan, 2010).

A muchos investigadores no les gusta la definición de eventos en términos de deseables-indeseables y satisfactorios-insatisfactorios, y por ello tradicionalmente, a un evento “deseable” se le ha llamado reforzador. Un reforzador se define como un evento que incrementa la probabilidad de una respuesta, cuando es presentado después de ella (Lieberman, 2012). Sin embargo, recientemente Baum (2005) ha propuesto una definición más general. Podemos entender a los reforzadores como eventos que filogenéticamente han sido importantes. Ejemplos de reforzadores son la comida, el sexo y refugio. El intervalo de eventos que pueden ser reforzantes es muy amplia.

La ley del efecto de Thorndike fue la base sobre la cual B. F. Skinner fundamentó el estudio de los principios del reforzamiento. Skinner fue quien introdujo las cajas operantes o también llamadas de Skinner, las cuales actualmente todavía son muy utilizadas

para hacer investigación sobre los patrones conductuales que se dan en condicionamiento operante. Las cajas tradicionales tienen una sola palanca que al ser presionada entrega una cantidad pequeña de comida (Boring, 2006). En la figura 1 podemos ver un ejemplo de una caja de Skinner. Con este instrumento Skinner proporcionó una manera de estudiar cómo los animales aprenden, en donde la unidad de análisis es cada respuesta a la palanca. Esta discretización de la conducta en apretar la palanca una vez permitió tener un registro de la conducta sencillo y elegante.

En sus inicios, Skinner escogió a ratas como sujetos experimentales y la presión de una palanca como respuesta, en busca de los principios que gobiernan el comportamiento de los animales. Skinner consideraba que partiendo de un escenario tan sencillo, sería fácil encontrar principios que generalizaran a situaciones más complejas fuera del laboratorio, de hecho fue Skinner quien descubrió muchas de las propiedades básicas del reforzamiento (Mazur, 1998).

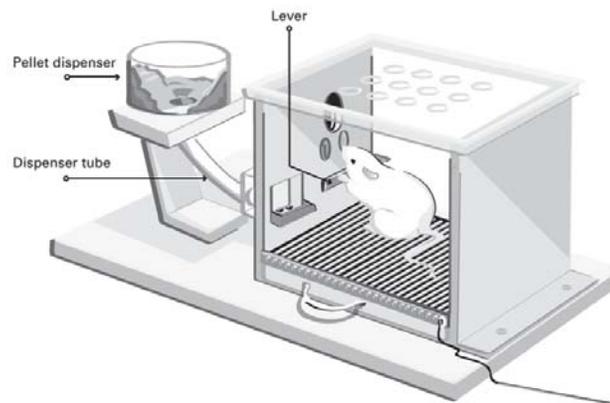


Figura 1: **Caja de Skinner**. Tomada de Lieberman (2012).

Skinner notó que el comportamiento es un continuo y que la conducta no se da en unidades tal cual, para ello propuso el concepto de **operante** como una manera de dividir el comportamiento en unidades significativas (Domjan, 2010). De acuerdo a Skinner la esencia del condicionamiento instrumental es que la conducta es controlada por sus consecuencias, de tal forma que la operante involucra tres elementos: 1) la presencia de estímulos

contextuales, 2) la ejecución de la respuesta y 3) la entrega del reforzador. Podemos entender esto como que en la presencia de ciertos estímulos, el reforzador será entregado si y solo si ocurre la respuesta adecuada. Debido a los tres elementos, Skinner llamó a esta relación la triple contingencia (Mazur, 1998). En el caso de las cajas operantes, el estímulo podría ser una luz sobre la palanca, la respuesta sería la acción de apretar la palanca y el reforzador podría ser comida o agua.

La innovación de Skinner con estas cajas fue hacer uso de una respuesta que los sujetos pudieran ejecutar de manera repetida, permitiendo un registro sencillo y continuo de la conducta. Con este cambio Skinner hizo uso de la tasa de respuesta como variable dependiente, a diferencia de las variables de interés de los estudios de su tiempo que se concentraban más en el tiempo que transcurría hasta una conducta (latencia). Así fue posible registrar los cambios en la tasa de respuesta mientras un sujeto aprendía (Mazur, 1998). Medidas como el tiempo de latencia no permiten conocer la probabilidad con la que ocurre una respuesta, sin embargo, la tasa de ocurrencia de un evento sí nos da una medida directa de dicha probabilidad (Domjan, 2010).

En resumen, los investigadores que estudian condicionamiento operante, creen que en general el reforzamiento tiene un poder sencillo pero fuerte sobre la conducta: los comportamientos que son reforzados incrementarán en frecuencia y aquellos que no lo son, disminuirán y eventualmente desaparecerán.

1.3. Secuencias conductuales

En el Análisis Experimental de la Conducta la unidad de análisis utilizada para estudiar la adquisición de patrones conductuales todavía suele ser una sola respuesta a una palanca en una caja operante. Algunos críticos del paradigma Skinneriano suelen indicar que esta medida de comportamiento es muy artificial. La crítica no es infundada, pero cabe mencionar que Skinner buscaba una conducta que no estuviera ya en el repertorio de los animales, para poder registrar el aprendizaje de una conducta novedosa. De ahí la elección aparentemente arbitraria de apretar una palanca (Lieberman, 2012). Sin

embargo, es claro que en el ambiente natural los organismos no solamente realizan una acción y obtienen comida u otro reforzador, ya que generalmente, se requieren de patrones más complejos que llevan un orden específico.

Otra crítica que se le puede hacer a esta unidad de análisis es que no hay razón para asumir que el efecto del reforzador se restringe a la respuesta discreta inmediatamente previa que lo produjo. La demora entre algunas respuestas y el reforzador puede variar, pero no por ello el reforzador no actúa sobre ellas (Catania, 1971). Asimismo, las contingencias de reforzamiento no tienen porqué fortalecer sólo aquellas operantes establecidas por el experimentador, sino que al mismo tiempo muchas otras conductas pueden experimentar cambios en su fuerza (Schick, 1971). Sin embargo, es muy común que únicamente se mida la conducta de apretar la palanca una vez, dejando de lado cualquier otra conducta. La contigüidad temporal entre la respuesta y el reforzador es fundamental, sin embargo, una sola respuesta podría ser parte de una unidad más grande de varias respuestas y al suponer que sólo se fortalece la más contigua al reforzador, se podrían estar perdiendo otras relaciones ordenadas (Shimp, 1976).

De acuerdo con Timberlake (1993), la conducta de los organismos puede ser clasificada en sistemas conductuales, los cuales están organizados alrededor de funciones importantes, como alimentación o reproducción. De esta manera, la presentación de comida de manera periódica no sólo afectará una conducta, probablemente varias conductas asociadas con la alimentación se presentarán en diferentes frecuencias. Entonces, el substrato sobre el cual el reforzamiento ejerce control no es únicamente una acción discreta, como apretar una palanca, sino patrones de respuestas más generales. Timberlake sugirió que el modelo simple de reforzamiento respuesta-reforzador podría estar sobre-simplificando un fenómeno biológico mucho más complejo.

El dividir el continuo de comportamiento en unidades de análisis que tengan significado conductual es complicado. Evidentemente no cualquier patrón de comportamiento puede funcionar como una unidad de conducta, por ello, son necesarios algunos criterios para establecer qué es una unidad conductual. Operacionalmente, en el AEC una unidad queda definida como aquella clase de comportamiento que se establece como requisito

para la obtención del reforzador. La probabilidad de ocurrencia de una unidad conductual debe de ser afectada por sus consecuencias, es decir, debe de ser condicionable (Zeiler, 1983). Asimismo, se ha sugerido que una manera de definir unidades conductuales es demostrando que las relaciones funcionales entre ellas y ciertos eventos del ambiente son similares a las encontradas con otras unidades claramente establecidas (Branch, 1977).

De acuerdo con lo anterior, muchas respuestas pueden ocurrir altamente organizadas, pero no por ello ser unidades condicionables. Al usar secuencias de respuestas el reforzador se hace depender no sólo de una respuesta, sino de una serie de respuestas en un orden temporal específico (Mazur, 1998). Entonces, para demostrar que una secuencia de respuestas puede funcionar como una unidad, se debe probar que su probabilidad de ocurrencia depende de sus consecuencias.

Grayson y Wasserman (1979) dieron una clara demostración de que un reforzador puede controlar la probabilidad de un patrón un poco más complejo que una sola respuesta. Estos investigadores llevaron a cabo un experimento en el que reforzaron una secuencia de dos respuestas a dos palancas durante varias fases. Los resultados demostraron que una secuencia de dos respuestas es una propiedad diferenciable del comportamiento de los animales. En cada fase, la frecuencia de la secuencia reforzada aumentó y la de las otras secuencias disminuyó, lo que parece indicar que el reforzamiento fortaleció a la secuencia como una unidad, diferenciándola de las demás posibles secuencias. No obstante, un resultado interesante fue que al reforzar una secuencia heterogénea como izquierda (I) derecha (D), una secuencia que se presentó con mucha frecuencia fue DD. Se ha sugerido que esto ocurre porque al reforzar una secuencia como ID, el animal le asigna mayor crédito a la respuesta más contigua al reforzamiento, que en este caso sería la respuesta D (Catania, 1971). Esto mostró que aunque en efecto parecía que la secuencia como patrón general había sido reforzado, también hubo un efecto diferencial sobre las respuestas individuales que conformaban a la secuencia. Este experimento apoya la idea de que las secuencias pueden funcionar como unidades funcionales y demuestra cómo el reforzamiento afecta la distribución de varias secuencias.

En otra serie de estudios, se ha encontrado que aun cuando el procedimiento no re-

quiere la organización de la conducta en unidades, los animales desarrollan secuencias de respuestas dominantes y altamente estereotipadas. Por ejemplo, Schwartz (1980) utilizó un procedimiento en el que palomas tenían que picar dos teclas (I y D), cuatro veces cada una, en el orden que fuera, con el fin de mover una luz en una matriz de focos de 5x5 del extremo superior izquierdo al extremo inferior derecho. Cada respuesta en la tecla I movía la luz en la matriz hacia el foco de la derecha. Cada respuesta en la tecla D movía la luz hacía el foco de abajo. Aunque había 70 posibles secuencias que llevaban a la obtención del reforzador, todos los sujetos desarrollaron sólo una secuencia estereotipada que dominó su comportamiento. De manera que el reforzador parece haber formado una unidad que no sólo constaba de cada respuesta individual, sino de un patrón más general. Cuando se cambiaron las contingencias de reforzamiento, la secuencia dominante mantuvo su organización interna. Esto invita a inferir que la secuencia dominante se volvió una unidad integrada. No obstante, un detalle importante de este estudio es que cada respuesta fue señalada por un estímulo discriminativo, que fue el movimiento de las luces en la matriz de focos 5x5. Al quitar la matriz de luces, se encontró un aumento en la variabilidad y se perdió la organización interna de la secuencia dominante, lo que sugiere que la conducta estaba bajo el control de los estímulos discriminativos, no que se haya conformado en una unidad conductual.

No obstante, Reid *et al.* (2001) encontraron que aún en un procedimiento de operante libre¹, reforzando dos secuencias de dos respuestas con la misma probabilidad, los sujetos desarrollaron una secuencia dominante altamente organizada. Al analizar los tiempos entre respuestas (TER), Reid *et al.* (2001) encontraron que los TER entre secuencias fueron más largos que los TER entre respuestas dentro una secuencia. Lo cual sugiere que la conducta se organizó en “conjuntos” de respuestas diferenciables de acuerdo a las contingencias de reforzamiento, apoyando la idea de que una secuencia se puede volver una porción diferenciable del flujo conductual de los animales.

En la investigación sobre secuencias conductuales, se ha hecho una diferencia en cuanto

¹procedimiento en el que no existe ningún estímulo discriminativo que indique el inicio o final de un ensayo o de una respuesta

al método usado para estudiarlas. Existen dos maneras de entrenar las secuencias, en una las respuestas son señaladas por estímulos, en este caso se habla de cadenas. En el otro procedimiento cada respuesta de la secuencia no es señalada por estímulos, en este caso hablaremos de secuencias tal cual. En la investigación con cadenas, cada respuesta es señalada por un estímulo que guía la conducta de los animales y no se permite la ejecución completa de las respuestas incorrectas, por lo que es imposible monitorear todas las posibles secuencias de respuestas. Esto imposibilita especificar si un patrón de respuestas se ha diferenciado de otro. En cambio en la investigación con secuencias, se permite a los sujetos la ejecución de todas las posibles combinaciones de respuestas, aunque no sean la secuencia reforzada y no se guía la ejecución por medio de estímulos (Grayson y Wasserman, 1979). Los diferentes procedimientos han dado resultados diferentes en cuanto a la conformación de unidades conductuales.

Se ha encontrado que si se entrenan cadenas, al quitar los estímulos guía la conducta se desordena, como en el experimento de Schwartz (1980) mencionado previamente. Esto podría implicar que aunque se realice muy bien la secuencia de respuestas bajo el control de estímulos, no significa que ésta se haya vuelto una unidad funcional. Con secuencias, dado que no hay estímulos discriminativos explícitos, es más factible afirmar que el individuo ha integrado una serie de respuestas individuales en una unidad más grande. Sin embargo, en el caso de las secuencias, podría ser que responder en cada palanca tuviera por sí mismo propiedades discriminativas que influyeran en la probabilidad de las siguientes respuestas. En este trabajo nos centraremos en las secuencias de respuestas, en vez de en cadenas.

1.3.1. Estructura de la conducta

Como ya se mencionó la unidad de análisis en el AEC, la operante, está definida en términos del efecto que tiene sobre su ambiente, de manera que actividades que tengan el mismo efecto sobre el ambiente son consideradas instancias de la misma operante (Domjan, 2010). Esta definición es funcional, de manera que deja por completo de lado la estructura de la conducta, y dado que la respuesta medida suele ser una sola respuesta a una palanca, la estructura conductual se reduce a la topografía que ésta puede tener.

Por ejemplo, la rata podría apretar la palanca con su pata izquierda, con la derecha o con su cola, pero estas diferencias en la topografía física de las respuestas corresponderían a una misma operante, ya que son funcionalmente equivalentes, es decir, llevan a la misma consecuencia.

No obstante, el uso de secuencias de respuestas hace posible situaciones en las que la estructura de la conducta puede estudiarse directamente, manipulando el orden temporal de las respuestas de una secuencia. Uno de los primeros estudios en los que se encontró un efecto diferencial según la estructura de las secuencias de respuestas fue hecho por Polidora (1963). Polidora entrenó a un grupo de ratas a realizar una serie de secuencias de respuestas con diferente estructura (manipulando el largo de la secuencia y el número de alternaciones entre palancas). Cuando las respuestas de las ratas alcanzaron un estado estable² en las diferentes secuencias entrenadas, se les inyectó la droga JB 329 en tres diferentes dosis: 0.25, 0.5 y 1 mg/kg. Se encontró que el efecto de la droga dependía de la interacción entre la dosis y la estructura de la secuencia entrenada. El desempeño de secuencias “sencillas” (las más cortas y con menos alternaciones) fue menos afectado por las diferentes dosis administradas que el desempeño de secuencias “difíciles”.

A partir de estos resultados, Polidora propuso que el uso de secuencias de respuestas provee un medio para estudiar un continuo de conducta compleja, y a partir de las diferentes estructuras planteó el siguiente gradiente de complejidad: 1) secuencias más cortas son más sencillas que secuencias más largas, 2) secuencias que requieren la alternación entre palancas son más complejas y 3) las secuencias más complejas son las “secuencias de conteo” (e.g. AAB, porque implica el proceso cognitivo de contar dos respuestas en un mismo operando (A) y luego cambiar a otro operando (B)). No obstante, Polidora no llevó a cabo un estudio sistemático tomando en cuenta todas las posibles combinaciones de secuencias, sino que escogió al azar algunas de las posibles secuencias de dos, tres y cuatro respuestas, por lo que no pudo estudiar el panorama completo de todas las posibles estructuras.

²el criterio para establecer que la conducta había llegado a estado estable fue que durante tres sesiones consecutivas la tasa de respuestas y la proporción de secuencias correctas se mantuvieran constantes

Más recientemente, Neuringer, Kornell y Olufs (2001) en un experimento sobre variabilidad con secuencias de tres respuestas a tres palancas, en donde había 27 posibles secuencias, también sugirieron que parece haber un gradiente de dificultad en la ejecución de secuencias. Estos autores realizaron una serie de experimentos en los que en una primera fase reforzaron variabilidad, de tal manera que una secuencia de respuestas era reforzada sólo si había sido realizada con una frecuencia relativa menor o igual a .05 en los ensayos anteriores. En una segunda fase, ninguna conducta fue reforzada. Se observó que durante la primera fase se desarrolló una jerarquía de ejecución según la estructura de las secuencias, en donde secuencias de respuestas aparentemente más sencillas (como dar tres respuestas en una misma palanca) ocurrieron más frecuentemente que secuencias difíciles (como alternar las respuestas entre las tres palancas). Al pasar a una fase sin reforzamiento, aunque hubo un aumento en la variabilidad, la distribución de las posibles secuencias se mantuvo constante, es decir, las secuencias más probables durante reforzamiento, también lo fueron en la segunda fase donde ninguna conducta fue reforzada. Lo que se aprendió durante la fase de reforzamiento se vio reflejado en la segunda fase. Algo muy importante de este estudio fue que aunque las secuencias ya no estaban siendo reforzadas en la segunda fase, mantuvieron la integridad y organización interna que habían desarrollado durante el reforzamiento.

Los resultados de Neuringer *et al.* (2001) sugieren que el reforzamiento pudo haber estado actuando a nivel de las secuencias completas, en vez de actuar sobre las respuestas individuales que las componían. No obstante, Reid, Dixon y Gray (2008) mencionaron que esto implicaría que cada una de las 27 posibles secuencias de tres respuestas a las tres palancas se agruparon como una unidad conductual, lo cual es muy poco probable, ya que implicaría un proceso cognitivo bastante costoso. Es posible que algunas de las secuencias más frecuentes sí se conformaran en unidades conductuales, pero algunas de las secuencias fueron ejecutadas muy poco y casi no fueron reforzadas. Al re-analizar los datos, Reid *et al.* (2008) sugirieron que la distribución de secuencias observada en el experimento de Neuringer *et al.* (2001) se pudo haber debido a procesos a nivel de las respuestas individuales más que a nivel de las secuencias como unidades.

Una explicación alternativa a los resultados de Reid *et al.* (2008) es que ambos procesos estuvieran ocurriendo simultáneamente, es decir, que el reforzador tuviera un efecto sobre las secuencias como unidad, y sobre las respuestas individuales de cada secuencia. Bachá-Méndez, Reid y Mendoza-Soylovna (2007) llevaron a cabo un experimento en el que encontraron que una secuencia de dos respuestas a dos palancas puede volver a aparecer como una unidad después de un periodo en el que no fue reforzada y su frecuencia disminuyó a cero, lo cual indicaría que la secuencia se conformó como unidad. Sin embargo, también encontraron procesos a nivel de las respuestas individuales de las secuencias. El experimento que estos investigadores llevaron a cabo constó de 4 fases. En la primera fase se reforzó una secuencia heterogénea (ID o DI). En la segunda y tercera fases se reforzó una secuencia homogénea (II o DD) diferente para cada fase. En la última se reforzó otra secuencia heterogénea. En las fases 1 y 4 en las que una secuencia heterogénea estaba siendo reforzada, se observó que el error más frecuente fue repetir la respuesta más cercana a la entrega del reforzador³. Esto parece haber sido controlado por procesos a nivel de respuesta, ya que la posición ordinal de las respuestas en la secuencia afectó diferencialmente la producción de los errores.

Es posible dividir el procedimiento realizado por Bachá-Méndez *et al.* (2008) en tres fases. En una primera fase se reforzó una secuencia, después dicha secuencia dejó de recibir el reforzador al ser ejecutada y otra secuencia fue reforzada. Por ello las ratas dejaron de ejecutar la secuencia entrenada en la fase 1 y comenzaron a ejecutar la nueva secuencia ahora reforzada. Finalmente esa segunda secuencia también dejó de ser reforzada. Ante esta situación, si las ratas tuvieran que encontrar ahora qué nueva conducta ejecutar, resultaría exitoso que empezaran a variar su conducta buscando variantes que funcionaran. No obstante, si en el pasado (Fase 1) un comportamiento ya había sido exitoso, entonces resultaría adaptativo que la variabilidad conductual no fuera completamente azarosa, sino que se le diera preferencia a conductas que en el pasado ya han funcionado, que en este caso sería la secuencia reforzada en la primera fase. En efecto, esto fue lo que se encontró, ya que todos los sujetos mostraron resurgimiento íntegro de la secuencia

³e.g. si la secuencia ID estaba siendo reforzada, una secuencia errónea muy frecuente era DD

heterogénea entrenada en la primera fase, lo que indicó que resurgimiento es un proceso que operó a nivel de la secuencia como unidad. Estos resultados sugieren que los sujetos emitieron la secuencia reforzada como si fuera una unidad conductual, afectada por las contingencias como un todo. No obstante, fue posible observar tanto procesos a nivel de las respuestas individuales de las secuencias como a nivel de secuencia completa, operando simultáneamente dentro de una unidad conductual claramente establecida.

Dada la evidencia experimental, parecería que hay razón para tomar a una secuencia de respuestas como unidad de análisis. Al reforzar una sola respuesta, la variabilidad que se puede llegar a estudiar está limitada por los parámetros físicos de la respuesta. Asimismo, al usar unidades conductuales pequeñas podríamos estar limitando la generalidad de nuestras leyes conductuales (Shimp, 1976). En cambio, al utilizar secuencias de respuestas se puede cuantificar el conjunto de todas las posibles secuencias de respuestas y con ello estudiar la distribución y variabilidad de la conducta de manera sistemática. Las secuencias como unidad de análisis, permiten cuantificar cómo el reforzador afecta no sólo a la secuencia sujeta a reforzamiento, sino a todas las posibles secuencias y tomar el papel de la estructura de la conducta en consideración. Esto abre el intervalo de fenómenos que se pueden estudiar, ya que con las preparaciones clásicas, el hecho de medir una única conducta limita el estudio de patrones conductuales más complejos y generales.

Capítulo 2

Modelos lineales generalizados (GLM)

2.1. Introducción

En muchas disciplinas, tales como la biología, la psicología y la medicina, existe un gran interés por analizar diversos fenómenos a partir de un número reducido de variables. Los modelos estadísticos nos permiten encontrar patrones de asociación e interacción entre variables de manera ordenada, y nos proveen de una aproximación simplificada de la realidad.

Unos de los modelos más importantes y usados en estadística son sin duda los modelos lineales, en parte gracias a su fácil implementación e interpretación. Todo modelo lineal tiene dos componentes fundamentales: una variable de interés y con cierta distribución, y una ecuación que liga dicha variable con una combinación lineal de variables explicativas, con el propósito de explicar y predecir a y (Dobson, 2002).

La variable a modelar y , llamada usualmente variable dependiente o de respuesta, puede ser continua o categórica. Una variable categórica es aquella cuya escala de medición consiste en un conjunto de categorías. Este tipo de variables se divide en dos: las *variables ordinales*, las cuales tienen un orden natural; y las *variables nominales*, en las cuales el orden de las categorías no es relevante, podríamos cambiar el orden de las categorías y

los resultados no cambiarían (Agresti, 2007). En cuanto a las variables definidas como continuas, son aquellas que pueden tomar una infinidad de valores posibles, que pueden ser todos los reales, o un subconjunto de ellos (Lindsey, 1997).

Se llama lineal a un modelo cuando la variable de respuesta es una función lineal de los parámetros desconocidos, esta restricción no aplica a las variables explicativas, las cuales pueden ser funciones no lineales y esto no afecta la linealidad del modelo. Una de las razones de la popularidad de estos modelos es que la estimación de los parámetros desconocidos β_0, \dots, β_p se puede hacer de manera directa y sencilla a través del método de mínimos cuadrados, aunque también existen otras formas de estimación (Myers, Montgomery, Vining, y Robinson, 2010).

La forma general en que los modelos lineales relacionan a la variable dependiente y con otras variables es:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (2.1.1)$$

donde y es la variable a modelar, x_1, \dots, x_p es el conjunto de variables explicativas, β_0, \dots, β_p es el conjunto de parámetros desconocidos y ϵ es el término de error aleatorio.

El análisis de regresión lineal y el análisis de varianza (ANOVA por sus siglas en inglés) son los modelos lineales clásicos, los cuales han tenido un uso extenso y han sido de gran utilidad en diversas disciplinas. No obstante, tienen dos supuestos que representan una limitación importante: 1) asumen que la variable dependiente es función lineal de un conjunto de variables explicativas; y 2) asumen que la variable respuesta es continua, con distribución normal y varianza constante.

El problema es que existen muchas situaciones prácticas en las que estos supuestos no se cumplen, ya que es muy común que las variables de interés sean medidas en diversas escalas y que la relación entre la variable dependiente y las independientes no sea lineal. Por ejemplo, en muchas aplicaciones en las ciencias sociales y biológicas, la variable dependiente es categórica o consiste de conteos o incluso es continua pero no normal.

La relación entre modelos lineales con variables dependientes con diversas distribuciones y escalas, es clara a la luz de los **modelos lineales generalizados** (GLM por sus siglas en inglés). Esta familia de modelos provee de un marco unificador para muchas

técnicas estadísticas, de hecho, los ya mencionados modelos de regresión lineal y ANOVA, son casos específicos de ella. Los GLM fueron introducidos por Nelder y Wedderburn en 1972, y como su nombre lo indica, son generalizaciones del modelo clásico de regresión lineal (Dunteman y Ho, 2006). Nelder y Wedderburn unificaron a esta familia de modelos de regresión que incluye modelos con variables dependientes tanto normales como no normales, superando las limitaciones mencionadas de los modelos lineales clásicos.

Resumiendo, los GLM permiten usar métodos análogos a aquellos desarrollados para los modelos lineales, en las siguientes situaciones más generales:

- Cuando la variable respuesta tiene una distribución diferente de la normal y no necesariamente es una variable continua.
- Cuando la relación entre la variable respuesta y las variables explicativas no necesariamente es de la forma lineal.

Esta generalización es posible en parte, gracias a que muchas de las propiedades de la distribución normal son compartidas por una clase más amplia de distribuciones llamada *familia exponencial*. En los GLM todas las distribuciones de probabilidad de las variables respuesta pertenecen a esta familia, la cual incluye distribuciones como la normal, la binomial y la poisson, entre otras (Dobson, 2002).

Dentro de los GLM se encuentran la regresión logística, la regresión poisson, los modelos log-lineal, etc., algunos de los cuales serán descritos en secciones posteriores. Estos modelos permitieron cambiar la manera de pensar de los analistas de datos, liberándolos de tener que buscar transformaciones de la variable dependiente que simultáneamente logaran la linealidad de los predictores y la normalidad de la distribución de la variable dependiente (Myers et al., 2010).

2.2. Características generales de los GLM

Un modelo lineal generalizado está formado de tres componentes principales (Agresti, 2007):

1. *Componente aleatorio*: se refiere a las n observaciones independientes y_1, \dots, y_n de la variable respuesta, las cuales se asume comparten la misma distribución perteneciente a la familia exponencial.
2. *Componente sistemático*: se refiere al conjunto de variables explicativas x_1, \dots, x_p , las cuales entran al modelo como una combinación lineal de los parámetros β_1, \dots, β_p de la siguiente manera:

$$\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

3. Una *función liga* g que relaciona el componente aleatorio con el sistemático.

Comenzando por el componente aleatorio, si consideramos una sola variable aleatoria y con parámetro θ , su distribución pertenece a la familia exponencial si puede ser expresada de la siguiente manera (Dobson, 2002):

$$f(y; \theta) = s(y)t(\theta)e^{a(y)b(\theta)} \quad (2.2.1)$$

donde a, b, s y t son funciones conocidas.

La ecuación 2.2.1 suele ser escrita en su forma exponencial:

$$f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)] \quad (2.2.2)$$

donde $s(y) = \exp(d(y))$ y $t(\theta) = \exp(c(\theta))$.

La forma general de los modelos lineales generalizados que liga a la variable respuesta y con el componente sistemático es:

$$g[E(y_i)] = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (2.2.3)$$

en donde $E(y_i) = \mu_i$ es el valor esperado o esperanza de la variable respuesta, g es la función liga, la cual es una función monótona y diferenciable, como la identidad o el logaritmo. Las x_j con $j = 1, \dots, p$, son las variables explicativas y las β_j con $j = 1, \dots, p$, son los p parámetros a estimar, los cuales nos hablan de la relación entre la variable dependiente y las variables explicativas. Resulta relevante mencionar que la función liga

es una transformación de la esperanza o media poblacional, no de los datos. A diferencia de las transformaciones clásicas, la función liga se aprovecha de la distribución natural de la variable respuesta (Myers et al., 2010).

La parte $\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ es llamada componente lineal y nos proporciona una función del valor esperado de cada y_i de acuerdo al modelo. Este mapeo de los valores esperados de la variable de respuesta y al componente lineal es parte integral de los modelos lineales generalizados (Dobson, 2002).

Como ejemplo, podemos observar que la regresión lineal es un caso particular de estos modelos, ya que en primera instancia, la distribución normal forma parte de la familia exponencial. En este caso la función liga es la identidad $g(E[y]) = E[y]$, y es posible modelar de manera directa la esperanza de y con una función lineal, sin ninguna transformación del valor esperado. Entonces, el modelo lineal queda especificado como:

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (2.2.4)$$

que es la forma clásica de los modelos de regresión para respuestas continuas.

Para otros modelos el valor esperado no puede ser escrito directamente como una función lineal de los parámetros β_j $j = 1, \dots, p$, pero sí mediante alguna función g no lineal de la esperanza. La función liga g es muy importante, ya que transforma la esperanza de la variable dependiente de manera que el valor transformado, $g(E[y])$, sí es una función lineal de los parámetros de regresión (Dunteman y Ho, 2006; Agresti, 2007).

Es posible apreciar que los modelos lineales generalizados nos permiten una mayor flexibilidad en cuanto a la variable dependiente y a su relación con el componente lineal. No obstante, están restringidos a una estructura lineal para las variables explicativas y a una sola respuesta dependiente (Lindsey, 1997).

2.3. Estimación de los parámetros

La estimación de los parámetros β_j $j = 1, \dots, p$, de los modelos lineales generalizados está fundamentada en métodos de máxima verosimilitud. La estimación por máxima verosimilitud se basa en el concepto de que los parámetros estimados deberían de ser aquellos

para los cuales la probabilidad de los datos observados tome el mayor valor posible. Es decir, condicionando a la muestra obtenida, la estimación máximo verosímil encuentra el valor de los parámetros que más probablemente generaron las observaciones de la muestra (Agresti, 2007).

Para estimar los parámetros de regresión y sus errores estándar es necesario especificar la distribución de probabilidad adecuada, ya que a partir de ella se especifica la función de verosimilitud. Esta función se refiere a la función de probabilidad de los datos observados, especificada en términos de los parámetros (Agresti, 2007). Los estimadores máximo verosímiles se encuentran maximizando esta función de verosimilitud. Es por ello que la especificación correcta de la función de probabilidad de los datos es de suma importancia (Dobson, 2002).

Los estimadores obtenidos por métodos de máxima verosimilitud son muy usados en estadística, ya que son estimadores muy precisos, en términos de su error estándar. Asimismo, con muestras grandes estos estimadores tienden a distribuirse aproximadamente normal, lo que facilita su evaluación (Agresti, 2007).

Los estimadores máximo verosímiles, con excepción del modelo de regresión normal, no se pueden obtener mediante métodos analíticos estándar, debido a la complejidad del sistema de ecuaciones resultante. Es por ello que se usan métodos numéricos iterativos (Dunteman y Ho, 2006).

La mayoría de los paquetes estadísticos incluyen procedimientos para ajustar modelos lineales generalizados que usan algoritmos que comienzan con una aproximación inicial de las β_j , y a partir de ahí iteran hasta que se alcanza una convergencia adecuada. El algoritmo usado para realizar la estimación es el de mínimos cuadrados ponderados, el cual sólo funciona en la familia exponencial (Lindsey, 1997).

Máxima verosimilitud es un método que da muy buenos resultados bajo los supuestos de que los datos en efecto se distribuyen como la distribución asumida y que existe independencia entre las observaciones. No obstante, es común que aunque las respuestas sí sean independientes, no se distribuyan como alguna de las distribuciones de la familia exponencial. También suele suceder que las observaciones estén correlacionadas. En estos

casos una opción es usar métodos de estimación basados en quasi-verosimilitud (Myers et al., 2010).

2.4. Evaluación de los modelos

2.4.1. Devianza

Una vez formulado un modelo, es necesario evaluar su ajuste a los datos observados. En general, se buscan aquellos modelos que describan bien los datos y que sean lo más sencillo posible. De acuerdo al principio de parsimonia, no se deberían agregar más variables que aquellas necesarias para explicar el efecto observado. Es así que un modelo más sencillo y parsimonioso es preferible a uno más complicado que deja poca variabilidad no explicada (Dobson, 2002). Asimismo, se busca describir y predecir el fenómeno general, no las variaciones aleatorias particulares de cada muestra, ya que al sobreajustar, el modelo pierde la capacidad de generalización.

Una manera de evaluar al modelo es compararlo con el modelo más complejo posible, el cual contenga el número máximo de parámetros estimables. Dicho modelo es llamado *modelo completo o saturado* y tiene tantos parámetros como observaciones muestreadas y por lo tanto, ajusta perfectamente a los datos (Dunteman y Ho, 2006).

En todo caso, la función de verosimilitud del modelo saturado $L(b_{max}; y)$, siempre será mayor a la de cualquier otro modelo que nos interese evaluar, ya que provee la descripción más completa de los datos (Dobson, 2002). Si llamamos a $L(b; y)$ a la máxima verosimilitud del modelo de interés, entonces la razón de estas verosimilitudes:

$$\lambda = \frac{L(b, y)}{L(b_{max}; y)}$$

provee una manera de evaluar la bondad de ajuste del modelo de interés. En la práctica, menos dos veces la diferencia de las log-verosimilitudes del modelo saturado y el modelo a evaluar, se usa como indicador de qué tan bien el modelo se ajusta a los datos. Se usa esta cantidad debido a que tiene la propiedad de distribuirse aproximadamente como una ji-cuadrada, con grados de libertad igual a la diferencia en el número de parámetros

entre los dos modelos (Agresti, 2007). Este estadístico es llamado *devianza*, el cual queda expresado de la siguiente manera:

$$D = -2[l(b; y) - l(b_{max}; y)] \sim \chi_{n-p}^2$$

con n =tamaño de muestra, p =número de parámetros del modelo.

La devianza es un estadístico de prueba para la hipótesis de que todos los parámetros que están en el modelo saturado, pero no en el modelo de interés, son iguales a 0. En este caso buscamos que la devianza del modelo de interés sea pequeña, ya que queremos que la verosimilitud de nuestro modelo, el cual contiene una cantidad reducida de parámetros, se parezca lo más posible a la verosimilitud del modelo saturado. Es decir, buscamos que los valores de p no sean significativos.

Si la devianza del modelo seleccionado es grande y estadísticamente significativa al comparar con una ji cuadrada con los grados de libertad adecuados, esto podría indicar sobredispersión. Este fenómeno se refiere a que la varianza estimada es mayor de lo que se esperaría para la distribución de probabilidad usada en el modelo (Dunteman y Ho, 2006). Sobredispersión no es un problema en la regresión lineal clásica, ya que el parámetro de varianza está separado del valor esperado, pero en el caso de muchas otras distribuciones de la familia exponencial, la varianza depende de la media. En tal caso, se puede re-especificar la función de probabilidad para los datos (Agresti, 2007).

La devianza también se puede usar para comparar la bondad de ajuste de dos modelos diferentes. Estos deben estar anidados o ser jerárquicos, es decir, deben tener la misma distribución de probabilidad y todas las variables contenidas en el predictor lineal del modelo con menor número de parámetros M_0 deben estar en el modelo más complejo M_1 (Dobson, 2002).

Para comparar dos modelos, primero se calcula la devianza de cada uno de los modelos. Evidentemente la devianza del modelo reducido siempre será mayor a la devianza del modelo con más parámetros; no obstante, la pregunta es si la diferencia entre ellas es suficientemente grande para justificar la inclusión de parámetros extras al modelo. Dado que la devianza de cada modelo se distribuye como una ji-cuadrada, la diferencia de las devianzas también se distribuirá como una ji cuadrada, con grados de libertad igual a

la diferencia de parámetros entre los dos modelos. Si esta diferencia es grande y cae en la zona de rechazo, entonces hay evidencia de que el modelo más complejo M_1 provee una descripción significativamente mejor de los datos (Dobson, 2002; Dunteman y Ho, 2006). En conclusión, la diferencia de devianzas es una buena manera de hacer pruebas de hipótesis para modelos anidados.

Una alternativa para cuando los modelos no están anidados es el uso del Criterio de Información Akaike (AIC por sus siglas en inglés), el cual representa una penalización en la log-verosimilitud de acuerdo a la complejidad del modelo. AIC propone tomar en cuenta el ajuste general del modelo y los grados de libertad. El AIC está definido como

$$AIC = 2k - 2\log L$$

donde k es el número de parámetros. Cuando se están comparando modelos, un AIC más pequeño indica un modelo preferible sobre un AIC mayor (VonEye y Mun, 2013).

Otra opción es la prueba de Wald, la cual se basa en las propiedades de normalidad asintóticas de los estimadores máximo verosímiles. Este estadístico parecido a la prueba t es útil para probar los coeficientes de manera individual. Usando los errores estándar de los coeficientes es posible hacer el contraste de hipótesis:

$$H_0 : \beta_j = 0 \text{ vs } H_1 : \beta_j \neq 0$$

mediante el estadístico:

$$Z_0 = \frac{\hat{\beta}_j}{\hat{se}(\hat{\beta}_j)}$$

La distribución para este estadístico bajo la hipótesis nula es la normal estándar. Algunos paquetes estadísticos elevan al cuadrado a Z_0 y lo comparan con una distribución ji-cuadrada con un grado de libertad. No obstante, esta prueba también se puede hacer con el método descrito previamente usando la diferencia de devianzas (Myers et al., 2010).

A partir del estadístico de Wald descrito en el párrafo anterior, la construcción de los intervalos de confianza para las β es de forma directa si consideramos su distribución asintóticamente normal, quedando de la siguiente manera:

$$b_j - Z_{\alpha/2}\hat{se}(b_j) \leq \beta_j \leq b_j + Z_{\alpha/2}\hat{se}(b_j) \quad (2.4.1)$$

Si el intervalo construido incluye al 0, entonces no rechazamos la hipótesis nula de que el parámetro es igual a 0, es decir, el efecto de la variable asociada a dicho parámetro no es estadísticamente significativo.

2.4.2. Análisis de residuales

Otra manera de evaluar los modelos es a través de los residuales. La diferencia entre el valor observado y y el ajustado por el modelo $\hat{\mu}_i$ es llamado residual. Los residuales son una parte importante de los modelos, ya que con ellos es posible revisar que algunos supuestos hechos al plantear un modelo en efecto se cumplen (Dobson, 2002). Asimismo, nos pueden servir para detectar problemas en el modelo, por ejemplo, con ellos podemos detectar la falta de un término cuadrático en el modelo, observaciones donde el ajuste fue muy malo, la falta de independencia entre las observaciones o si la liga elegida fue adecuada (Dunteman y Ho, 2006).

El residual más sencillo para evaluar la bondad del ajuste de modelos lineales generalizados es el residual de Pearson (Dunteman y Ho, 2006), el cual es:

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{var}(\hat{\mu}_i)}}$$

Los residuales de Pearson fluctúan alrededor de cero y aunque su distribución no es normal, una gráfica cuantil-cuantil comparando con la normal nos puede indicar si los supuestos del modelo se cumplen. En la gráfica los puntos deberían de caer cercanos a una línea recta de 45 grados como evidencia de normalidad (Dobson, 2002).

Otro residual usado para analizar el ajuste del modelo es el residual de la devianza. La devianza está compuesta de la suma de las devianzas al cuadrado de cada observación. Con estos residuales, podemos analizar la contribución de cada observación particular a la devianza, mientras mayor sea, peor ajuste tiene el modelo en esa observación. Dichos residuales son de la siguiente forma:

$$r_i = \text{signo}(y_i - \hat{\mu}_i) \sqrt{2(l_i(y_i) - l_i(\mu_i))}$$

En donde el término de signo indica si el residual es positivo o negativo, $l(y_i)$ es el valor de la log-verosimilitud cuando el valor esperado es el valor observado del individuo y $l(\hat{y}_i)$

es la log-verosimilitud cuando el valor esperado es substituido por el valor generado por el modelo. El término debajo de la raíz cuadrada es la contribución de la i -ésima observación a la devianza (Dunteman y Ho, 2006). La información recabada de los residuales de la devianza es similar a la de los residuales de Pearson.

Finalmente, un análisis gráfico de los residuales es importante. Graficar los residuales contra los valores ajustados por el modelo nos provee de mucha información. En esta gráfica buscamos que los residuales tengan un patrón aleatorio, es decir, se espera que no haya patrones sistemáticos que puedan indicar la falta de una variable explicativa. Asimismo, en esta gráfica es posible detectar cambios en la varianza, por ejemplo, un incremento en la dispersión de los residuales al final del intervalo de valores ajustados podría indicar sobredispersión (Faraway, 2006). Asimismo, se debería realizar una gráfica de los residuales según el orden en el que los valores de y_i fueron obtenidos en el tiempo, espacio o cualquier otra variable que podría causar alguna dependencia entre las observaciones. Si los residuales son independientes, entonces los puntos deberían de fluctuar aleatoriamente, sin ningún patrón sistemático (Dobson, 2002).

Ya con los modelos lineales generalizados definidos y descritos en términos generales, es posible pasar a describir modelos específicos de esta familia que son de nuestro interés.

Capítulo 3

Modelos para conteos

En muchos estudios se miden variables nominales u ordinales con varias categorías, en los cuales se registra el número de observaciones, conteos o frecuencias de cada una de ellas. Como los conteos, las variables Poisson pueden tomar cualquier número entero no-negativo, lo que las hace adecuadas para modelar este tipo de datos. Si Y es el número de ocurrencias del evento, entonces su distribución de probabilidad puede escribirse como (Dobson, 2002):

$$f(y) = \frac{\mu^y e^{-\mu}}{y!} \quad (3.0.1)$$

donde μ es el promedio de ocurrencias y es tanto su esperanza como su varianza, es decir, si Y_i variable aleatoria tiene una distribución Poisson, entonces $E(Y) = Var(Y) = \mu$. Esto nos dice que si los conteos son más grandes, en promedio también tenderán a ser más variables.

Podemos reescribir la ecuación 3.0.1 en su forma exponencial:

$$f(y, \mu) = \exp(y \log \mu - \mu - \log y!) \quad (3.0.2)$$

la cual tiene la misma forma que la ecuación 2.2.2, por lo que es posible observar que la distribución Poisson es parte de la familia exponencial. Asimismo, los supuestos de independencia y misma distribución para todas las Y_i de los GLM se mantienen para estos modelos. Generalmente, si no se cumplen estos supuestos se obtendrán estimaciones sesgadas (VonEye y Mun, 2013).

Resulta relevante comentar que de acuerdo al diseño del estudio, el modelo de probabilidad asumido para los datos será diferente. Por ejemplo, si se asume que todas las frecuencias Y_i , $i = 1, \dots, N$, de una tabla de contingencia de N celdas, son variables aleatorias independientes y no se pone ninguna restricción a las frecuencias de las celdas, entonces la distribución de probabilidad conjunta sería:

$$f(y; \mu) = \prod_{i=1}^N \mu_i^{y_i} e^{-\mu_i} / y_i!$$

Sin embargo, si la única restricción que se tiene es que la suma de las Y_i sea n , entonces se usaría la distribución multinomial. En otros casos, por ejemplo si el número de observaciones por categoría de una variable se fijó de antemano, la distribución multinomial producto sería más adecuada. No obstante, todos estos modelos de distribución son equivalentes en verosimilitud a la Poisson, de manera que para cualquiera de las distribuciones mencionadas, los estimadores de máxima verosimilitud son los mismos. Por lo tanto, para propósitos de estimación, la distribución Poisson siempre puede ser asumida como la distribución de probabilidad de la muestra (Dobson, 2002).

Es así que es posible modelar el efecto de las variables explicativas sobre los conteos Y_i a través del parámetro μ de la distribución Poisson, lo cual se puede dar en dos situaciones diferentes. En la primera, el número de eventos se relacionan con diferentes cantidades de exposición, las cuales deben ser tomadas en cuenta para modelar la tasa de ocurrencia del evento. Por ejemplo, no es lo mismo 3 accidentes de avión en 100 vuelos, que 3 accidentes de avión en 10 vuelos. La otra situación es aquella en la que la exposición es igual para todos los datos recolectados, y por ello no resulta relevante incluirla en el modelo. Si todas las variables son categóricas, se llaman *modelos log-lineales*, cuyo nombre hace referencia a la función liga usada (Dobson, 2002). En resumen, un modelo log-lineal es un GLM que independientemente del diseño del estudio asume la distribución Poisson para las frecuencias Y_i , y es el modelo en el que nos centraremos en esta sección.

3.1. Modelos log-lineales

El término “modelo log-lineal” apareció por primera vez en 1969 y desde entonces es uno de los métodos más populares para el análisis de frecuencias. Específicamente, los modelos log-lineal se utilizan para analizar tablas de contingencia multivariadas y nos ayudan a describir patrones de asociación entre variables categóricas (VonEye y Mun, 2013).

Primero definamos las tablas de contingencia. Considerando dos variables categóricas, X y Y , cada una con I y J categorías correspondientemente, se le llama *tabla de contingencia* a la tabla con I renglones y J columnas, la cual tiene $I \times J$ celdas que representan todas las posibles combinaciones de resultados (Agresti, 2007). En la siguiente tabla vemos la estructura general de estas tablas:

Tabla 1: Ejemplo de tabla de contingencia de $I \times J$.

		variable Y		
variable X	y_1	\cdots	y_J	
x_1	n_{11}	\cdots	n_{1J}	$n_{1.}$
\vdots	\vdots	\ddots	\vdots	\vdots
x_I	n_{I1}	\cdots	n_{IJ}	$n_{I.}$
	$n_{.1}$	\cdots	$n_{.J}$	n

n es el total de observaciones, las n_{ij} , $i = 1, \dots, I$ $j = 1, \dots, J$, son las frecuencias de cada celda, la última columna y la última fila representan los totales marginales por renglón ($n_{i.}$) y por columna ($n_{.j}$), respectivamente.

En un modelo log-lineal la variable a modelar son las frecuencias de cada celda de una tabla de contingencia, por lo que los podemos usar para (VonEye y Mun, 2013):

- Examinar la distribución conjunta de varias variables
- Evaluar la dependencia o independencia entre dos o más variables

- Estudiar los patrones de asociación entre dos o más variables

El modelo más sencillo que se puede ajustar a una tabla es el *modelo nulo*, el cual propone que no existen efectos. Si esta proposición es correcta, la probabilidad esperada para cada celda es $\hat{\pi}_{ij} = 1/(I \times J)$, en donde todas las celdas tienen la misma probabilidad sin contemplar su pertenencia a las diversas categorías (Dobson, 2002).

Un modelo un poco más complejo que el nulo, en el que se incluyen los efectos de cada variable por separado, es el de independencia. Si consideramos una tabla de contingencia $I \times J$, las variables son estadísticamente independientes si la probabilidad conjunta de cada celda π_{ij} está determinada por las probabilidades marginales de la siguiente manera

$$\pi_{ij} = \pi_{i.}\pi_{.j} \quad (3.1.1)$$

donde $\pi_{i.} = n_{i.}/n$ se refiere a la probabilidad marginal de pertenecer al renglón i $i = 1, \dots, I$, y $\pi_{.j} = n_{.j}/n$ a la probabilidad marginal de pertenecer a la columna j , $j = 1, \dots, J$. Los modelos log-lineal usan como frecuencia esperada de las celdas $\mu_{ij} = n\pi_{ij}$, de manera que si sustituimos la ecuación 3.1.1 en esta definición de frecuencia esperada, nos queda que (Agresti, 2007):

$$\mu_{ij} = n\pi_{i.}\pi_{.j} \quad (3.1.2)$$

La función liga natural de la distribución Poisson es el logaritmo. Si le sacamos logaritmo a la ecuación 3.1.2 nos queda el siguiente modelo log-lineal aditivo

$$\log(\mu_{ij}) = \log n + \log \pi_{i.} + \log \pi_{.j} \quad (3.1.3)$$

Que podemos reescribir como:

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y \quad (3.1.4)$$

en donde λ_i^X es el efecto de la variable renglón y λ_j^Y es el efecto de la variable columna. En este primer modelo sólo se incluyeron los efectos principales y como ya se mencionó representa el modelo log-lineal de independencia. Si consideramos una tabla de 2×2 , es decir, una tabla con 4 celdas, el modelo de independencia tendría los siguientes 5 parámetros: $\lambda, \lambda_1^X, \lambda_2^X, \lambda_1^Y$ y λ_2^Y . Es fácil observar que el modelo estaría sobreparametrizado, ya

que hay más parámetros que probabilidades (VonEye y Mun, 2013). En tablas $I \times J$ hay $(I-1)(J-1)$ parámetros no redundantes, por lo que generalmente se especifica uno de los parámetros de cada variable como cero y es el que suele quedar como referencia (Agresti, 2007). Más adelante, se explicará la interpretación de los parámetros.

A partir de la definición de $\mu_{ij} = n\pi_i \cdot \pi_j$ obtenemos que la frecuencia esperada para el modelo (3.1.4) es $\hat{\mu}_{ij} = n \left(\frac{n_i}{n}\right) \left(\frac{n_j}{n}\right) = n_i \cdot n_j / n$. Estas frecuencias esperadas son iguales a las de la muy conocida prueba estadística ji-cuadrada, X^2 , la cual se utiliza para probar independencia entre variables. Si se rechaza el modelo de independencia, entonces se concluye que las dos variables están asociadas, de manera análoga a como se hace en la prueba X^2 . En contraste, si el modelo ajusta bien los datos entonces se asume que las variables son independientes (VonEye y Mun, 2013). Vemos que para tablas $I \times J$ la prueba X^2 es análoga a ajustar el modelo log-lineal de independencia.

Para variables estadísticamente dependientes tendríamos que incluir un término que represente dicha relación. El modelo saturado para tablas de 2×2 que satisface esta condición es

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \quad (3.1.5)$$

en donde el parámetro λ_{ij}^{XY} refleja la interacción entre las variables X y Y , de manera que las frecuencias esperadas de una variable dependen del nivel de la otra variable. El modelo saturado tiene 0 grados de libertad, ya que todos los posibles efectos están presentes en él.

Los modelos log-lineal son jerárquicos, lo cual significa que si una interacción o componente de orden mayor está en el modelo, todos los términos de orden menor relacionados con él, también deben estar presentes. Por ejemplo, si la interacción λ_{ij}^{XY} está en el modelo, entonces forzosamente λ_i^X y λ_j^Y también deben estar en él. Y lo mismo aplica para interacciones de orden mayor (Agresti, 2007).

Para tablas de dos dimensiones sólo podemos probar independencia o dependencia entre las variables, prueba que como ya vimos, podemos hacer con la X^2 . Sin embargo, cuando existen más variables, el paso de la independencia a la dependencia es más gradual y es posible, mediante los modelos log-lineal, probar diferentes patrones de asociación entre

las variables. A continuación se comentará para el caso en el que tenemos 3 variables.

Consideremos una tabla de contingencia de tres variables X, Y y Z . El modelo sólo con efectos principales se llama *modelo de independencia mutua*, y trata a cada par de variables como independientes. Sin embargo, este modelo es muy sencillo y es común que se necesiten meter algunos términos de asociación. Si tomamos el siguiente modelo

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} \quad (3.1.6)$$

vemos que permite las asociaciones XZ y YZ , pero no contiene la asociación entre X y Y , de manera que representa la *independencia condicional* entre X y Y , controlando por Z . Este modelo nos permite ver si existe una tercera variable Z que al ser incluida en el modelo hace que desaparezca la relación entre X y Y .

Otra posibilidad es que el modelo permita las tres interacciones dobles

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ij}^{XY} \quad (3.1.7)$$

A este modelo se le llama de *asociación homogénea*, en donde la relación entre cualesquiera dos variables del modelo es la misma en cada nivel de la tercera variable. Finalmente, el modelo que incluye la triple interacción es el modelo saturado y tiene un ajuste perfecto (Agresti, 2007). Vemos que de acuerdo a los parámetros de interacción incluidos en los modelos, los patrones de asociación difieren. Estos modelos son extendibles a tablas con más de tres variables, pero la interpretación se va complicando.

Asimismo, mediante modelos log-lineal se pueden probar modelos de simetría. Uno de los más famosos es el modelo de simetría axial, el cual propone equiprobabilidad para las celdas que son simétricas con respecto a la diagonal, es decir, las probabilidades de un lado de la diagonal son un espejo del otro lado, $p_{ij} = p_{ji}$ con $i \neq j$ (Agresti, 2007). La figura 2 muestra el concepto de simetría con respecto a la diagonal.

Para probar este modelo de simetría se incluye una variable explicativa artificial en el modelo, en donde se genera un vector de contraste. Por ejemplo, para el modelo de simetría axial se puede asignar un 1 a las celdas c_{ij} y a las c_{ji} y ceros en la diagonal.

Otro modelo que se puede poner a prueba es el de simetría puntual, en este caso para la celda de referencia c_{ij} , se propone que $c_{i+k,j+l} = c_{i-k,j-l}$. Esto implica que las celdas que

m_{11}	m_{12}	m_{13}
m_{21}	m_{22}	m_{23}
m_{31}	m_{32}	m_{33}

Figura 2: **Simetría axial**. Tomada de VonEye y Mun (2013).

estén a k pasos de la celda de referencia son equiprobables. Al igual que con el ejemplo anterior, se agrega un vector de contraste en el que a las celdas que correspondan a la hipótesis de simetría puntual se les asigna un 1, y al resto un 0 (VonEye y Mun, 2013). En la tabla 2 se ve un ejemplo de simetría puntual para una tabla de contingencia de 5×5 . Las celdas con el mismo símbolo son simétricas con respecto a la celda de referencia. Es posible apreciar que podemos incluir muchos tipos de variables artificiales al modelo para poner a prueba diferentes hipótesis en las tablas de contingencia.

Tabla 2: Ejemplo de simetría puntual con respecto a c_{33} . Tomada de VonEye y Mun (2013).

m	e	f	g	h
l	d	a	b	i
k	c	c_{33}	c	k
i	b	a	d	l
h	g	f	e	m

3.1.1. Interpretación de los parámetros

La interpretación de los parámetros es más sencilla cuando vemos una respuesta en función de otras. La diferencia entre parámetros se puede interpretar como el logaritmo de momios (si sólo estamos analizando una variable) o como el logaritmo de la razón de

momios (cuando analizamos la interacción de dos variables) (Agresti, 2007). Definamos lo que es un momio y una razón de momios. Si un evento ocurre con probabilidad p , entonces el momio para ese evento es:

$$\Omega = \frac{p}{1-p}$$

Los momios se pueden comparar unos con otros. Supongamos que tenemos Ω_1 y Ω_2 , una manera de compararlos es a través de una razón:

$$\theta = \frac{\Omega_1}{\Omega_2}, \quad \Omega_2 > 0$$

en donde θ es la razón de momios, la cual nos indica qué tanto es más probable el primer evento que el segundo. Los momios y sus razones se pueden interpretar en términos de qué tanto más posible es un evento en un grupo en comparación con otro (VonEye y Mun, 2013).

Supongamos una tabla de $I \times J$, en donde la variable Y tiene dos categorías, $J = 2$. Bajo el modelo de independencia, el momio para las categorías de la variable Y sería:

$$\log \left(\frac{\mu_{i1}}{\mu_{i2}} \right) = \log \mu_{i1} - \log \mu_{i2} = (\lambda + \lambda_i^X + \lambda_1^Y) - (\lambda + \lambda_i^X + \lambda_2^Y) = \lambda_1^Y - \lambda_2^Y$$

Esto lo interpretamos como que el logaritmo del momio de las categorías de Y es igual a $\lambda_1^Y - \lambda_2^Y$, de manera que si exponenciamos esta diferencia obtenemos $\Omega_1 = \exp(\lambda_1^Y - \lambda_2^Y)$. Con Ω_1 podemos saber qué tanto más posible es la categoría 1 de Y con respecto a la categoría 2. En este ejemplo consideramos que Y es independiente de X , de manera que esperaríamos que los momios de Y fueran iguales en todos los niveles de X . Generalmente, una categoría es asignada como referencia y se le da el valor de 0, de manera que los parámetros estimados de las demás categorías resultan en el logaritmo de los momios con respecto a dicha categoría (Agresti, 2007).

En el caso de los parámetros de interacción, existe una relación directa entre ellos y el logaritmo de la razón de momios. Por ejemplo, en una tabla 2×2 el logaritmo de la razón de momios sería:

$$\log \theta = \log \left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}} \right) = \log \mu_{11} + \log \mu_{22} - \log \mu_{12} - \log \mu_{21} = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}$$

Si el modelo log-lineal de independencia se mantiene, entonces $\theta = 1$ y por lo tanto $\log \theta = 0$, es decir, el término de interacción no debe ir en el modelo.

3.1.2. Evaluación del modelo

El ajuste del modelo se puede evaluar con la devianza, como fue mencionado en el capítulo anterior. Otra opción es el estadístico ji cuadrada, X^2 , el cual evalúa la bondad de ajuste del modelo comparando el valor ajustado para cada celda con el conteo observado

$$X^2 = \frac{\sum (n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}$$

donde $\hat{\mu}_{ij}$ es el valor predicho por el modelo para una tabla de $I \times J$. El valor de este estadístico se compara con una χ^2 con grados de libertad igual al número de celdas menos el número de parámetros del modelo. Los modelos que no ajusten bien a los datos tendrán valores de X^2 grandes y por lo tanto valores de p pequeños (Agresti, 2007). No obstante, este estadístico es muy sensible al tamaño de muestra, aun cuando la relación en la tabla sea la misma, si el tamaño de la muestra aumenta, el valor de la X^2 también aumentará. Por otro lado, aunque el valor de la devianza también depende del tamaño de la muestra, se ve menos afectado por él. Esto resulta en que con tamaños de muestra muy grandes, todos los modelos que se ajusten serán rechazados (VonEye y Mun, 2013). Por lo tanto, hay que realizar la evaluación del modelo a partir de diversas evidencias.

Como se explicó en el capítulo anterior, es posible hacer los intervalos de confianza de los parámetros usando su error estándar. En este caso es posible exponenciar los límites del intervalo para generar el intervalo de confianza para los momios y las razones de momios.

También es posible ajustar el modelo mínimo o nulo $\log(\mu_{ij}) = \lambda$ y observar la diferencia entre las devianzas, para ver qué tanto se redujo la devianza al incluir los parámetros e interacciones escogidas. Asimismo, se puede calcular la pseudo- R^2 la cual es

$$pseudo - R^2 = 1 - \frac{\log l(b_m)}{\log l(b_{min})}$$

en donde $l(b_{min})$ es la máxima verosimilitud del modelo nulo y $l(b_m)$ es la máxima verosimilitud del modelo de interés. Esta medida indica qué tanto el modelo de interés mejoró el ajuste a los datos con respecto al modelo nulo. Valores mayores de la $pseudo - R^2$ sugieren que el modelo ajusta bien a los datos (Dobson, 2002; VonEye y Mun, 2013).

Asimismo, una mirada a los residuales de Pearson y a los de la devianza también puede dar una idea sobre qué tan adecuado es el modelo celda a celda. Ya que es posible que el modelo ajuste bien en general por excepción de una celda. Cuando un modelo tiene muchas celdas es posible que algunos residuales sean más grandes de lo esperado debido a procesos aleatorios (Dobson, 2002).

3.1.3. Sobredispersión

En estos modelos no es raro que ocurra sobredispersión, ya que los conteos tienden a variar más de lo que se esperaría si la distribución de la respuesta fuera verdaderamente Poisson, en donde $var(Y_i) = E(Y_i)$. La distribución binomial negativa puede resultar una distribución alternativa adecuada, ya que también involucra a los enteros no-negativos y tiene un parámetro adicional D que permite que la varianza sea mayor que la esperanza de la siguiente manera (Agresti, 2007):

$$E(Y) = \mu, \quad Var(Y) = \mu + D\mu^2$$

$D > 1$ es llamado parámetro de dispersión. Mayor sobredispersión en la Poisson resulta en un valor D mayor. Mientras D se acerque más a 0, entonces $Var(Y)$ será más parecida a la $E(Y)$, y la distribución binomial negativa convergerá hacia la Poisson. La sobredispersión se puede deber a una falta de independencia entre las observaciones, en tal caso, una opción viable son los modelos multinivel, los cuales se discutirán en el siguiente capítulo.

Capítulo 4

Modelos multinivel

Los modelos lineales se han extendido a situaciones en donde las respuestas no son independientes. Esto puede ocurrir en el caso de datos longitudinales, donde una variable es medida varias veces en un mismo sujeto, o cuando se mide una variable en varios sujetos que están agrupados de alguna manera. Para tratar con este tipo de datos se han generado dos métodos: las ecuaciones de estimación generalizadas (GEE's por sus siglas en inglés) y los modelos multinivel o mixtos (Dobson, 2002). En el presente trabajo únicamente se revisarán los últimos.

Un supuesto de los modelos de regresión clásicos es que las unidades medidas son independientes. Sin embargo, existen situaciones en las que resulta imposible satisfacer dicha condición. En estas situaciones utilizar modelos multinivel representa una ventaja, ya que se relajan algunos de los supuestos de los modelos estándar, lo que permite manejar estructuras de datos más complejas de una manera más flexible. Estos modelos representan una aproximación diferente ya que toman en consideración la estructura jerárquica de los datos, permitiendo varias maneras en que las observaciones se correlacionen entre sí, sin que las estimaciones se vean afectadas (Brown y Prescott, 2006).

El término de datos jerárquicos se refiere a unidades agrupadas en diversos niveles. Generalmente, el nivel de observación más bajo en la jerarquía es referido como nivel 1 y los grupos o variables de agrupamiento como nivel 2, en donde cada unidad de nivel 1 pertenece a una única unidad de nivel 2. Por ejemplo, en medidas longitudinales, las medidas

repetidas representan el nivel 1, y los individuos medidos el nivel 2. Cabe mencionar que se pueden tener más de dos niveles. Debido al agrupamiento, las unidades dentro de cada grupo tenderán a ser más similares y por lo tanto a estar más correlacionados entre sí que con unidades de otros grupos. Es así que la definición de la unidad de análisis resulta muy importante e ignorar la estructura anidada de los datos puede llevar a estimaciones incorrectas de los errores estándar, intervalos de confianza y pruebas de significancia (Goldstein, 2011).

Al considerar que las observaciones de un estudio son independientes, se asume que cada observación provee 100 % de información nueva con respecto a las otras. No obstante, cuando en realidad existe correlación entre las medidas, cada observación provee menos del 100 % de información nueva. Esto se ve reflejado en que cuando se hace un análisis “ingenuo”, es decir, sin tomar en consideración la estructura de los datos, los errores estándar tienden a ser subestimados o sobrestimados, lo que puede llevar a interpretaciones incorrectas (Twisk, 2006). Una estimación correcta se obtendrá únicamente si la variación entre los grupos es permitida en el modelo. Los modelos multinivel proveen una manera de investigar la naturaleza de la variabilidad entre grupos y el efecto de características grupales (Goldstein, 2011). En la siguiente sección se describirán en términos generales estos modelos.

4.1. Características generales

Si tenemos un modelo de efectos fijos, se dice que los niveles de las variables explicativas están fijos porque no son seleccionados aleatoriamente y únicamente se hacen inferencias sobre los niveles especificados. En estos modelos se asume que los parámetros a estimar son constantes desconocidas (Myers et al., 2010). No obstante, una alternativa es considerar a los niveles de un factor como realizaciones de una distribución de probabilidad, es decir, que los niveles de una variable son una muestra aleatoria de una población de posibles niveles. En tal caso, se dice que el término es un efecto aleatorio. De manera que los niveles específicos ya no son de interés y en contraste con los efectos fijos, los efectos aleatorios

son representados por variables aleatorias, no por constantes. De aquí viene el nombre de los modelos mixtos, ya que contienen tanto efectos fijos como aleatorios (West, Welch, y Galecki, 2007).

Es así que los efectos fijos son usados para modelar el valor esperado de y , es decir, describen la relación entre la variable dependiente y las independientes. Mientras que los efectos aleatorios se ocupan de la estructura de varianzas-covarianzas de y , describiendo la desviación específica de cada grupo con respecto a los coeficientes fijos (Myers et al., 2010).

Consideremos el modelo lineal más sencillo que describe una relación en un solo nivel

$$y_i = \beta_0 + e_i \quad (4.1.1)$$

en donde y_i $i = 1, \dots, n$ es la i -ésima observación, β_0 es el intercepto y e_i $i = 1, \dots, n$ es el i -ésimo error, los cuales generalmente se asume siguen una distribución normal. En este caso ajustamos un único intercepto para todos los individuos, ya que se asume que el parámetro a estimar es una constante desconocida.

Sin embargo, si se considera que los datos están clasificados en grupos y que los grupos son una muestra aleatoria de una población más grande de grupos, entonces su efecto se puede modelar como aleatorio. Supongamos que en efecto los datos están agrupados en J grupos, si se permite que el intercepto del modelo 4.1.1 sea un efecto aleatorio, ajustando un intercepto por grupo, entonces el modelo multinivel más sencillo que podemos ajustar es (Goldstein, 2011):

$$y_{ij} = \beta_0 + u_{0j} + e_{ij} \quad (4.1.2)$$

$$\text{var}(e_{ij}) = \sigma_e^2, \quad \text{var}(u_{0j}) = \sigma_{u_0}^2$$

en donde y_{ij} es el valor de la i -ésima observación en el j -ésimo grupo $i = 1, \dots, n$ y $j = 1, \dots, J$, u_{0j} es una variable aleatoria que denota el error a nivel del grupo y es el efecto aleatorio, con esperanza cero y varianza $\sigma_{u_0}^2$. El parámetro β_0 es el intercepto promedio de todos los grupos y e_{ij} representa a los errores individuales, que en este caso son la diferencia entre el valor individual de cada observación y la media de su grupo. Por su parte, u_{0j} es la diferencia entre la media de cada grupo y la media general, de manera que

$\sigma_{u_0}^2$ representa qué tan variables son los interceptos de un grupo a otro. Estos residuales de nivel 2 suelen ser graficados junto con sus intervalos de confianza en lo que se conoce como gráfica de oruga, podemos observar un ejemplo en la figura 3.

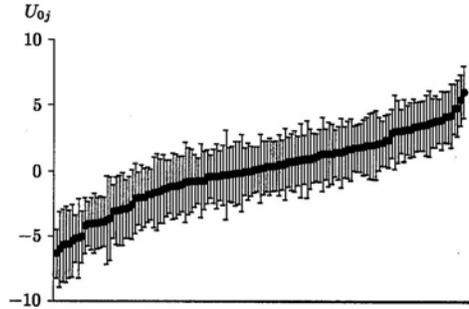


Figura 3: **Interceptos con su intervalo de confianza.** Tomada de Snijders y Bosker (1999).

Las variables aleatorias de este modelo son los errores, en donde e_{ij} es el error a nivel 1 y u_{0j} a nivel 2. Los errores en ambos niveles se consideran siguen una distribución normal. Lo que distingue a los modelos multinivel de los modelos clásicos es la presencia de varios términos de errores, razón por la que también se les conoce como modelos de componentes de varianza (Myers et al., 2010). Por ejemplo, la varianza de las respuestas para el modelo 4.1.2 quedaría de la siguiente manera:

$$\text{var}(y_{ij}) = \text{var}(u_{0j} + e_{ij}) = \sigma_e^2 + \sigma_{u_0}^2$$

esta es la suma de la varianza en ambos niveles, a σ_e^2 se le llama varianza intra-grupo (nivel 1) y a $\sigma_{u_0}^2$ varianza inter-grupos (nivel 2), ya que refleja la dispersión de un grupo a otro. Se asume que los errores de los diferentes niveles son independientes. Un índice comúnmente reportado en estos modelos es el Coeficiente de Correlación Intraclase (ICC, por sus siglas en inglés), el cual se refiere a la correlación de las observaciones dentro de cada grupo. El ICC se define como la varianza entre grupos entre la varianza total, esta última equivale a la suma de las varianzas entre-grupos e intra-grupo. Entre más pequeña sea la varianza intra-grupo mayor será el ICC (Twisk, 2006).

Al ajustar este tipo de modelos, la pregunta que surge es si realmente es necesario incluir el intercepto como un término aleatorio, es decir, si es necesario permitir que los

interceptos varíen de grupo a grupo. Es posible resolver esta pregunta comparando verosimilitudes, como fue explicado en el primer capítulo. En el caso de un modelo de intercepto aleatorio, se compara $-2 \log$ verosimilitud del modelo con el intercepto aleatorio con $-2 \log$ verosimilitud del modelo sin el intercepto aleatorio. Recordemos que esta diferencia se distribuye como una ji-cuadrada con grados de libertad igual a la diferencia de parámetros a estimar de los dos modelos comparados (Twisk, 2006).

Podemos extender el modelo 4.1.2 para que incluyan variables explicativas, las cuales pueden entrar tanto a nivel 1 como a nivel 2. Si consideramos una variable x_{ij} como efecto fijo, entonces el modelo sería

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{ij}$$

en donde la relación entre x y y es lineal y el efecto aleatorio es el del intercepto, ya que el efecto de la variables x se considera fijo. La inclusión de covariables en el modelo nos puede dar información sobre si ciertas variables explican las diferencias observadas en los niveles de agrupación. Esto lo podemos analizar, viendo el cambio en la varianza entre grupos después de incluir una variable (Twisk, 2006). En la figura 4 vemos un ejemplo de cómo quedaría dicho ajuste, en donde el intercepto es diferente para cada grupo, mientras que la pendiente es la misma para todos.

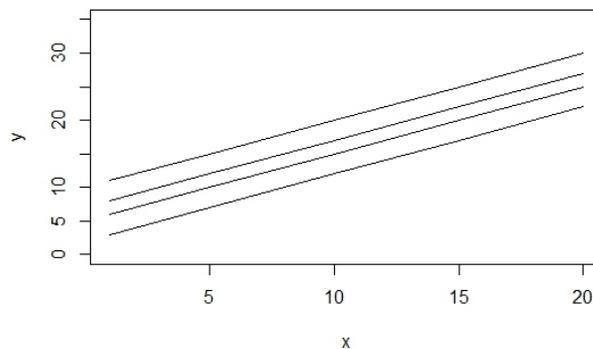


Figura 4: **Modelo de interceptos aleatorios.**

En este ejemplo se asume que la relación entre x y y es igual en todos los grupos. Podemos relajar este supuesto si la variable explicativa entra al modelo como efecto aleatorio,

de manera que la pendiente varía entre los grupos. En tal caso el modelo sería (Goldstein, 2011):

$$\begin{aligned}
 y_{ij} &= \beta'_0 + \beta'_1 x_{ij} + e_{ij} \\
 \beta'_0 &= \beta_0 + u_{0j} \\
 \beta'_1 &= \beta_1 + u_{1j}
 \end{aligned}$$

Lo cual lo podemos reescribir agrupando los efectos aleatorios de la siguiente manera

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + (u_{0j} + u_{1j} z_{ij} + e_{ij}) \quad (4.1.3)$$

en donde se diferencia entre las covariables x asociadas a los efectos fijos y las covariables z , las cuales se asocian con los efectos aleatorios. La pendiente promedio es β_1 , de manera que la pendiente para cada grupo j sería $\beta_1 + u_{1j}$. Cabe mencionar que también es posible agregar términos de interacción de la misma manera como se hace en regresión lineal clásica, sólo que en estos modelos es posible agregar interacciones “entre-niveles”. Así el modelo queda expresado como la suma de una parte fija que especifica la relación entre el valor esperado de y y las variables explicativas, y una parte aleatoria que contiene los residuales de ambos niveles.

Como tanto el intercepto como la pendiente son aleatorios cada uno añade un término de error. En la figura 5 podemos ver un ejemplo de cuando se deja que tanto los interceptos como las pendientes sean aleatorios.

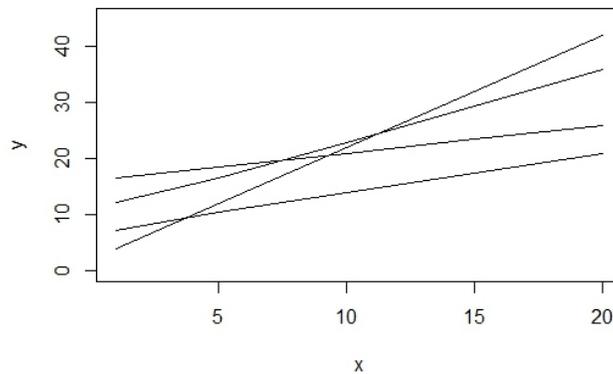


Figura 5: Modelo de interceptos y pendientes aleatorias.

Así como en el modelo con intercepto aleatorio, para probar si es necesario permitir en el modelo variabilidad en la pendiente, hay que comparar la log verosimilitud de un modelo que sólo incluya el intercepto aleatorio, con la log verosimilitud de un modelo que permita que tanto el intercepto como la pendiente sean efectos aleatorios. En este caso la diferencia de estas cantidades también sigue una distribución ji-cuadrada, con 2 grados de libertad, ya que en el modelo más complejo se agregan dos parámetros: la varianza de las pendientes y la covarianza de los interceptos con las pendientes (Twisk, 2006).

En los modelos de regresión lineal clásicos dado que se asume que todas las observaciones tienen varianza igual y que no están correlacionadas, la matriz de varianzas para y es la matriz diagonal $\text{var}(y) = \sigma_e^2 \mathbf{I}_n$. Sin embargo, cuando se asume que la estructura de los datos incluye agrupaciones y existen efectos aleatorios en el modelo, la matriz de varianzas para y es diferente. Es por ello que resulta importante distinguir claramente entre la matriz de covarianzas de los errores para cada grupo la cual representaremos como R_j y la matriz de covarianzas entre los efectos aleatorios D (West et al., 2007).

La matriz D puede tomar diversas formas, por ejemplo, podríamos pedir no correlación entre los efectos aleatorios. En dicho caso, para el modelo con intercepto y pendiente aleatorio se pediría que la covarianza fuera $\sigma_{u01} = 0$. En algunos casos es plausible asumir que los efectos aleatorios son independientes, sin embargo, hay situaciones en las que es mejor permitir una correlación entre los coeficientes que describen al mismo grupo (Myers et al., 2010). En el modelo 4.1.3 se asume que u_{0j} y u_{1j} siguen una distribución normal con media cero, varianza σ_{u0}^2 , σ_{u1}^2 respectivamente y covarianza σ_{u01} . Entonces la matriz de covarianzas de los efectos aleatorios D para este modelo sería (West et al., 2007):

$$\mathbf{D} = \text{var} \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} = \begin{pmatrix} \sigma_{0j}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{1j}^2 \end{pmatrix} \quad (4.1.4)$$

Una covarianza positiva entre los interceptos y las pendientes indica que grupos con interceptos más altos tienden a tener pendientes más pronunciadas, mientras que grupos con interceptos pequeños tendrían pendientes más horizontales. Esta interpretación se debe hacer siempre tomando en cuenta el signo de los coeficientes.

Finalmente, cabe mencionar que los parámetros β y las varianzas y covarianzas de

los efectos aleatorios suelen ser estimados mediante métodos de máxima verosimilitud o máxima verosimilitud restringida (Agresti, 2007).

Resumiendo, los modelos lineales mixtos son una herramienta muy útil, ya que extienden las capacidades de los modelos lineales clásicos permitiendo (West et al., 2007):

- Datos con estructura jerárquica
- Datos desbalanceados
- Matrices de covarianzas estructuradas tanto para los efectos aleatorios como para los errores

Un modelo con una estructura de varianzas y covarianzas definida que ajusta bien a los datos, en general llevará a estimaciones más precisas.

4.2. Modelos lineales generalizados mixtos

Como se ha mencionado en capítulos anteriores los GLM extienden la regresión clásica permitiendo respuestas no-normales y funciones liga del valor esperado. Con los modelos lineales generalizados mixtos (GLMM por sus siglas en inglés) podemos hacer una extensión más que permite modelar además de efectos fijos, efectos aleatorios en el predictor lineal. Los efectos aleatorios pueden ser incluidos en los GLMM de manera muy parecida a como son incluidos en los modelos lineales mixtos normales (Brown y Prescott, 2006).

Consideremos las observaciones y_{ij} en donde la i -ésima observación pertenece al j -ésimo grupo. Un GLMM tiene una forma muy parecida a un GLM con un término extra (Agresti, 2007):

$$g(\mu) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad (4.2.1)$$

en donde μ es el valor esperado, \mathbf{X} y \mathbf{Z} son las matrices de diseño para los efectos fijos y aleatorios respectivamente, y $\boldsymbol{\beta}$ y \mathbf{u} son los vectores de los efectos fijos y aleatorios. Los efectos aleatorios u se asume siguen una distribución Normal centrada en 0.

La regresión logística es parte de los modelos lineales generalizados, lo cual se verá en la siguiente sección. Si le agregamos efectos aleatorios, entonces los modelos logísticos multinivel siguen los mismos principios generales descritos en la sección (4.1), pero en este caso se trabaja con una variable dependiente dicotómica. Empecemos por describir la regresión logística.

4.2.1. Regresión logística

Uno de los tipos de datos categóricos más comunes es el de las variables binarias. La regresión logística es uno de los métodos estadísticos más usados para modelar este tipo de datos. Consideremos una variable de interés cuyos resultados sean dos categorías, las cuales suelen ser llamadas éxito y fracaso, con una probabilidad de éxito π y de fracaso $1 - \pi$. Si tenemos una serie de n ensayos independientes de dicha variable dicotómica, la variable aleatoria Y que representa al número de éxitos en los n ensayos, se considera tiene una distribución binomial (Dobson, 2002):

$$f(y; \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$$

con $y = 0, 1, 2, \dots, n$. Podemos reescribir esta función en su forma exponencial de la siguiente manera:

$$f(y; \pi) = \exp \left[y \log \left(\frac{\pi}{1 - \pi} \right) + n \log(1 - \pi) + \log \binom{n}{y} \right]$$

por lo tanto, es posible ver que la distribución binomial también forma parte de la familia exponencial.

Si consideramos el caso general de N variables aleatorias independientes Y_1, \dots, Y_N correspondientes al número de éxitos en N subgrupos, entonces $Y_i \sim \text{binomial}(n_i, \pi_i)$; y lo que nos interesa modelar son las proporciones de éxito $p_i = Y_i/n_i$ de cada subgrupo. En el caso más simple, la probabilidad de ocurrencia del evento es modelada directamente con una función lineal $\pi = \mathbf{x}^t \beta$. Sin embargo, el problema con este modelo lineal es que la relación de π con alguna variable x no tiende a ser lineal, ya que un cambio en x puede

tener un impacto diferente cuando π está cercana a 0, que cuando π está a la mitad del intervalo. Asimismo, tendríamos el problema de que los valores predichos pudieran caer fuera del intervalo (0,1), lo cual no tendría mucho sentido dado que se está tratando de modelar una probabilidad (Agresti, 2007).

Generalmente, una mejor opción para asegurarse de que la probabilidad esté restringida en un intervalo plausible y tenga una forma más adecuada, es modelarla mediante una función curva (creciente o decreciente) con forma de S (Myers et al., 2010). Para lograr esto, una posibilidad es que en vez de usar la probabilidad de un evento π tal cual, se modelen sus momios $\pi/(1 - \pi)$. A diferencia de las probabilidades, los momios pueden tomar cualquier valor de 0 a infinito. Si además transformamos los momios aplicándoles logaritmo, ahora el intervalo de los valores que puede tomar $\log(\frac{\pi}{1-\pi})$ es todos los reales. A esta transformación se le suele llamar $\text{logit}(\pi)$ (Snijders y Bosker, 1999). La figura 6 muestra la función logit .

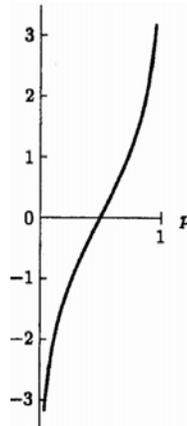


Figura 6: **Función logit**. Tomada de Snijders y Bosker (1999).

El modelo de regresión logística es un modelo donde $\text{logit}(\pi)$ es una función lineal de variables explicativas. El modelo más sencillo con una sola variable explicativa sería (Dobson, 2002):

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_i \quad (4.2.2)$$

La inversa de la función logaritmo es la función exponencial, de manera que la probabilidad

π_i queda (Snijders y Bosker, 1999):

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \quad (4.2.3)$$

Aunque la función liga logit es una de las más usadas, también se pueden considerar otras transformaciones que podrían llevar a modelos que ajusten mejor los datos. Por ejemplo, la liga probit o la liga log-log complementaria también pueden ser usadas (Collett, 2003a). Sin embargo, en el presente trabajo nos centraremos únicamente en la función logit.

Ya con el modelo básico de regresión logística definido, podemos ver que es parte de los GLM ya que: (a) y se distribuye como una binomial, que como ya se mostró es parte de la familia exponencial; (b) la función liga es la función $\text{logit} = \log(\pi_i/(1 - \pi_i))$; y (c) las variables explicativas entran al modelo como una combinación lineal de las β 's. Aunque π_i en efecto está restringida a tomar valores entre 0 y 1, $\text{logit}(\pi_i)$ puede tomar cualquier valor real (Agresti, 2007).

Si consideramos a x como una variable continua, entonces el modelo 4.2.2 implica que por cada aumento de una unidad en x , $\log[\pi(x)/(1 - \pi(x))]$ aumenta en cierta cantidad determinada por β_1 , en donde $\pi(x)$ es la probabilidad de éxito en el valor x . El signo de β_1 nos indica si la curva es ascendente (cuando β_1 es mayor a 0) o descendente (cuando β_1 es menor a 0). En el caso de que β_1 sea igual a 0, esto nos indica que X y la respuesta binaria Y son independientes, es decir, $\pi(x)$ es igual en todos los valores de x (Agresti, 2007).

Una interpretación importante de los modelos logísticos se basa en la razón de momios. Si tomamos el modelo 4.2.2 y exponenciamos ambos lados obtenemos:

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\beta_0 + \beta_1 x) = e^{\beta_0} (e^{\beta_1})^x \quad (4.2.4)$$

Entonces la interpretación de β_1 es en términos de momios. Veamos con más cuidado cómo se hace dicha interpretación para una variable explicativa continua. Consideremos los valores ajustados para x_i y para $x_i + 1$ dados por el modelo 4.2.2 (Myers et al., 2010):

$$\begin{aligned} \log(\text{momios}_{x_i}) &= \log\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1 x_i \\ \log(\text{momios}_{x_i+1}) &= \log\left(\frac{\pi(x_i + 1)}{1 - \pi(x_i + 1)}\right) = \beta_0 + \beta_1 (x_i + 1) \end{aligned}$$

la diferencia entre los dos valores predichos sería

$$\begin{aligned}\log(\text{momios}_{x_i+1}) - \log(\text{momios}_{x_i}) &= \beta_0 + \beta_1 x_i + \beta_1 - \beta_0 - \beta_1 x_i \\ &= \log\left(\frac{\text{momios}_{x_i+1}}{\text{momios}_{x_i}}\right) = \beta_1\end{aligned}\tag{4.2.5}$$

Si le sacamos exponencial a ambos lados de la ecuación 4.2.5 obtenemos que e^{β_1} es igual a la razón de momios, la cual la podemos interpretar como el incremento en las posibilidades de éxito asociado con un cambio de una unidad de la variable explicativa. Podemos generalizar eso a un incremento asociado a un cambio en r unidades de la variable explicativa como $e^{(r\beta_1)}$.

Es posible generalizar el modelo 4.2.2 para que tenga múltiples variables explicativas. Con k variables independientes, el modelo logístico general es (Agresti, 2007):

$$\text{logit}[\pi_i] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\tag{4.2.6}$$

Los parámetros β_i son el efecto de cada una de las variables x_i sobre el logaritmo de los momios de que ocurra un éxito, controlando por las demás x 's. Cuando hay más de una variable independiente, la interpretación de e^{β_1} es similar, lo entendemos como la razón de momios para un cambio en la variable x_1 , asumiendo que todos los demás predictores se mantienen constantes. También se pueden incluir predictores categóricos metiendo variables dummies al modelo. Generalmente, se toma una categoría de referencia a la cual se le asigna el valor de 0, de manera que si tenemos una variable con p categorías, necesitaremos incluir $p - 1$ variables indicadoras. En este caso también se pueden interpretar los coeficientes como el logaritmo de la razón de momios entre la categoría de referencia y las demás categorías de la variable (Brown y Prescott, 2006). Aunque este modelo no tiene términos de interacción, es posible incluirlos agregando términos del producto de las variables que se cree interaccionan.

Así como en regresión lineal clásica, también podemos obtener un intervalo de confianza para los coeficientes $\hat{\beta} \pm z_{\alpha/2}(SE(\beta))$, y si exponenciamos los límites obtenemos un intervalo de confianza para la razón de momios. Asimismo, podemos hacer pruebas de Wald a los coeficientes individuales, así como comparar las log-verosimilitudes de modelos

con y sin los parámetros de interés ($H_0 : \beta_1 = \dots = \beta_k = 0$) para determinar si el efecto de al menos una de las variables es significativo (Myers et al., 2010).

El formato de los datos es relevante para la evaluación de los modelos logísticos. Existen dos maneras en que se pueden dar los datos, una de ellas es agrupados por las diversas categorías de las variables explicativas de interés. De esta forma, lo que se tiene es el total de éxitos y fracasos para cada combinación de los valores de las variables predictoras categóricas. La otra posibilidad es que los datos tengan un formato no-agrupado, y se tengan las observaciones crudas de 1 y 0. Aunque para ambas maneras de presentar los datos los estimadores máximo verosímiles son iguales, algunos estadísticos como la X^2 y la devianza no pueden ser calculadas para el caso de datos desagrupados, ya que no se aproximarán a una distribución ji cuadrada. Esto suele pasar cuando las variables explicativas son continuas y toman valores diferentes para cada observación y por ello la manera de agruparlas no es inmediata. En estos casos la prueba de Hosmer-Lemeshow, la cual divide a los datos en grupos de acuerdo a las probabilidades estimadas, es una opción viable (Agresti, 2007; Myers et al., 2010).

Si cuando se ajusta un modelo logístico, éste describe adecuadamente las probabilidades observadas, se esperaría que la devianza fuera parecida a sus grados de libertad. Sin embargo, puede pasar que se crea que el modelo es correcto, pero la devianza sea mucho más grande que sus grados de libertad, en tal caso el supuesto de que la variabilidad se da como en una binomial puede que no sea válido. Cuando esto pasa se dice que existe sobredispersión. Una devianza grande se puede dar por diversas razones, por ejemplo, podría ser porque el componente lineal no sea adecuado, por la presencia de observaciones discrepantes, por el uso de una función liga inadecuada, etc. Si se ha revisado que estos problemas no estén en el modelo, entonces es razonable suponer que el valor grande de la devianza se debe a una variación mayor a la esperada o a una correlación entre las observaciones. Una posible solución a este problema es agregar efectos aleatorios al modelo (Collett, 2003a). Ya revisadas las características generales de la regresión logística, veamos su extensión como modelo mixto.

4.2.2. Regresión logística multinivel

La estructura básica de una regresión logística mixta con dos niveles involucra m grupos, es decir, unidades de nivel 2, cada una con n unidades de nivel 1. En donde la variable dependiente Y_{ij} es codificada como 0 y 1, y los subíndices i y j hacen referencia a los niveles 1 y 2, respectivamente. En un modelo de regresión logística con efectos aleatorios los grupos se consideran como una muestra de una población de grupos.

Los resultados de la regresión logística mixta pueden ser expresados e interpretados de la misma manera que en la regresión logística con sólo efectos fijos. La función liga logit sigue siendo utilizada y por lo tanto, los parámetros al ser exponenciados siguen siendo expresados en términos de razones de momios (Brown y Prescott, 2006).

El modelo logístico con intercepto aleatorio expresa al logaritmo de los momios como una combinación lineal de r variables explicativas y la desviación aleatoria de cada grupo u_{0j} (Snijders y Bosker, 1999):

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \sum_{k=1}^r \beta_k x_{kij} + u_{0j} \quad (4.2.7)$$

donde u_{0j} se asume sigue una distribución normal con media 0 y varianza $\sigma_{u_0}^2$. Un cambio de unidad en x_k entre los individuos de un mismo grupo, se asocia con una diferencia de magnitud β_k en el logaritmo de sus momios.

Así como en los modelos lineales mixtos el modelo de interceptos aleatorios puede ser extendido a un modelo con pendientes aleatorias. Consideremos el modelo con una pendiente aleatoria (Snijders y Bosker, 1999):

$$\text{logit}(\pi_{ij}) = \beta_0 + \sum_{k=1}^r \beta_k x_{kij} + u_{0j} + u_{1j} z_{1ij} \quad (4.2.8)$$

Ahora hay dos efectos aleatorios, uno para el intercepto y uno para la pendiente. Los efectos fijos se pueden probar de la misma manera que se ha comentado previamente, por ejemplo, utilizando la prueba de Wald. Los parámetros de varianza pueden ser evaluados comparando la devianza de un modelo que incluya al término de varianza de interés con la de uno que no lo incluya.

Dado que en los modelos logísticos multinivel la varianza total no es directamente estimable, se usa una manera alternativa de calcular la correlación intra-clase:

$$ICC = \frac{\sigma_{entre}^2}{\sigma_{entre}^2 + \left(\frac{\pi^2}{3}\right)} \quad (4.2.9)$$

Resulta matemática y computacionalmente difícil estimar los parámetros de los modelos logísticos multinivel, en especial, las varianzas y covarianzas de los efectos aleatorios. Para ello, se puede usar el método de máxima quasi-verosimilitud o el de quasi-verosimilitud penalizada. Los diferentes métodos suelen diferir en sus estimaciones, sobre todo en la estimación de los parámetros de la parte aleatoria.

Capítulo 5

Análisis de supervivencia

Análisis de supervivencia es el término que se le ha dado a los métodos de análisis de datos que tienen como variable dependiente el tiempo hasta un evento, por ejemplo, el tiempo hasta que un foco se funde o el tiempo que pasa hasta que un paciente muere. Los métodos usados en el análisis de supervivencia se ajustan al marco general de los modelos lineales generalizados, aunque la distribución de probabilidad asumida para los tiempos hasta el evento de interés no necesariamente pertenezca a la familia exponencial.

El análisis de supervivencia es muy usado en estudios epidemiológicos en los que se busca ver cuánto tiempo sobrevive un individuo, por lo que la ocurrencia del evento suele ser la muerte de un individuo o la incidencia de una enfermedad, y por ello se suele referir al evento como la ocurrencia de un fracaso. No obstante, esto no significa que el evento esté restringido a este tipo de sucesos. Asimismo, aunque múltiples eventos pueden ser de interés, en este caso nos centraremos en el análisis de la ocurrencia de un solo evento (Kleinbaum y Klein, 2012).

El tiempo de supervivencia t de una observación, se refiere al tiempo que pasa desde un momento inicial hasta la ocurrencia de un evento de interés, y puede ser considerado como un valor de la variable aleatoria T , la cual puede tomar cualquier valor no-negativo (Collett, 2003b). Existen ocasiones en donde las unidades bajo estudio nunca presentan el evento de interés durante el tiempo en que son estudiadas. En este caso, su tiempo de supervivencia es *censurado*, de manera que no conocemos el tiempo de supervivencia

verdadero, pero al menos sabemos que el evento no ocurrió antes del final del periodo de observación (Lee y Wang, 2003). Los datos que son censurados proporcionan información incompleta, pero aún así proveen información valiosa que no debería de ser descartada.

Aunque el tiempo hasta un evento puede considerarse como una variable continua, los procedimientos estándar que se usan para analizar este tipo de variables no son la mejor opción, ya que los datos de supervivencia no tienden a distribuirse simétricamente y por ello no sería adecuado recurrir a la distribución normal. Esto se podría solucionar transformando los datos, sin embargo, resulta mejor buscar otras distribuciones. Asimismo, la presencia de tiempos censurados hace inapropiados a los métodos comúnmente usados para datos continuos (Collett, 2003b).

Si se conoce qué distribución tienen los tiempos de supervivencia, entonces se puede hacer un análisis paramétrico, de otra manera, un análisis no paramétrico es una mejor opción (Lee y Wang, 2003). Varias distribuciones son usadas para caracterizar los tiempos de supervivencia, sin embargo, las más comúnmente empleadas son la exponencial y la Weibull. Aunque las distribuciones asumidas para los tiempos de supervivencia no siempre cumplen con todos los supuestos de los modelos lineales generalizados, la estimación de los parámetros basada en máxima verosimilitud funciona bien para muestras grandes de estas distribuciones (Dobson, 2002; Dunteman y Ho, 2006).

La distribución de supervivencia es comúnmente caracterizada por tres funciones: la función de densidad de probabilidad, de las cuales ya se mencionaron algunas comúnmente usadas, la función de supervivencia y la función de riesgo, estas últimas dos serán definidas en la siguiente sección (Lee y Wang, 2003).

5.1. Función de supervivencia y de riesgo

Como ya se mencionó en el análisis de supervivencia existen dos funciones muy importantes para caracterizar al tiempo hasta un evento: la función de supervivencia y la función de riesgo. Es importante describir estas funciones ya que en términos generales los propósitos de hacer un análisis de supervivencia son (Kleinbaum y Klein, 2012):

- Estimar e interpretar las funciones de supervivencia y riesgo
- Comparar las funciones de supervivencia y/o de riesgo
- Evaluar el efecto de variables explicativas sobre el tiempo de supervivencia

Dada la variable aleatoria T representando los tiempos de supervivencia, con $f(t)$ como su función de probabilidad, la función de supervivencia $S(t)$ se refiere a la probabilidad de que el tiempo de supervivencia T , sea igual o mayor al tiempo específico t . Sabemos que la probabilidad de que ocurra el evento de interés antes de un tiempo t está dada por la distribución de probabilidad acumulada (Dobson, 2002):

$$F(t) = Pr(T < t)$$

donde $F(t)$ es la función de distribución, e indica el área bajo la función de densidad que está a la izquierda de t . Como el área bajo todo el intervalo de la función de densidad es 1, entonces la función de supervivencia debe ser (Dunteman y Ho, 2006):

$$S(t) = Pr(T \geq t) = 1 - F(t) \tag{5.1.1}$$

Entonces podemos interpretar a $S(t)$ como la probabilidad de que un individuo sobreviva después del tiempo t .

En la figura 7 podemos ver las siguientes características teóricas de las funciones de supervivencia: (1) $S(t)$ es una función no creciente del tiempo t ; (2) cuando $t = 0$ la probabilidad de supervivencia es 1, es decir, la probabilidad de sobrevivir al menos hasta el tiempo cero necesariamente es 1; y (3) cuando $t = \infty$ entonces la probabilidad de supervivencia es 0, ya que teóricamente, todas las observaciones presentarán el evento en algún momento y la probabilidad de sobrevivir un tiempo infinito es cero (Kleinbaum y Klein, 2012). Una curva de supervivencia con una pendiente muy pronunciada indica una tasa de supervivencia baja o un tiempo de supervivencia corto. Una curva más aplanada y gradual se puede interpretar como una tasa y tiempo de supervivencia más altos (Lee y Wang, 2003).

Por otro lado, la función de riesgo es usada para expresar el riesgo de que el evento ocurra en algún tiempo t , es decir, es la probabilidad de que el evento se presente en el

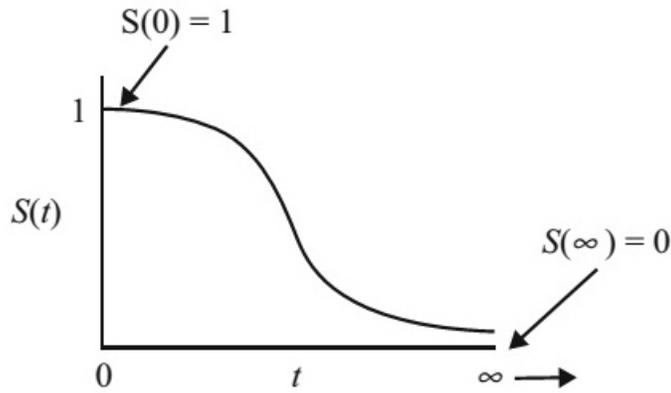


Figura 7: **Función de supervivencia.** Tomada de Kleinbaum y Klein (2012).

tiempo t dado que se ha sobrevivido hasta dicho tiempo. Si consideramos la probabilidad de que el tiempo de supervivencia esté en el intervalo de tiempo $[t, (t + \Delta t)]$, condicional a que se sobrevivió al menos hasta dicho tiempo, esto lo podemos escribir como $Pr(t \leq T < t + \Delta t \mid T \geq t)$. Podemos expresar esta probabilidad condicional por unidad de tiempo dividiéndola entre el intervalo de tiempo Δt , y lo que obtenemos es una tasa. Entonces la función de riesgo está dada por el límite (Collett, 2003b):

$$\lim_{\Delta t \rightarrow 0} \frac{Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

Podemos reescribir a la función de riesgo como

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad (5.1.2)$$

Vemos que la función de riesgo está relacionada con la función de supervivencia. La función de riesgo también es conocida como tasa de fracaso instantánea, tasa de mortalidad condicional o fuerza de mortalidad.

En la práctica, cuando no hay datos censurados, la función de riesgo se estima como la proporción de observaciones que presentan el evento en el intervalo por unidad de tiempo, dado que han sobrevivido hasta el inicio del intervalo (Lee y Wang, 2003). La función de riesgo $h(t)$ se puede interpretar como la probabilidad instantánea de que el evento pase en el tiempo t , dado que la observación ha sobrevivido hasta ese momento (Dunteman y Ho, 2006). En la figura 8 vemos ejemplos de varias funciones de riesgo y es posible observar

que en todos los casos, para un valor en específico de t , la función de riesgo siempre es igual o mayor a 0 y no tiene un límite superior, ya que es una tasa, no una probabilidad.

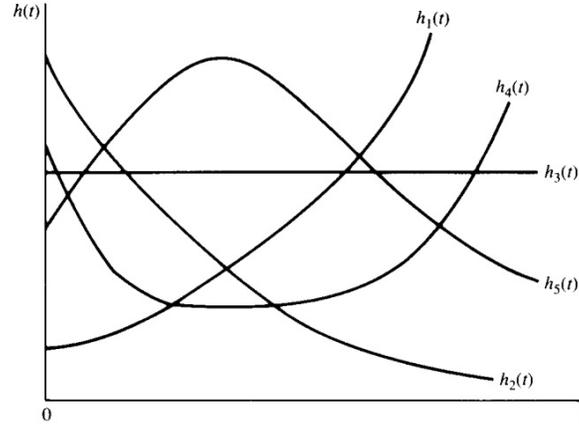


Figura 8: **Función de riesgo.** Tomada de Lee y Wang (2003).

Asimismo, la función de riesgo puede ser constante, es decir, independientemente del tiempo t considerado, el valor de $h(t)$ siempre es el mismo. Cuando esto pasa el modelo de supervivencia es exponencial y su función de riesgo sería la $h_3(t)$ en la figura 8 (Kleinbaum y Klein, 2012). En este caso vemos que en efecto la función de riesgo no depende del tiempo t , de manera que la probabilidad del evento en cualquier intervalo no se relaciona con cuánto tiempo se haya logrado sobrevivir. La función de riesgo también puede ser creciente, decreciente o incluso indicar un patrón más complejo como es mostrado por las funciones $h_4(t)$ y $h_5(t)$ (Dobson, 2002; Lee y Wang, 2003).

También resulta útil considerar la función de riesgo acumulativa, la cual está dada por (Lee y Wang, 2003):

$$H(t) = \int_0^t h(x)dx \quad (5.1.3)$$

Existe la siguiente relación de esta función con la función de supervivencia

$$H(t) = -\log[S(t)] \quad (5.1.4)$$

$H(t)$ es el total de riesgo acumulado hasta el tiempo t , de manera que en el tiempo 0 $H(t) = 0$ y en el tiempo ∞ $H(t) = \infty$.

Independientemente de cual de las funciones se analice, hay una clara relación entre ellas. Por un lado, la función de supervivencia se enfoca en el aspecto de que no suceda el fracaso o evento, es decir, de sobrevivir; mientras que la función de riesgo se enfoca en la ocurrencia del evento. Podría decirse que la función de riesgo nos da la información opuesta a la dada por la función de supervivencia. De hecho, a partir de $S(t)$ se puede derivar $h(t)$ y viceversa (Dunteman y Ho, 2006).

Es posible usar las distribuciones de Weibull o la exponencial para analizar las funciones de supervivencia y riesgo, sin embargo, existen procedimientos de estimación no paramétricos. En la siguiente sección describiremos uno de los más usados en datos de supervivencia, las estimaciones Kaplan Meier.

5.2. Curvas Kaplan Meier

Un problema básico en el análisis de tiempos hasta un evento es estimar la función de supervivencia a partir de los datos obtenidos y hacer inferencias de los patrones de supervivencia. Como ya se mencionó, dada la variable aleatoria de tiempos de supervivencia T , la función calculada de supervivencia es un estimado de la probabilidad de sobrevivir más allá del tiempo t . Una manera de calcular la probabilidad de supervivencia que permite la presencia de tiempos censurados es mediante el método de Kaplan Meier, en el cual se estiman las probabilidades de supervivencia usando la fórmula del producto límite. Esta fórmula involucra el producto de probabilidades condicionales, en donde cada término que se multiplica es la probabilidad de exceder un tiempo específico t_i dado que se ha sobrevivido hasta ese tiempo. La fórmula de Kaplan Meier es (Kleinbaum y Klein, 2012):

$$\hat{S}(t_f) = \prod_{i=1}^f \hat{Pr}(T > t_i | T \geq t_i) \quad (5.2.1)$$

donde t_f es el tiempo en el que ocurrió el evento. La fórmula para la probabilidad de supervivencia está limitada al producto de los términos hasta el tiempo t_i especificado. Por ejemplo, la probabilidad de sobrevivir hasta el tiempo t_k se expresaría como:

$$\hat{S}(k) = p_1 \times p_2 \times \cdots \times p_k$$

en donde p_1 es la proporción de observaciones que sobrevivieron al menos hasta el tiempo 1, p_2 es la proporción de observaciones que sobrevivieron hasta el segundo tiempo dado que sobrevivieron hasta el tiempo 1 y p_k es la proporción de observaciones que sobrevivió hasta el tiempo k , dado que ya habían sobrevivido hasta el tiempo $k - 1$. Cabe mencionar que los estimados mediante la fórmula de producto límite son estimadores de máxima verosimilitud (Lee y Wang, 2003).

En la práctica, se calcula $\hat{S}(t)$ para cada tiempo de supervivencia diferente. En los intervalos en los que no ocurre ningún evento, $\hat{S}(t)$ se mantiene constante, por lo que la gráfica suele ser escalonada. Cuando se grafica $\hat{S}(t)$ contra t , se pueden obtener los diversos percentiles del tiempo de supervivencia. Uno de los más reportados es la mediana, la cual es el valor del tiempo de supervivencia en donde $\hat{S}(t) = 0.5$. Podemos comparar de manera general dos poblaciones de tiempos de supervivencia a partir de sus medianas u otros percentiles. Generalmente se reporta la mediana en vez de la media de los tiempos de supervivencia porque su distribución suele ser asimétrica (Lee y Wang, 2003).

Otra medida descriptiva que puede ser usada es la tasa promedio de riesgo. Este promedio se obtiene dividiendo el total de fracasos o presentaciones del evento, entre la suma de los tiempos de supervivencia observados. Mientras más grande sea la tasa de riesgo, más baja será la probabilidad de supervivencia (Kleinbaum y Klein, 2012). Estas dos medidas, la mediana de los tiempos de supervivencia y la tasa promedio de riesgo, nos dan una manera de hacer comparaciones globales entre diferentes curvas, pero no nos permiten comparar las curvas en varios momentos, lo cual si se puede hacer a partir de las gráficas de Kaplan Meier descritas.

A las curvas de supervivencia Kaplan Meier se les pueden calcular intervalos de confianza. Como con cualquier otro estimador la fórmula para calcular el intervalo al 95 % es (Kleinbaum y Klein, 2012):

$$\hat{S}(t) \pm 1.96\sqrt{Var(\hat{S}(t))}$$

en donde $\hat{S}(t)$ es la supervivencia estimada por medio de Kaplan Meier en el tiempo t , y la parte dentro de la raíz es su varianza.

El método de Kaplan Meier es uno de los métodos más usados en el análisis de datos de supervivencia, ya que provee de un método muy útil para estimar probabilidades de supervivencia y para presentar gráficamente las curvas de supervivencia.

5.2.1. Prueba log-rank

Comparar dos curvas de supervivencia es un problema que suele surgir en muchas investigaciones con datos de supervivencia. Esta comparación se puede hacer dibujando las curvas e inspeccionándolas, sin embargo, esto solo da una idea intuitiva de las posibles diferencias, pero no nos indica si dicha diferencia es significativa o simplemente producto del azar. La prueba log-rank es una de las más usadas para comparar curvas de Kaplan Meier. Esta prueba nos sirve para ver si dos o más curvas son estadísticamente iguales mediante la comparación entre observado y esperado en cada tiempo en el que ocurrió el evento (Lee y Wang, 2003).

Empezando por el caso más sencillo en donde sólo se tienen dos curvas a comparar, los valores esperados para la curva 1, están dados por (Kleinbaum y Klein, 2012):

$$e_{1i} = \left(\frac{n_{1i}}{n_{1i} + n_{2i}} \right) \times (m_{1i} + m_{2i})$$

en donde e_{1i} es el valor esperado de observaciones que presentarán el evento en el i -ésimo tiempo, $n_{1i}/(n_{1i} + n_{2i})$ es la proporción de sujetos en riesgo, es decir, que han sobrevivido hasta el i -ésimo tiempo y $(m_{1i} + m_{2i})$ es el número total de fracasos observados de ambos grupos en el tiempo i . Para comparar los dos grupos la prueba log-rank es:

$$Log - rank = \frac{(\sum_{i=1}^k (m_{gi} - e_{gi}))^2}{Var(\sum_{i=1}^k (m_{gi} - e_{gi}))}, \quad g = 1, \dots, g=no. \text{ grupos} \quad (5.2.2)$$

en donde k es el número de tiempos en los que se presentó el evento o el fracaso. Este estadístico se obtiene dividiendo el cuadrado de la suma de las diferencias $m_{gi} - e_{gi}$, entre la varianza de dicha suma. La suma se hace sobre todos los tiempos en los que se presentó el evento de interés. La diferencia entre esperado y observado se puede hacer para cualquiera de los grupos y el resultado será el mismo. La hipótesis nula es que no hay diferencias entre las dos curvas de supervivencia, $S_1(t) = S_2(t)$. Bajo esta hipótesis, el estadístico log-rank se distribuye aproximadamente como una ji-cuadrada con un grado de libertad.

Este estadístico se puede generalizar a más de dos grupos, en donde la hipótesis nula es que todas las curvas de supervivencia son iguales. En este caso, el estadístico no es tan sencillo de calcular, ya que involucra varianzas y covarianzas de las sumas de las diferencias entre observado y esperado para cada grupo. Al igual que con dos curvas, bajo la hipótesis nula la prueba log-rank se distribuye como una ji cuadrada con $G - 1$ grados de libertad, en donde G es el número de curvas a comparar ([Kleinbaum y Klein, 2012](#)).

Cuando existen datos censurados hay otras pruebas que se pueden utilizar en la comparación de dos curvas de supervivencia además de la log-rank, como la prueba de Cox-Mantel, la Tarone-Ware y la Peto. Todas estas pruebas son variaciones de la prueba de log-rank, por lo que para el presente estudio sólo se comentarán como posibles opciones. Asimismo, si los datos no tienen tiempos censurados, entonces pruebas no paramétricas estándar como la prueba de Wilcoxon o la de Mann-Whitney pueden ser usadas ([Lee y Wang, 2003](#)).

Capítulo 6

Resultados

6.1. Descripción del experimento

Los datos analizados en el presente estudio fueron el resultado de un experimento que se realizó en la Facultad de Psicología de la UNAM en el laboratorio de Adaptación Animal del Dr. Gustavo Bachá. Empezaremos por describir dicho experimento.

6.1.1. Sujetos experimentales

Se utilizaron 16 ratas hembras de la cepa Wistar experimentalmente ingenuas¹, con aproximadamente 4 meses de edad al inicio del experimento. Fueron mantenidas al 80 % de su peso *ad libitum* durante el transcurso de todo el experimento y se alojaron en cajas habitación individuales con acceso libre a agua. Las sesiones experimentales se llevaron a cabo 5 días por semana siempre a la misma hora. Dos horas después de cada sesión experimental se alimentaba a las ratas para mantener su peso corporal.

6.1.2. Material

Se utilizaron cuatro cámaras experimentales de condicionamiento operante (MED Associates), cada una de ellas ubicada dentro de una caja sono amortiguadora. En la figura

¹ratas que nunca habían participado en un experimento

9 se muestra una caja operante. Todas las cajas estuvieron conectadas a una interfase MED Associates que controló cada sesión experimental.



Figura 9: **Caja operante.**

En el panel posterior de la caja se encontraba una luz general de 28 v. En el panel frontal se encontraban dos palancas: una del lado izquierdo y una del derecho, y arriba de cada una se encontraba un foco de 28 v. Entre las dos palancas había una abertura cúbica de 4 cm por lado, con un orificio en la parte inferior a través del cual subía una copa que estaba en el extremo de un brazo que era activado dependiendo del programa de reforzamiento establecido. La copa proporcionaba 0.1 ml de leche ultra-pasteurizada a las ratas.

6.1.3. Fases

Moldeamiento. Fase en la que se enseñó a las ratas a presionar las palancas para obtener el reforzador, que en este caso fue 0.1 ml de leche ultra-pasteurizada. Dado que las ratas fueron mantenidas al 80 % de su peso, estaban motivadas a buscar comida. Para entrenarlas a responder, se les puso en un programa en donde toda respuesta de apretar cualquiera de las dos palancas de la caja iba seguida de un reforzador.

Entrenamiento. En esta fase se enseñó a las ratas a ejecutar secuencias de respuestas en vez de una sola respuesta. De aquí en adelante llamaremos a las palancas de acuerdo a su localización: izquierda (I) y derecha (D). Para esta fase, primero se entrenó a las ratas a ejecutar secuencias de dos respuestas a dos palancas. Una vez que esto había sido logrado,

se les reforzaba cualquier secuencia de tres respuestas a dos palancas que involucrara al menos una alternación entre palancas, antes de pasarlas a su fase experimental.

Fase experimental. En esta fase primero se asignó aleatoriamente a las ratas a uno de cuatro grupos: AAA, ABB, ABA y AAB. De manera que cada grupo quedó conformado por cuatro sujetos. Después se asignó aleatoriamente a cada rata a una de las 4 cajas experimentales, con la restricción de que a no más de dos ratas del mismo grupo les correspondiera la misma caja. La asignación a los diferentes grupos definió qué subconjuntos de las posibles secuencias de tres respuestas serían reforzadas.

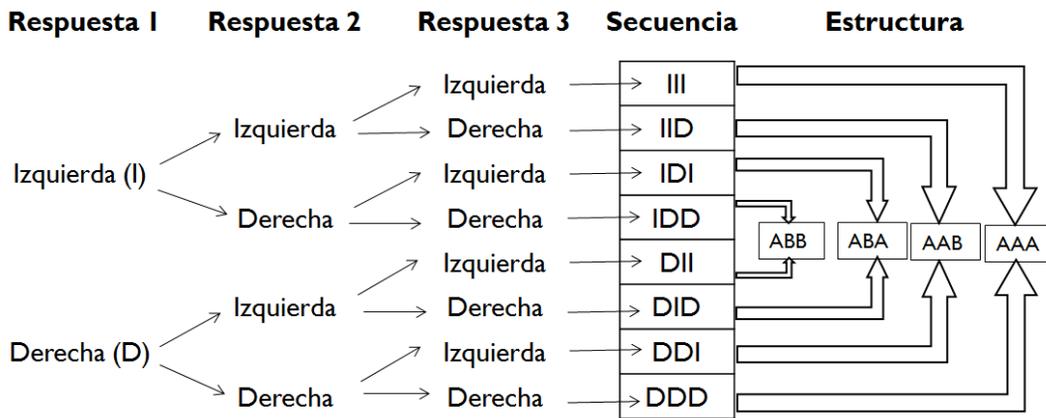


Figura 10: Secuencias de tres respuestas a dos palancas.

En la figura 10 podemos ver un esquema de todas las posibles secuencias de 3 respuestas que se pueden ejecutar en dos palancas. Asimismo, se puede observar la agrupación de las secuencias en 4 posibles conjuntos de acuerdo a su estructura espacio temporal. De manera que las secuencias III y DDD tienen la misma estructura AAA, ya que sólo involucran repetir la misma respuesta en una de las palancas. Las secuencias con estructura ABA involucran realizar una doble alternación entre las palancas, de manera que tanto IDI como DID tienen la misma estructura. El mismo razonamiento aplica para las demás estructuras. Resumiendo, en la tabla 3 se puede ver qué secuencias específicas fueron reforzadas según el grupo.

Si la rata ejecutaba alguna de las secuencias reforzadas en su grupo, entonces sonaba un tono durante 5 segundos y se entregaba 0.1 ml de leche. Si la secuencia ejecutada no

Tabla 3: Secuencias reforzadas.

Grupo	Secuencias reforzadas
AAA	III y DDD
ABB	IDD y DII
ABA	IDI y DID
AAB	IID y DDI

era la correcta, entonces se apagaban todas las luces de la cámara experimental durante 2 segundos, no se entregaba el reforzador al animal y las respuestas emitidas durante este periodo no tenían ningún efecto programado. Esto se hizo con el propósito de demarcar mediante estímulos visuales los ensayos, para que las ratas pudieran agrupar su conducta en “subconjuntos” de secuencias de tres respuestas. Se corrieron múltiples sesiones hasta que las ratas acumularan 2,000 reforzadores. El número de reforzadores se estableció para permitir que los animales llegaran a un estado estable en su conducta. El número de sesiones necesarias para llegar a los 2,000 reforzadores varió de rata a rata. Hubo algunas que en 15 sesiones consiguieron todos los reforzadores establecidos, mientras que otras tardaron más de 40 sesiones. Esta variabilidad, también se notó entre los diferentes comportamientos registrados.

6.1.4. Estructura de los datos

Para entender la estructura de los datos es necesario describir en qué consistieron las sesiones experimentales. Cada sesión inició con la luz general y las luces sobre las palancas prendidas hasta que el sujeto realizaba una secuencia de tres respuestas de la manera que quisiera. Cada secuencia de tres respuestas, ya fuera reforzada o no, fue considerada como un ensayo. Se corría una sesión al día, y cada sesión consistía de 150 ensayos aproximadamente. Es decir, durante una sesión, las ratas tenían la oportunidad de realizar 150 secuencias de respuestas. Un esquema de los datos se muestra en la figura 11.

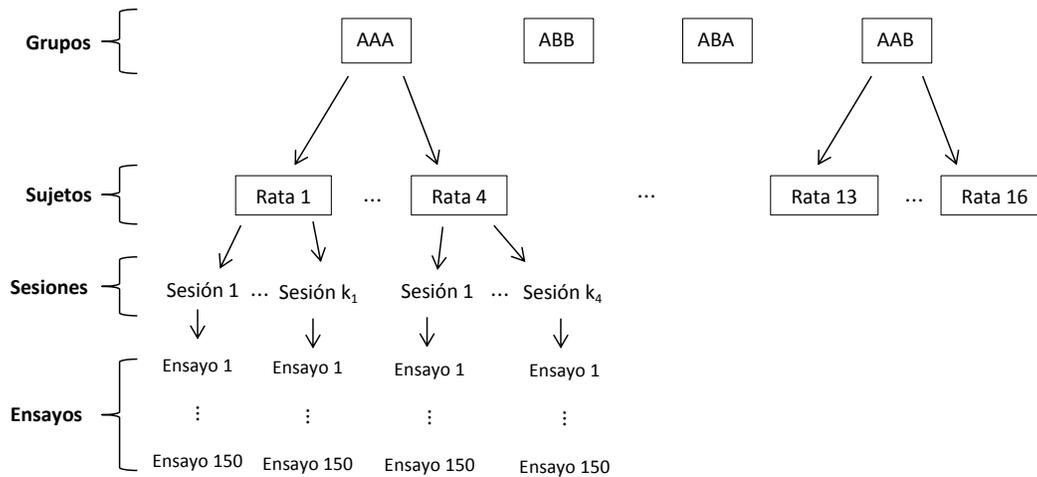


Figura 11: **Estructura de los datos.**

En cada sesión se midió cuántas veces ejecutaron cada una de las posibles estructuras conductuales, reforzadas y no reforzadas. De manera que se tuvo la frecuencia de las 4 estructuras en cada una de las sesiones hasta que las ratas acumularon todos los reforzadores establecidos. Asimismo, se midió la secuencia temporal con la que se ejecutaron las secuencias dentro de cada sesión. De manera que se obtuvo también, una base de datos más larga con una variable binaria que tomó el valor de 1 en los ensayos en los que la estructura reforzada había sido ejecutada y 0 en los ensayos en los que cualquiera de las otras estructuras fue ejecutada.

Es así que las variables dependientes a analizar fueron: 1) la frecuencia de cada una de las 4 estructuras conductuales en cada sesión y 2) la variable dicotómica indicando la ejecución o no de la estructura reforzada en cada ensayo. Las variables explicativas que fueron analizadas son: el grupo, las sesiones, la conducta en el ensayo anterior y variables artificiales que se usaron para poner a prueba ciertas hipótesis que serán explicadas en las secciones posteriores.

6.1.5. Objetivos

El experimento tuvo como objetivo general explorar si la estructura espacio-temporal del comportamiento reforzado es una variable importante en la manera en que los animales distribuyen su conducta. Los objetivos específicos del presente estudio fueron:

- Analizar las curvas de adquisición de las conductas reforzadas según su estructura
- Estudiar si existe una relación entre la estructura reforzada y la distribución de las posibles conductas, tanto las reforzadas como la no reforzadas
- Probar si una explicación basada en la hipótesis de contigüidad explica bien la distribución de la conducta de los animales
- Analizar el efecto de variables moleculares (e.g. conducta en el ensayo anterior) y molares (e.g. pertenencia al grupo) sobre la conducta de los sujetos

6.2. Análisis de supervivencia

En esta sección describiremos los resultados del análisis de supervivencia llevado a cabo. Los análisis se corrieron en el programa R ([R Core Team, 2013](#)) con la librería `survival`. Para calcular los tiempos de supervivencia se utilizó la función `Surv`. Las curvas de Kaplan-Meier se calcularon mediante la función `survfit` y para hacer las pruebas log-rank para probar igualdad de las curvas de supervivencia se utilizó `survdif`.

Lo primero que se hizo fue realizar una curva Kaplan-Meier por sujeto para la estructura de conducta reforzada. Como a todas las ratas se les estableció el criterio de ejecutar 2,000 veces la estructura reforzada para que el experimento llegara a su fin, podríamos conceptualizar el análisis como que tenemos una población de 2,000 conductas y el objetivo es estudiar cómo esas conductas se fueron dando a lo largo de los ensayos en cada rata. En términos del análisis de supervivencia la falla o muerte se presenta cuando se realiza la conducta reforzada. Para ello, se construyó una variable en la que en cada ensayo, si la rata ejecutaba alguna de las secuencias de la estructura reforzada, se le asignaba el

número 1 y los ensayos en los que no ocurrían las conductas reforzadas no fueron considerados. No se asignó 0 a las conductas no reforzadas ya que R considera que 0 indica un tiempo de supervivencia censurado. De tal forma que el tiempo para la siguiente falla fue el número de ensayos necesarios para realizar nuevamente la conducta reforzada. Dado que el experimento se terminó de acuerdo a la ejecución de la estructura reforzada, cada rata podía terminar en tiempos diferentes y las conductas también podían ser ejecutadas en momentos diferentes del experimento. Estos análisis requieren de observaciones independientes, y dado que la ejecución de la conducta reforzada se midió muchas veces en una misma rata, dicho supuesto no se cumple. Sin embargo, este primer análisis fue meramente descriptivo y una aproximación inicial para conocer el comportamiento de los sujetos experimentales.

En la figura 12 podemos ver las curvas de Kaplan-Meier por sujeto. En el eje de las ordenadas está la probabilidad de supervivencia y en el eje de las abscisas están los ensayos. Los colores de las curvas hacen referencia a los grupos de acuerdo a la estructura reforzada. Por ejemplo, las 4 líneas azules representan las curvas de supervivencia de las 4 ratas a las cuales se les reforzó las secuencias con estructura AAA. Las 12 líneas restantes corresponden a las demás ratas a las cuales se les reforzaron las otras estructuras.

Es posible observar que hubo mucha variabilidad en cuanto a la ejecución de cada rata, sin embargo, la estructura sí parece haber tenido un efecto sobre la ejecución de la estructura reforzada. Para las ratas cuyas secuencias reforzadas eran de la estructura AAA las 2,000 conductas ocurrieron muy rápido, en poco más de 2,000 ensayos ya habían ocurrido todas las conductas. Esto nos indica que en prácticamente cada ensayo realizaron la conducta reforzada, teniendo una ejecución casi perfecta. Las cuatro curvas de este grupo (las líneas azules) se sobrelapan, lo que podría indicar que la ejecución de la estructura reforzada fue muy similar entre las ratas de este grupo.

Previamente se había mencionado que una medida para comparar las curvas era la mediana, en la tabla 4 podemos observar las medianas para las cuatro ratas del grupo AAA, junto con su intervalo de confianza al 95% y el número de ensayos ejecutados en total.

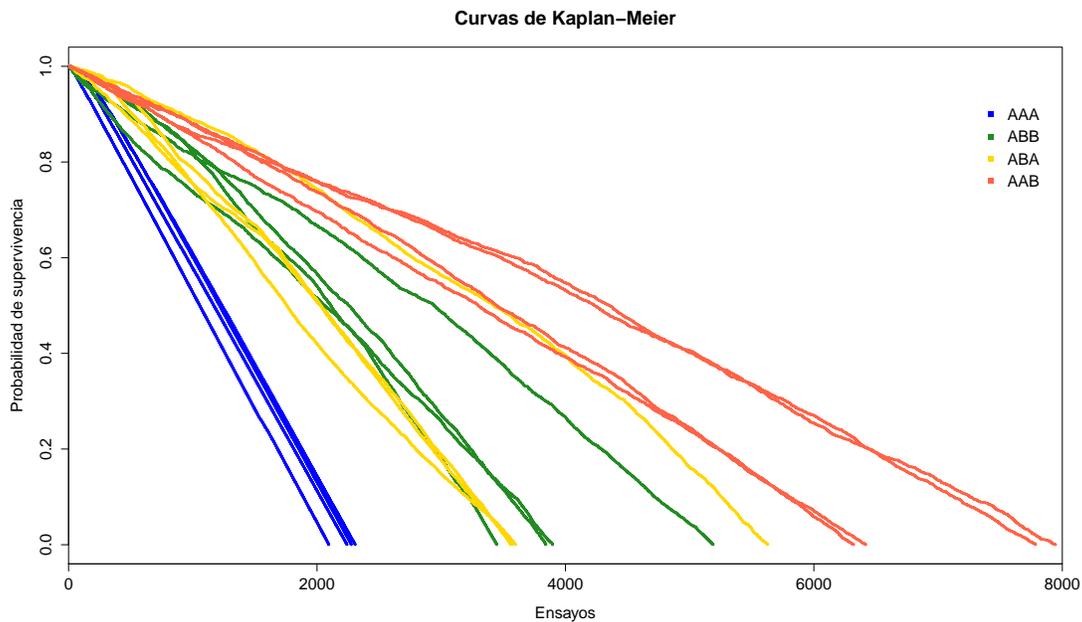


Figura 12: **Curvas Kaplan-Meier para la estructura reforzada por rata.**

Vemos que las ratas de este grupo tardaron entre 2,096 y 2,308 ensayos en completar la fase experimental y todas tienen su mediana alrededor de los mil ensayos. Es claro que en este grupo la conducta que ocurrió en prácticamente todos los ensayos tuvo la estructura AAA.

En cuanto a los grupos ABB y ABA vemos que las curvas de supervivencia de las ratas de ambos grupos se traslapan. En este caso vemos más variabilidad en cuanto a los ensayos que se necesitaron para que ocurrieran las 2,000 conductas de la estructura reforzada. En la tabla 5 y 6 vemos los ensayos, medianas y los intervalos de confianza al 95 % de las medianas de las curvas de los sujetos en estos dos grupos.

En estas dos tablas vemos que la mayoría de las ratas de ambos grupos tardaron entre 3,500 y 4,000 ensayos en obtener todos los reforzadores. Sin embargo, vemos que hubo una rata en cada grupo que tardó mucho más que las demás, aproximadamente 5,000 ensayos. Sin tomar en cuenta a estas dos ratas, vemos que las demás tienen una mediana de alrededor de 2,000 ensayos, momento en que la probabilidad de supervivencia era de 0.5 . Vemos que las medianas de las curvas de estos dos grupos son casi el doble de las

Tabla 4: Medianas de las curvas Kaplan-Meier del grupo AAA.

Sujeto	Ensayos	Mediana	Límite inferior	Límite superior
Rata 1	2096	1056	1012	1101
Rata 2	2277	1216	1172	1261
Rata 3	2308	1234	1189	1284
Rata 4	2242	1169	1124	1216

Tabla 5: Medianas de las curvas Kaplan-Meier del grupo ABB.

Sujeto	Ensayos	Mediana	Límite inferior	Límite superior
Rata 5	3900	2060	1973	2160
Rata 6	5190	2944	2797	3039
Rata 7	3845	2258	2164	2326
Rata 8	3450	2112	2050	2171

medianas del grupo AAA. Por lo que podríamos inferir que es más complicado ejecutar ABB y ABA con respecto a AAA. Entre los dos grupos, ABB y ABA no parece haber una diferencia notable, aunque las ratas 11 y 6 tardaron mucho más tiempo que las otras 6 ratas de sus grupos, cuyas curvas están muy juntas.

Finalmente, las curvas Kaplan-Meier para las ratas del grupo AAB se alargan hasta los 8,000 ensayos. En este caso vemos que hay más variabilidad entre las ratas ya que hubo un par de ratas que tardaron un poco más de 6,000 ensayos, mientras que las otras dos del grupo tardaron prácticamente 8,000 ensayos. En la tabla 7 se muestran ensayos y medianas para cada rata.

Las curvas de supervivencia de todas las ratas de este grupo están por arriba de las curvas de las ratas de los otros grupos. Es decir, estas ratas tardaron mucho más tiempo y ejecutaron la estructura reforzada más paulatinamente. En este grupo la mediana varió entre 3,200 y 4,300 aproximadamente. Las medianas de las curvas de este grupo son

Tabla 6: Medianas de las curvas Kaplan-Meier del grupo ABA.

Sujeto	Ensayos	Mediana	Límite inferior	Límite superior
Rata 9	3600	1758	1700	1816
Rata 10	3561	2030	1959	2090
Rata 11	5632	3422	3304	3566
Rata 12	3574	2032	1971	2095

Tabla 7: Medianas de las curvas Kaplan-Meier del grupo AAB.

Sujeto	Ensayos	Mediana	Límite inferior	Límite superior
Rata 13	6491	3297	3162	3435
Rata 14	7787	4357	4188	4472
Rata 15	6321	3444	3335	3594
Rata 16	7950	4234	4074	4360

más del triple de las medianas del grupo AAA. Asimismo, con respecto a los grupos ABB y ABA la mediana de este grupo estuvo entre 1,000 y 2,000 ensayos mayor. Definitivamente, es posible observar que las medianas entre los grupos fueron bastantes diferentes, lo que nos podría ayudar a definir un orden de dificultad de las estructuras. En la figura 13 se muestran las medianas de todos los sujetos graficadas junto con sus intervalos de confianza a manera de resumen.

Se realizaron pruebas log-rank dentro de cada grupo, para ver si era plausible la hipótesis de que las ratas dentro de cada grupo se comportaron de manera similar. Las pruebas llevadas a cabo para cada grupo resultaron altamente significativas, lo que indica que hubo mucha variabilidad entre la ejecución de las ratas. En la tabla 8 se puede ver el valor del estadístico, los grados de libertad (que en este caso fueron 3 para las 4 pruebas ya que había 4 sujetos por grupo) y el valor de p , que fue tan pequeño que R lo redondeó a 0. Se realizaron las mismas comparaciones con la prueba Kruskal Wallis, pero se encontró el

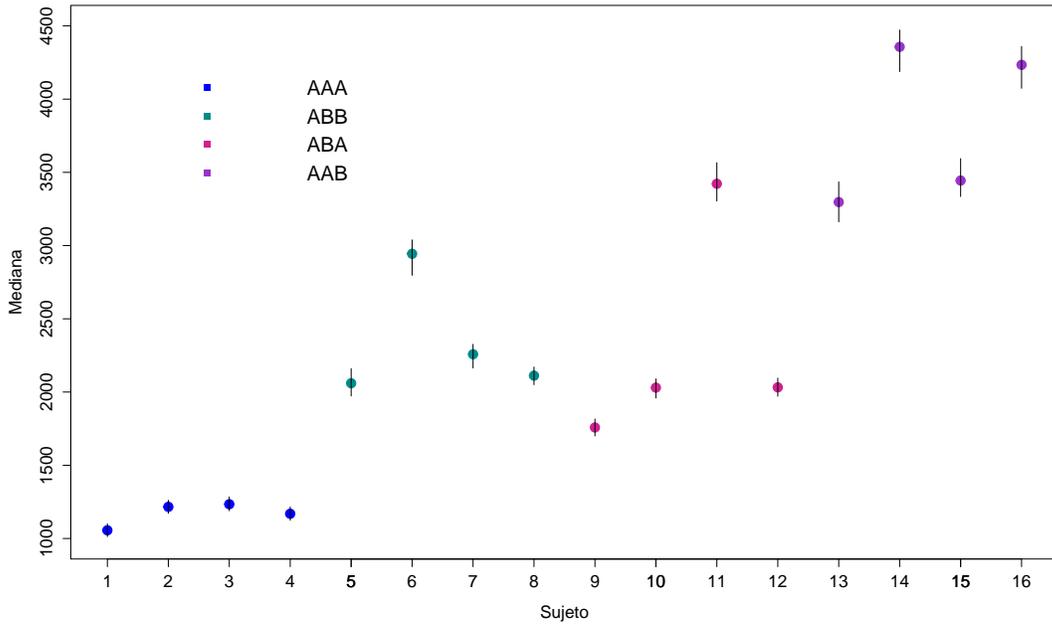


Figura 13: Medianas por sujeto con intervalo de confianza al 95 %.

Tabla 8: Prueba log-rank por grupo.

Grupo	Chisq	g.l.	p
AAA	189	3	<0.00001
ABB	2611	3	<0.00001
ABA	3571	3	<0.00001
AAB	1277	3	<0.00001

mismo resultado, es decir, no hay evidencia para decir que las curvas de un mismo grupo son similares.

Aunque resultó informativo analizar las curvas de la estructura reforzada, esto no nos dijo nada de cómo se comportaron las demás estructuras que no fueron reforzadas. Por ello se decidió realizar las curvas de supervivencia de las demás estructuras, que aunque no fueron reforzadas también se registraron. En la figura 14 se muestran 4 gráficas, cada

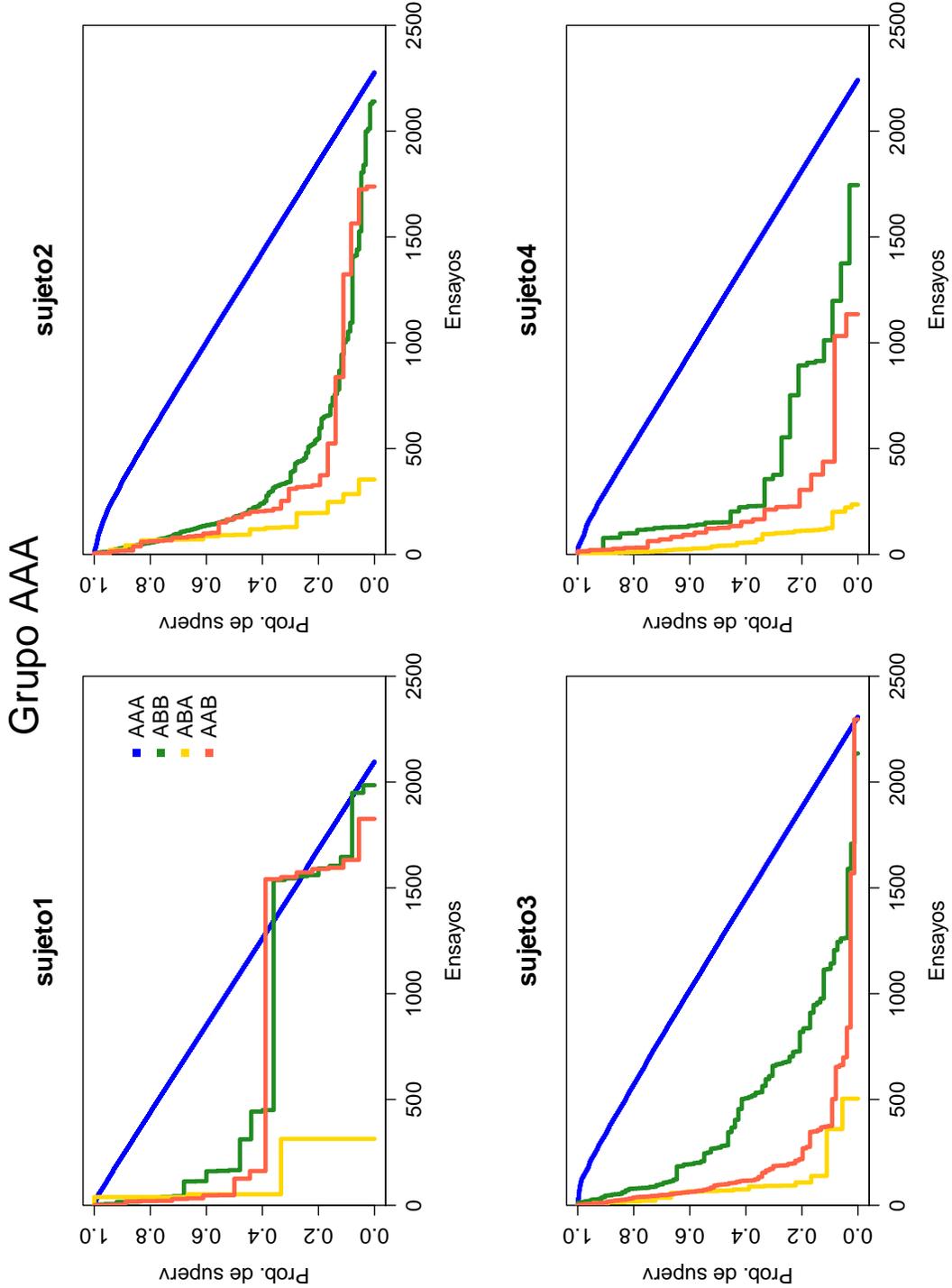


Figura 14: Curvas Kaplan-Meier para cada conducta por rata del grupo AAA.

una correspondiente a un individuo del grupo AAA, en donde se pueden observar las curvas de supervivencia de las 4 posibles estructuras de comportamiento que podían realizar los sujetos.

Resulta interesante observar en estas gráficas del grupo AAA cómo es que las otras posibles estructuras fueron ocurriendo a lo largo de los ensayos. Vemos que en todas las ratas, la estructura ABA (curva amarilla) fue la primera en dejar de ocurrir. De hecho, después de los 500 ensayos no volvió a ocurrir en ninguna de las ratas. Por otro lado las estructuras ABB y AAB siguieron siendo ejecutadas, aunque con mucha menor frecuencia que la estructura reforzada AAA. Aunque las curvas de supervivencia para ABB y ABA en estas ratas se traslaparon, parecería que la estructura ABB tardó más en dejar de ser ejecutada que AAB.

Estos datos nos pueden servir para poder generar un gradiente de similitud entre las estructuras. Resulta interesante que al reforzar AAA, estructura que no requiere de ninguna alternación entre las palancas, la estructura menos ejecutada sea la que requiere del mayor número de alternaciones ABA. Mientras que las estructuras que requieren una sola alternación ABB y AAB quedaron en medio en cuanto a su ejecución. Es muy interesante que las ratas sean sensibles a este gradiente en la estructura y que ello se vea reflejado en su distribución conductual.

Todas las ratas de este grupo mostraron un patrón muy similar en el acomodo de sus curvas de supervivencia. No obstante, cabe mencionar que la rata 1 muestra un patrón diferente a las demás ratas alrededor del ensayo 1,500. Esto probablemente se debió a un error al correr el experimento, en el que pudo haber un error en el programa de reforzamiento o pudo pasar que el dispensador del reforzador no funcionara en esa sesión.

En la figura 15 se muestran las curvas de Kaplan-Meier para todas las conductas de las ratas del grupo ABB. Es posible observar que la diferenciación en las curvas de supervivencia de todas las estructuras no se dio de la misma manera que en el grupo AAA. En este caso parecería que las ratas siguieron ejecutando todas las estructuras prácticamente durante toda la fase experimental. En las ratas 6, 7 y 8 la estructura reforzada ABB (curva verde) parece tener una ejecución por arriba de las demás estructuras, sin embargo, aún

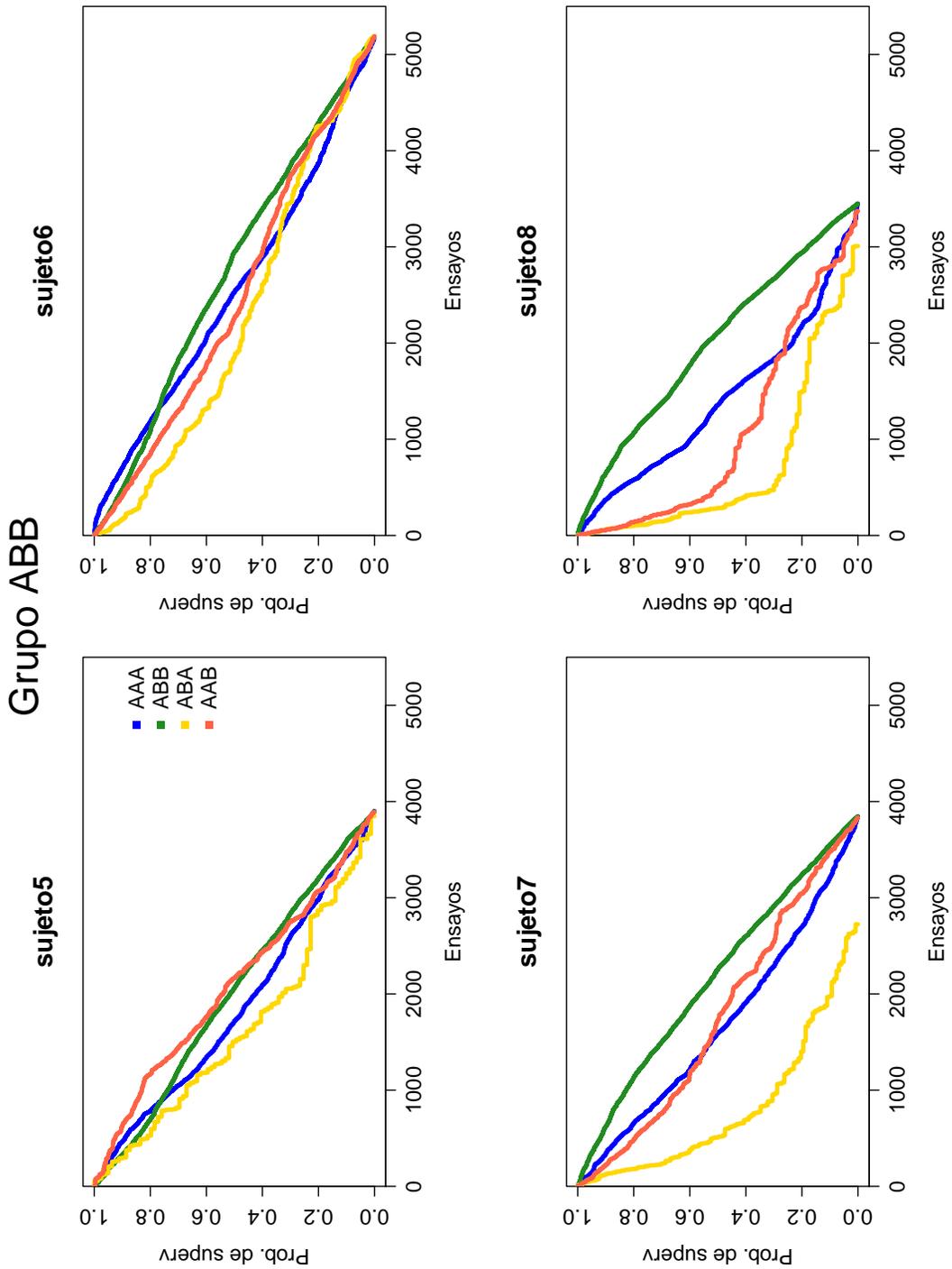


Figura 15: Curvas Kaplan-Meier para cada conducta por rata del grupo ABB.

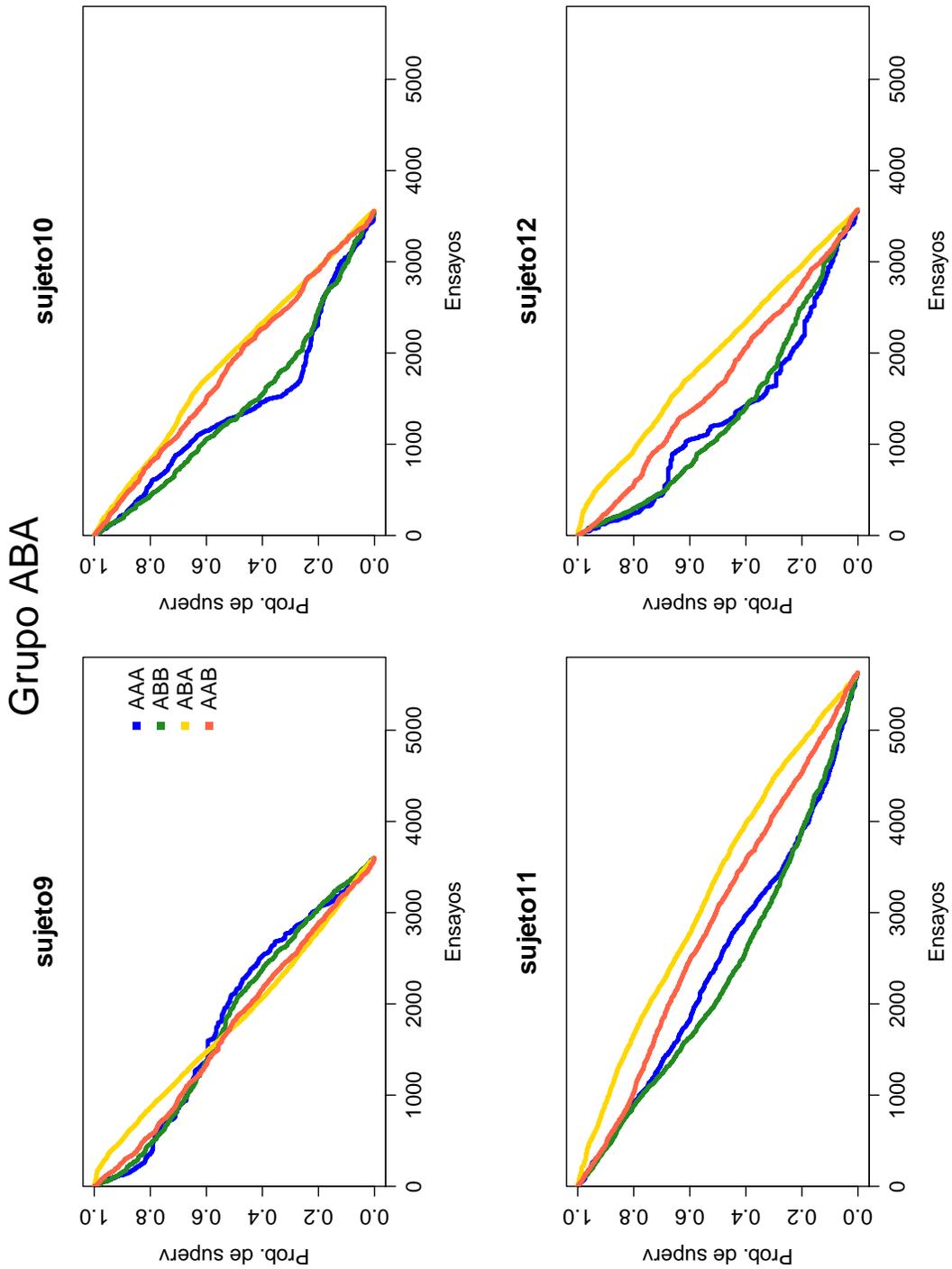


Figura 16: Curvas Kaplan-Meier para cada conducta por rata del grupo ABA.

así podemos ver un traslape. En todos los sujetos la estructura ABA fue la menos ejecutada y su curva de supervivencia está por debajo de todas las demás conductas. El patrón de las demás estructuras no es tan claro como en el grupo anterior, ya que tanto AAA como AAB se cruzan entre ellas y con las demás curvas.

En la figura 16 se muestran las curvas para todas las estructuras de las ratas del grupo ABA. Al igual que en el grupo ABB todas las posibles estructuras fueron ejecutadas hasta el final del experimento. Sin embargo, tres de las ratas, la 10, 11 y 12, parecen haber diferenciado su conducta de manera que las curvas de las estructuras ABA y AAB están por arriba de las curvas de las estructuras AAA y ABB. En estas tres ratas la estructura reforzada (línea amarilla) se encuentra por arriba de las demás en las tres ratas mencionadas, diferenciándose de las demás, al menos en este análisis. Para el otro sujeto, la rata 9 vemos que las curvas de todas las estructuras están muy cercanas las unas de las otras.

En la figura 17 vemos las gráficas por sujeto del grupo AAB. En este caso es todavía más complicado hacer alguna inferencia sobre la distribución de la conducta, ya que todas las estructuras fueron realizadas hasta el final de la fase experimental y se observa mucho traslape entre las curvas. Asimismo, la estructura reforzada no fue la más frecuente.

Aunque parecería no haber una clara diferenciación de la conducta reforzada, este análisis lo que nos indica es que los animales siguieron realizando prácticamente todas las estructuras durante toda la fase experimental, pero resulta complicado diferenciar cómo se distribuyeron. Aunque una manera de definir aprendizaje es por la diferenciación de la conducta reforzada, en este caso nos interesaba el patrón completo de la ejecución de todas las estructuras como una unidad mucho más general.

Este análisis nos permite tener un estudio fino de cómo fueron ocurriendo las posibles conductas ensayo a ensayo, de manera que fue posible observar que hubo mucha variabilidad en la ejecución de las ratas. Tanto en la manera en que fueron ejecutando las diferentes estructuras a lo largo de los ensayos, como en el número de ensayos que cada rata necesitó para terminar su fase experimental.

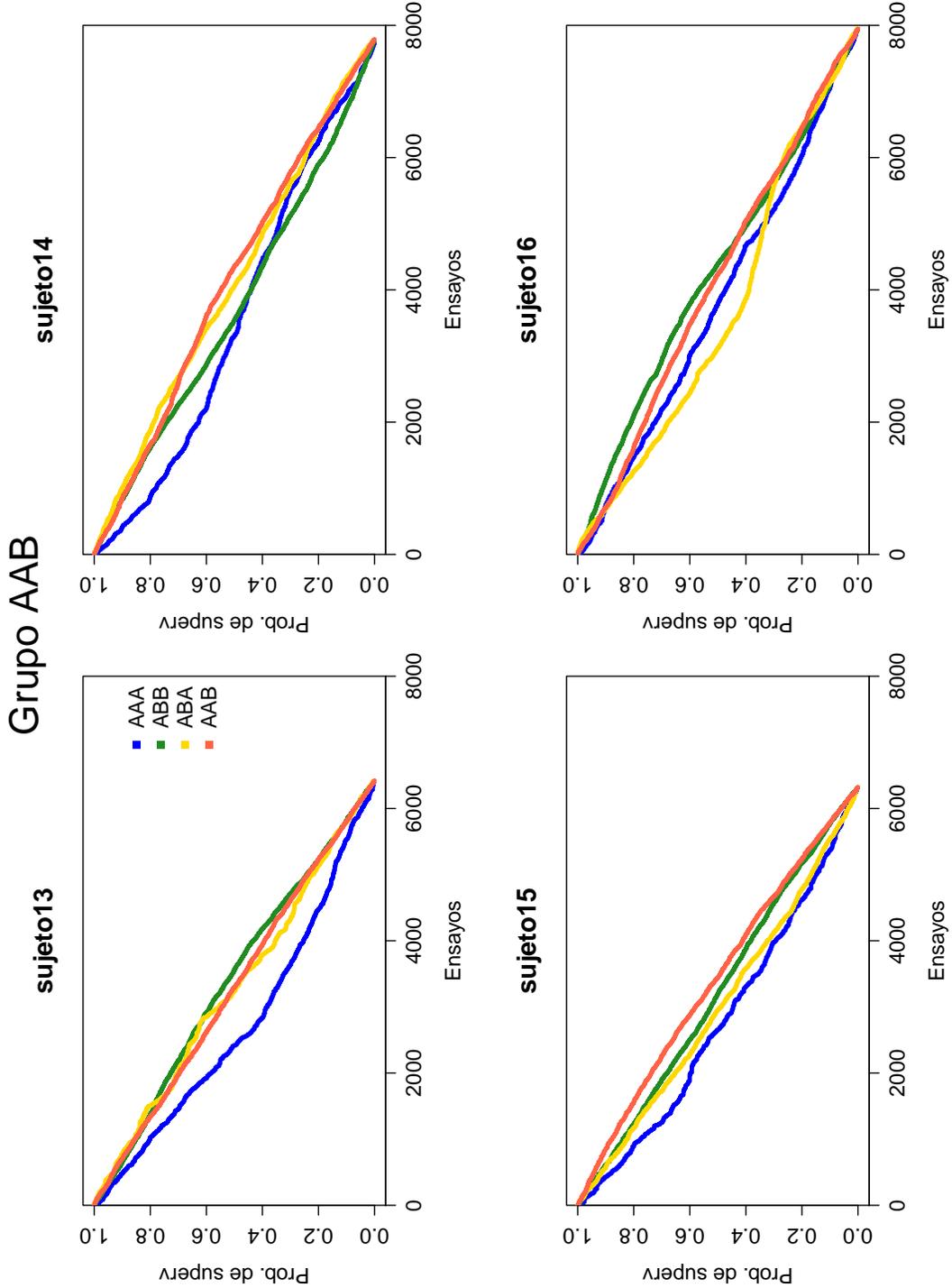


Figura 17: Curvas Kaplan-Meier para cada conducta por rata del grupo AAB.

6.3. Ajuste de un modelo log-lineal

Dada la naturaleza categórica de los datos, se decidió ajustar un modelo log-lineal a una tabla multivariada con las variables grupo, ejecución y sesión. La variable grupo hizo referencia a la pertenencia a alguno de los grupos AAA, ABB, ABA y AAB. La variable ejecución hizo referencia a cada una de las posibles estructuras ejecutables por las ratas; y la variable sesión se refirió a 5 sesiones que quedaron definidas como: la primera sesión y las sesiones en que los sujetos acumularon 500, 1,000, 1,500 y 2,000 reforzadores. La variable a modelar fue la frecuencia con que cada estructura fue ejecutada en cada grupo y en cada sesión considerada. Una vez más el análisis se corrió en el programa R, se usaron las funciones `glm` y `glm.nb` para ajustar los modelos. Inicialmente, todas las variables fueron consideradas categóricas, sin embargo, en modelos posteriores se consideró más adecuado meter a la variable sesión como continua.

Se empezó por construir una tabla de contingencia multivariada con las variables grupo, ejecución y sesión. La tabla resultante se muestra descompuesta por sesión en la figura 18. Los marginales por renglón de la tabla de cada sesión están fijos alrededor de 600, ya que cada rata tenía 150 ensayos de los cuales podía distribuir su conducta entre las 4 posibles estructuras. La suma marginal no siempre fue exactamente 600 porque algunos sujetos no siempre realizaban 150 ensayos en una sesión, aunque se les dejara mucho tiempo. Dado que fueron 4 ratas por grupo y su ejecución por estructura se sumó, la suma por renglón debería de ser alrededor de 600 conductas.

Estas tablas nos permiten observar cómo fueron cambiando cada 500 reforzadores las frecuencias con que los grupos ejecutaron cada una de las estructuras. Esto queda más claro en la figura 19 en donde se muestra el *mosaicplot*, el cual nos da una representación gráfica de la tabla multivariada en cuestión. Cada bloque “columna” de esta gráfica representa un grupo, dentro de los cuales está cada posible estructura. Los bloques “renglón” representan cada sesión. Así podemos observar claramente cómo fue cambiando la distribución de cada grupo a lo largo del experimento. Es fácil notar que en la primera sesión todos los grupos empezaron realizando todas las posibles estructuras con frecuencias parecidas. Aún el grupo AAA, en donde desde la primera sesión vemos que la estructura

	AAA	ABA	AAB	ABB
AAA	330	71	93	114
ABA	122	122	157	187
AAB	143	101	135	189
ABB	120	70	135	220

	AAA	ABA	AAB	ABB
AAA	582	1	2	17
ABA	73	320	100	107
AAB	75	170	138	162
ABB	229	19	75	277

(a) sesión 1 y sesión 2.

	AAA	ABA	AAB	ABB
AAA	581	0	0	9
ABA	61	365	114	60
AAB	70	120	203	207
ABB	196	16	77	311

	AAA	ABA	AAB	ABB
AAA	593	0	1	8
ABA	39	354	122	85
AAB	69	113	188	230
ABB	146	9	86	359

(b) sesión 3 y sesión 4.

	AAA	ABA	AAB	ABB
AAA	596	0	1	4
ABA	36	380	95	89
AAB	49	155	187	209
ABB	101	9	73	417

(c) sesión 5.

Figura 18: **Tablas de contingencia para las 5 sesiones consideradas.**

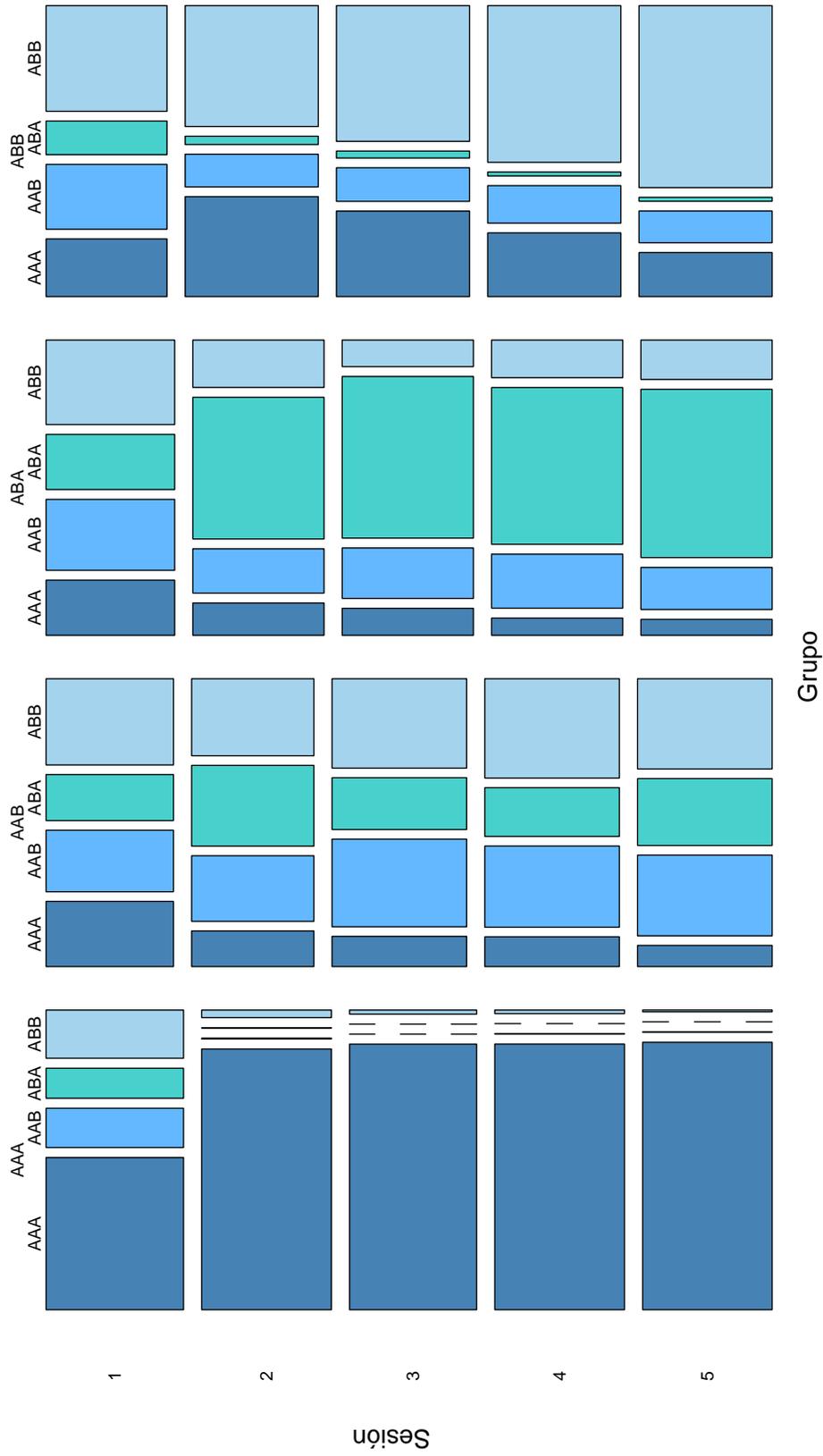


Figura 19: Mosaicplot.

reforzada es dominante, en esta primera sesión los sujetos todavía realizaron las otras 3 posibles estructuras. Entonces parece que los animales empezaron probando diferentes conductas, y posteriormente fueron ordenando su conducta de acuerdo a las contingencias programadas.

En las siguientes sesiones es posible observar que para los grupos AAA, ABA y ABB, la estructura reforzada fue incrementando y las demás estructuras fueron disminuyendo en su ejecución. En cuanto al grupo AAA la estructura reforzada dominó básicamente toda la conducta de las ratas. En las estructuras ABA y ABB, aunque la reforzada sí se volvió la más frecuente, otras estructuras siguieron siendo ejecutadas aún al final del experimento.

Finalmente, en la ejecución del grupo AAB se nota un cambio de la primera sesión con respecto a las demás, pero en este caso la secuencia reforzada no es la más realizada por el grupo. La estructura ABB es igual o más frecuente que la estructura AAB reforzada. Asimismo, la estructura ABA para este grupo también fue realizada bastante a lo largo de las sesiones. No podría decirse que las ratas no aprenden nada ya que su patrón conductual parece mantenerse constante y las ratas no varían entre las 4 posibles estructuras de manera completamente azarosa. Se observa una distribución estable entre las estructuras. Sin embargo, es claro que la ejecución de las secuencias con estructura AAB es algo que les resultó a los sujetos muy complicado y parece ser que realizaron el cambio de palanca antes de tiempo y por ello ABB es muy frecuente.

Una ejecución casi perfecta implicaría que la diagonal de las tablas de la figura 18 fuera aumentando en frecuencia con el paso de las sesiones, y que lo que está alrededor de la diagonal fuera disminuyendo considerablemente. No obstante, tanto en las tablas como en el *mosaicplot* es claro que aunque la estructura reforzada sí aumenta en la mayoría de los grupos, las demás estructuras siguen siendo ejecutadas. Un objetivo de este modelo fue probar que la ejecución de las estructuras se da en parte por un aprendizaje de la conducta reforzada, pero que también existe otro patrón en donde las estructuras no reforzadas se ejecutan de acuerdo a la hipótesis de contigüidad. Esta hipótesis sugiere que los animales asignan crédito a diversos componentes de su conducta de acuerdo con su cercanía con el reforzador. De manera que la respuesta más cercana a la entrega del reforzador (i.e.

la última en ser ejecutada), en teoría debería de tener más fuerza que la más lejana al reforzador (i.e. la primera en ser ejecutada).

Entonces para contrastar si estaba ocurriendo únicamente un proceso a nivel de la estructura reforzada o si la contigüidad con el reforzador estaba influyendo en la ejecución de las demás estructuras, se generaron dos variables que dividieron a las tablas para cada sesión como se muestra en la figura 20.

	AAA	ABA	AAB	ABB
AAA	2	0	1	0
ABA	0	2	0	1
AAB	0	1	2	1
ABB	1	0	1	2

(a) Variable de hipótesis de contigüidad.

	AAA	ABA	AAB	ABB
AAA	2	0	0	0
ABA	0	2	0	0
AAB	0	0	2	0
ABB	0	0	0	2

(b) Variable ejecución perfecta.

Figura 20: **Variables para poner a prueba diferentes ejecuciones.**

En la tabla (a) vemos que se generó una variable categórica de tres niveles a la cual se le asignó arbitrariamente el valor de 2 a la diagonal, el valor de 1 a lo que se considera sería la ejecución por la hipótesis de contigüidad y 0 a lo demás. Podríamos decir que la categoría 2 representa la ejecución de la estructura reforzada, la categoría 1 la de ejecución de errores por contigüidad y la categoría 0 la ejecución de errores no relacionados con contigüidad. La tabla (b) muestra la variable que representaría una ejecución perfecta, separando únicamente entre la diagonal y el resto de las celdas. Se esperaría que el peso de la categoría de la diagonal fuera mayor que el de las celdas de contigüidad, las cuales a su vez mayores a las celdas con 0.

Para poner a prueba esto se inició ajustando dos modelos con las variables grupo, ejecución, sesión y la variable hipotética correspondiente a cada hipótesis. Dado que no son modelos anidados, se compararon los valores del AIC. Para el modelo con la hipótesis de contigüidad el valor del AIC fue de 3364.75, mientras que para el modelo con la hipótesis de ejecución perfecta el valor fue de 4266.35. De un modelo a otro sólo hay un grado de libertad de diferencia. Asimismo, se obtuvo el valor de la pseudo R para cada modelo. Para el modelo de contigüidad fue de 0.69 mientras que para el otro modelo fue de 0.6. Estos resultados nos llevaron a quedarnos con la variable que representa el efecto de contigüidad en el modelo.

Todos los modelos mencionados a continuación se corrieron asumiendo la familia Poisson. Se comenzó ajustando el modelo que sólo tenía los efectos principales, hasta llegar a varios modelos con diferentes interacciones. En la tabla 9 se muestra el AIC, devianza, grados de libertad y el valor de la pseudo R de cada modelo corrido. La selección de las variables e interacciones que quedaron en el modelo estuvo guiada por los cambios en la devianza entre modelos anidados.

Tabla 9: Comparación de modelos.

Modelos	Términos	AIC	Devianza	g.l	pseudo R
Modelo 1	grupo, ejecución, sesión y contigüidad	3364.8	2860.3	67	0.69
Modelo 2	grupo, contigüidad y sesión*ejecución	3222.5	2694.1	55	0.70
Modelo 3	grupo, ejecución y sesión*contigüidad	2673	2152.5	59	0.75
Modelo 4	sesión, contigüidad y grupo*ejecución	2041.6	1523.1	60	0.81
Modelo 5	grupo*ejecución y sesión*contigüidad	1349.8	815.32	52	0.88

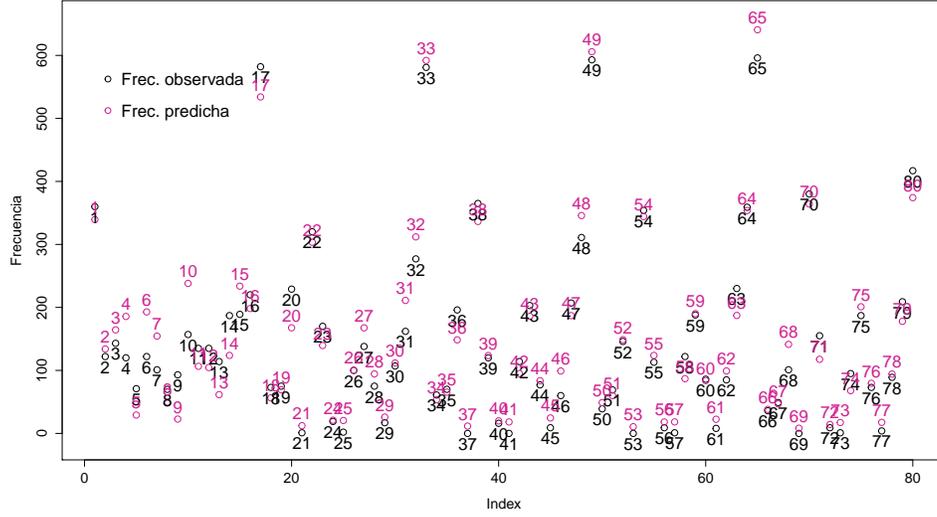
La devianza nula fue de 10,199.18 con 79 grados de libertad. Vemos que todos los modelos reducen mucho la devianza nula pasando de 10,000 a 2,000 o menos. Cuando se corrió el primero modelo sin interacciones la variable sesión no fue significativa, sin embargo, se planteó que podría ser que lo fuera en una interacción. Al incluir la interacción entre sesión y ejecución en el modelo 2, no hubo un aumento significativo en la pseudo R ni una disminución grande en la devianza, considerando que se perdieron muchos grados

de libertad. Entonces se probó otra interacción entre sesión y la variable contigüidad, suponiendo que el proceso de aprendizaje y de ejecución de las estructuras no reforzadas fue cambiando a lo largo de las sesiones. En este caso se encontró que el agregar dicha interacción reducía significativamente la devianza. En el modelo 4 se agregó al modelo 1 la interacción grupo con ejecución, la cual resultó muy significativa ya que redujo considerablemente la devianza con respecto al modelo 1. Finalmente se decidió correr un modelo con las dos interacciones que más redujeron la devianza. El valor de la devianza y AIC de este quinto modelo fueron los más pequeños de todos los modelos. Asimismo, su valor de pseudo R fue el más grande, explicando aproximadamente el 88 % de la varianza. Ya no se agregaron más términos de interacción dobles o triples en busca de un modelo parsimonioso.

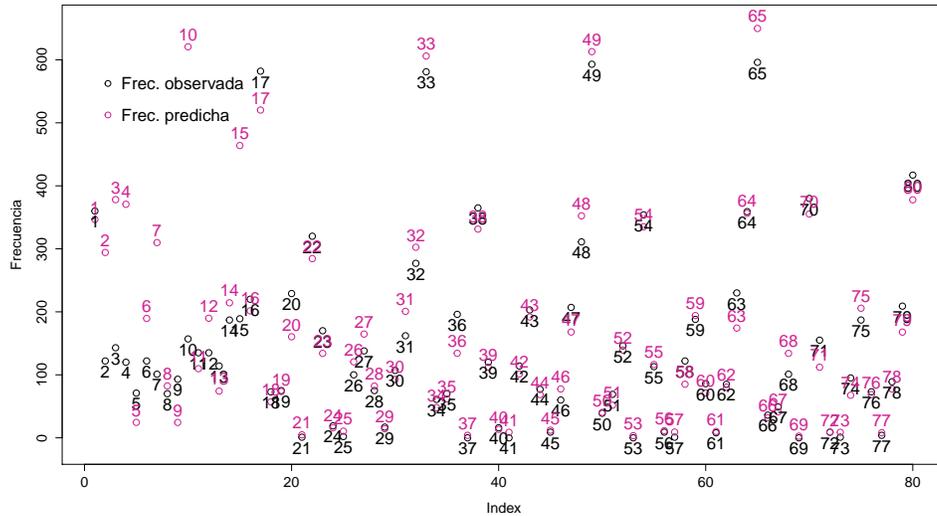
Aunque el modelo parece bueno en cuanto a la pseudoR, su devianza es mucho más grande que sus grados de libertad. Esto podría deberse a la presencia de sobredispersión. Entonces se realizó una prueba para comprobar si en efecto había sobredispersión, en donde la hipótesis nula fue que no existía sobredispersión. La prueba resultó altamente significativa, con una $p = 0.0004$, lo que llevó a rechazar la hipótesis de que no existe sobredispersión. Entonces se corrió un sexto modelo con los mismos términos que el modelo 5, pero ahora asumiendo una distribución binomial negativa para permitir mayor variabilidad. En este caso se obtuvo una devianza de 102.56 con 52 grados de libertad y un AIC de 868.69.

Aunque con la binomial negativa la devianza es solamente dos veces los grados de libertad, es decir, hubo una reducción importante en la devianza al cambiar de distribución, los valores ajustados por este modelo no predicen tan bien los valores observados como el modelo Poisson. En la figura 21 se muestra una gráfica para cada modelo con los valores predichos en morado y los valores observados en negro.

Es posible observar que ambos modelos tienen problemas en predecir las frecuencias de la primera sesión. Sin embargo, el modelo 6, que asume la distribución binomial negativa, predice frecuencias muy grandes en comparación con las observadas para la primera sesión. Por ejemplo, para el valor 10 el modelo con binomial negativa predice una frecuencia de



(a) Observado vs predicho modelo 5.



(b) Observado vs predicho modelo 6.

Figura 21: **Comparación de los valores predichos vs los observados.**

600, cuando el valor observado fue menor a 200. De los demás valores ambos modelos parecen predecir bastante bien las frecuencias, aunque en la inspección visual parece que las frecuencias predichas del modelo 6 están más cercanas a los datos observados. Asimismo, como ya se mencionó la devianza del modelo 6 es menor que la del modelo

5. Hasta cierto punto tiene sentido que ambos modelos fallen en predecir el inicio del experimento, ya que al principio los animales no han organizado su conducta de acuerdo a la entrega del reforzador, esto ocurre en sesiones posteriores.

Dado que se está suponiendo que en la primera sesión los animales estaban variando azarosamente y dado que las variables fueron pensadas para cuando la conducta de los sujetos ya se ha estabilizado en patrones predecibles, una opción es quitar los datos de esa primera sesión. Se procedió a reducir la base de datos quedando únicamente las tablas de la sesión 2 a la 5. Esto redujo mucho la devianza del modelo 5 con familia Poisson a 147.41 con 39 grados de libertad y un AIC de 568.09. Dado que en este caso ya no se tenía la primera sesión, se decidió correr el modelo con la variable sesión como continua, ya que los cambios entre sesiones a partir de la segunda sesión parecen ser constantes. Se corrió el modelo con sesión como continua encontrando una devianza de 153.36 con 45 grados de libertad y AIC: 562.04. La diferencia entre las devianzas es pequeña y al dejar la variable como continua ganamos varios grados de libertad. Por ello se decidió quedarse con sesión como una variable continua. Se volvió a hacer la prueba de sobredispersión en este último modelo y siguió saliendo significativa con una $p = 0.0013$.

Entonces con la nueva base de datos recortada se volvió a correr el modelo asumiendo una distribución binomial negativa. Lo cual dio una devianza de 59.835 con 45 grados de libertad y AIC: 524.74. El p -value de la devianza fue de 0.07. Si consideramos un criterio arbitrario de $p > 0.05$ para aceptar un modelo, podríamos considerar que este último modelo resulta ser el mejor. En la tabla 10 se muestran los resultados de dicho modelo. En esta tabla es posible ver que todos los coeficientes fueron significativos, tanto los efectos principales como las interacciones propuestas. Dado que las interacciones fueron significativas rechazamos el modelo de independencia.

Dado que el grupo y la estructura que quedaron como referencia fueron el grupo AAA y la estructura AAA las comparaciones se harán respecto a ellas. Asimismo, para la variable contigüidad la categoría de referencia fue lo codificado como 0 de la tabla (a) de la figura 20. En cuanto a la variable sesión, como fue considerada continua, los resultados se interpretarán como el cambio en los momios dado el cambio de una sesión

Tabla 10: Resultados del modelo final con distribución binomial negativa.

Coeficientes	Estimado	Error Est.	z	Pr(> z)
(Intercepto)	3.73602	0.18162	20.571	< 2e-16 ***
GrupoAAB	0.96764	0.15160	6.383	1.74e-10 ***
GrupoABA	0.72745	0.13677	5.319	1.04e-07 ***
GrupoABB	0.42119	0.09286	4.536	5.73e-06 ***
EjecucionAAB	-4.69382	0.51311	-9.148	< 2e-16 ***
EjecucionABA	-4.60776	1.00989	-4.563	5.05e-06 ***
EjecucionABB	-0.97496	0.07857	-12.408	< 2e-16 ***
Contigüidad1	1.13772	0.19436	5.854	4.81e-09 ***
Contigüidad2	2.40994	0.20911	11.525	< 2e-16 ***
Sesión	-0.15127	0.03885	-3.893	9.89e-05 ***
GrupoAAB:EjecucionAAB	2.53220	0.55011	4.603	4.16e-06 ***
GrupoABA:EjecucionAAB	5.44087	0.53101	10.246	< 2e-16 ***
GrupoABB:EjecucionAAB	3.93421	0.51774	7.599	2.99e-14 ***
GrupoAAB:EjecucionABA	3.88512	1.02920	3.775	0.000160 ***
GrupoABA:EjecucionABA	3.37207	1.03343	3.263	0.001102 **
GrupoABB:EjecucionABA	3.54628	1.01712	3.487	0.000489 ***
GrupoAAB:EjecucionABB	0.62802	0.15889	3.953	7.73e-05 ***
Contigüidad1:Sesión	0.09935	0.04822	2.061	0.039342 *
Contigüidad2:Sesión	0.21761	0.04781	4.552	5.32e-06 ***
Devianza nula: 3862.251 con 63 grados de libertad				
Devianza residual: 59.835 con 45 grados de libertad				
AIC: 524.74				
código de significancia: 0 '***' 0.001 '**' 0.01 '*'				

a otra, asumiendo que los cambios son iguales entre sesiones.

Para la variable de ejecución vemos que todos los coeficientes son negativos, es decir, en términos generales los momios de ejecutar cualquier estructura con respecto a

AAA son menores a 1. Esto nos indica que lo que más ejecutan los sujetos es AAA, ya que es una estructura muy sencilla. En cuanto a la variable de contigüidad vemos que los momios de ejecutar los errores por contigüidad (codificado como contigüidad1) son $\exp(1.1377) = 3.1195$. Lo que podemos interpretar como que los momios de cometer este tipo de error son 3 veces más posibles que los de cometer los otros errores no relacionados a la hipótesis de contigüidad (contigüidad0). Asimismo, con respecto a la diagonal, vemos que las posibilidades de ejecutar la estructura correcta son $\exp(2.4099) = 11.32$ veces las de cometer un error no relacionado con contigüidad.

En cuanto a las interacciones, éstas se interpretaron como razones de momios. La interacción contigüidad*sesión resultó significativa por lo que los cambios en las estructuras reforzadas y los errores por contigüidad fueron cambiando de sesión a sesión. Entonces el coeficiente contigüidad2*sesión nos indica que los momios de la diagonal respecto a los errores no relacionados a contigüidad son $\exp(0.2176) = 1.24$ al cambiar de una sesión a otra. Es decir, hay un incremento del 24 % en estos momios de una sesión a otra. Por otro lado, la razón de los momios de los errores de contigüidad con respecto a los errores de no contigüidad al aumentar en una sesión, son $\exp(0.099) = 1.10$. Considerando que las sesiones se eligieron por incrementos de 500 reforzadores acumulados, también podría decirse que cada 500 reforzadores acumulados las posibilidades de ejecutar los errores por contigüidad con respecto a los demás errores aumentan en 10 %.

La interacción significativa ente grupo y ejecución nos habla de que de acuerdo al grupo, la distribución de la frecuencia entre las diferentes estructuras es diferentes. Por ejemplo, los momios de ejecutar AAB con respecto a AAA para el grupo ABB son $\exp(3.93) = 50.90$ veces los momios estimados para el grupo AAA. Estas razones de momios son tan grandes porque se está comparando con el grupo AAA, el cual prácticamente no realizó ninguna conducta más que AAA. Por ello todos los términos de interacción son positivos.

En la figura 22 se muestran 4 gráficas, una por grupo, en las que se observan las frecuencias en el eje de las y , y las 4 sesiones consideradas en el eje de las x . Las frecuencias observadas para cada estructura fueron representadas por líneas continuas, mientras que

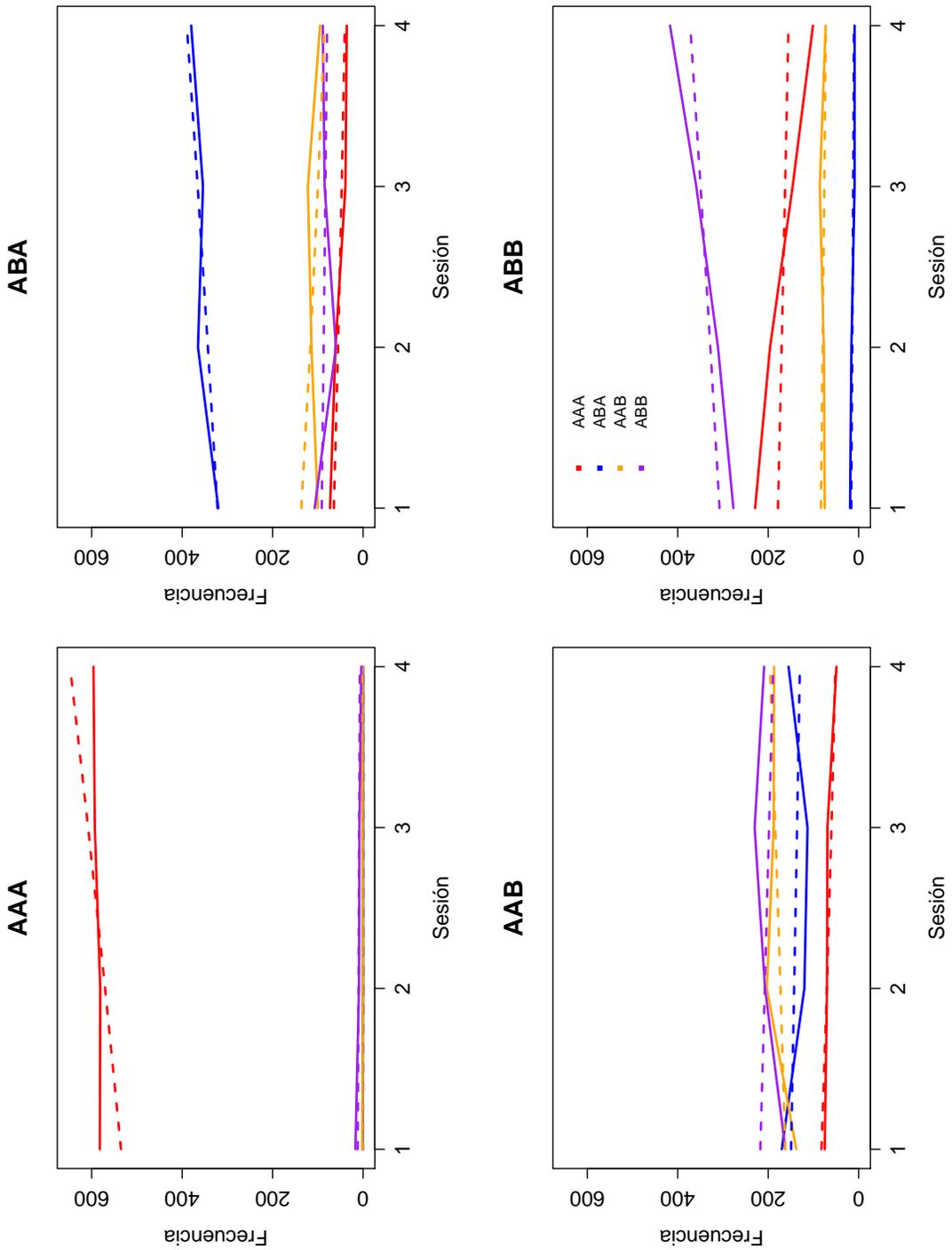


Figura 22: Valores ajustados y observados. Las líneas punteadas representan lo ajustado y las continuas lo observado.

las frecuencias ajustadas por el modelo por líneas punteadas. Al ver las gráficas parece que el ajuste es muy bueno, y que el modelo predice adecuadamente las frecuencias de las diferentes estructuras en cada grupo.

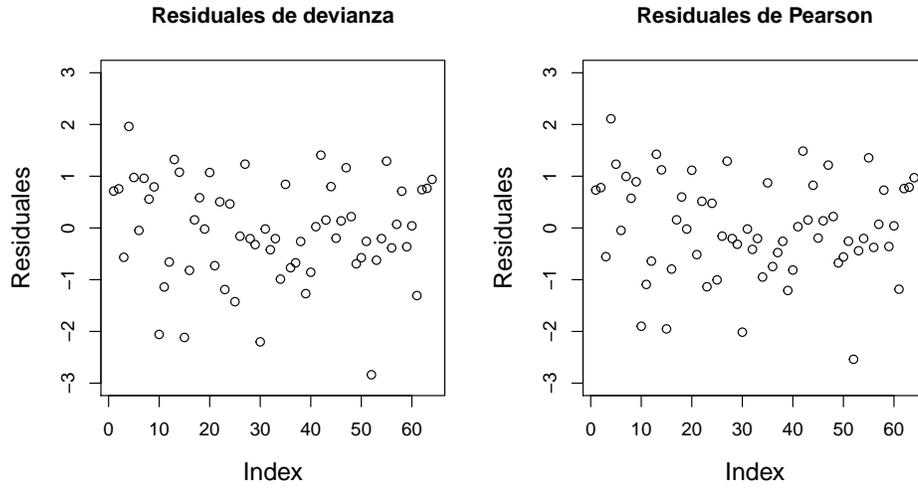
Asimismo en esta gráfica es claro que para los grupos AAA, ABA y ABB la estructura reforzada se diferenció de las demás estructuras con una mayor frecuencia. En el caso del grupo AAA la estructura reforzada fue prácticamente lo único que los sujetos hicieron. En el caso de los grupos ABA y ABB vemos que en ambos casos la conducta reforzada llegó a una frecuencia alrededor de 400 para cada grupo. Las demás estructuras siguieron siendo reforzadas, y el orden en que fueron ejecutadas difirió entre los grupos. Esto concuerda con la descripción de las curvas de supervivencia analizadas previamente, las cuales nos permitieron hacer inferencias parecidas. En cuanto al grupo AAB vemos que la estructura reforzada representada por la línea amarilla, está prácticamente encimada con la ejecución de la estructura ABB en morado. Sin embargo, las diferentes estructuras sí parecen tener un orden en la frecuencia con que fueron ejecutadas a lo largo del experimento.

Aunque ya tenemos el valor de la devianza en comparación con sus grados de libertad para analizar el ajuste de los datos, también es conveniente hacer el análisis de residuales. En la figura 23 se muestran en las gráficas (a) los residuales de Pearson y los de la devianza. En estas dos gráficas es posible observar que los residuales varían alrededor de 0 y la mayoría está entre 2 y -2, lo cual es bueno pues no parece haber algún residual que indique un ajuste muy malo en alguna celda. Asimismo, estos residuales no muestran algún patrón claro.

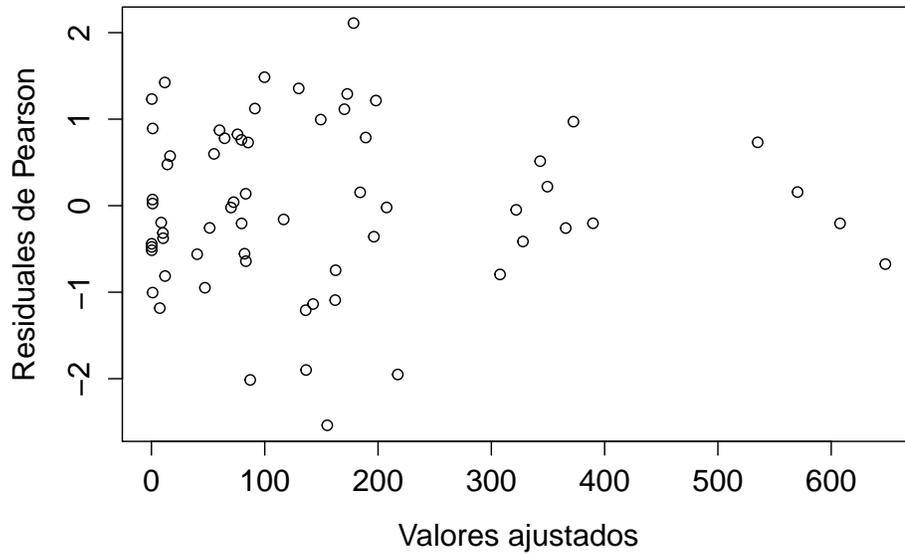
En la gráfica (b) de la figura 23 se muestra una gráfica en donde en el eje de las y están los residuales de Pearson y en el de las x los valores ajustados por el modelo. En esta gráfica se busca que no haya patrones. No obstante, en este caso hay menos datos hacia valores más grandes, y por ello mayor variabilidad en valores ajustados pequeños que en grandes.

El modelo propuesto parece ajustar bien y nos da mucha información acerca del aprendizaje de las ratas. No obstante, desde el análisis de supervivencia fue claro que había mucha variabilidad en cuanto a la ejecución de cada rata. Asimismo, la estructura jerárquica

de los datos puede representar un problema para el supuesto de independencia, es por ello que se llevó a cabo el análisis que se describirá en la siguiente sección.



(a) Residuales de Pearson y de devianza.



(b) Ajustados vs residuales de Pearson.

Figura 23: Gráficas del análisis de residuales.

6.4. Ajuste de una regresión logística multinivel

En esta sección se describirán los resultados del ajuste de un modelo de regresión logística multinivel a los datos. Se decidió realizar este análisis por la naturaleza longitudinal del estudio llevado a cabo. Se consideró una estructura jerárquica de los datos de dos niveles, en donde el nivel 2 fueron las 16 ratas y el nivel 1, las medidas repetidas ensayo a ensayo anidadas en cada rata. De manera que se consideraron 16 grupos, cada uno con diferente cantidad de medidas repetidas, ya que cada rata obtuvo los 2,000 reforzadores en diferentes tiempos. La variable dependiente fue un indicador dicotómico que en cada ensayo tomó el valor de 1 si la rata había ejecutado la estructura reforzada y 0 si había ejecutado cualquier otra estructura. De las variables independientes que se consideraron, el paso del tiempo fue la más importante y se midió con la variable continua sesión. Asimismo, se consideró como posible variable explicativa la conducta en el ensayo anterior, la cual se codificó en una variable que se llamó *status_1*. Esta última variable fue 1 si en el ensayo anterior se había realizado la conducta reforzada y 0 en otro caso. Finalmente, se tuvo la variable grupo para hacer referencia a la estructura reforzada en cada rata. Los modelos se corrieron en R usando la función *glmer* y la familia binomial para describir los datos. Todos los modelos que se mencionarán se corrieron asumiendo la función liga *logit*.

Primero veamos los datos reales. En la figura 24 se graficó por rata la proporción de respuestas correctas en cada sesión experimental. Los colores de las líneas corresponden a la estructura reforzada. Únicamente se muestran las curvas de aprendizaje de la estructura reforzada, las demás conductas no fueron consideradas en este análisis. Es posible observar que las ratas del grupo AAA aprenden en pocas sesiones y su conducta se estabiliza rápidamente. En cuanto al grupo ABB y ABA su conducta es más variable, pero vemos que tienen una proporción de respuestas correctas alrededor de 0.5 y 0.6. Finalmente, las ratas del grupo AAB son las que tienen una pendiente más horizontal, con una proporción de respuestas correctas alrededor de 0.3. Es posible observar que hubo mucha variabilidad en la ejecución de los sujetos, por lo que resulta prudente hacer un análisis de curvas de crecimiento para cada sujeto.

El análisis se inició ajustando el modelo más sencillo posible: con sólo el intercepto, el

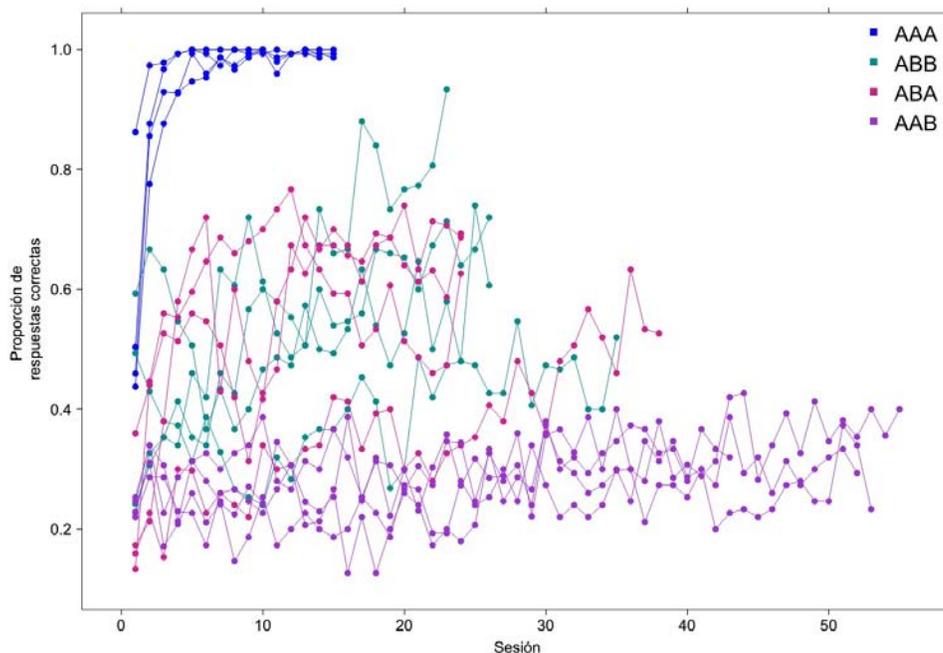


Figura 24: **Proporción de respuestas correctas por ensayo.**

cual se permitió fuera aleatorio. Este primer modelo quedó de la siguiente manera:

$$\text{logit}(\pi_{ij}) = 0.5517 + u_{0j} \quad (6.4.1)$$

La $\hat{\beta}_0$ se interpreta como el logaritmo de los momios de realizar la estructura reforzada en un sujeto promedio (con $u_{0j} = 0$). Si exponenciamos el valor de $\hat{\beta}_0$ obtenemos 1.73, es decir, en promedio es 73% más posible la ocurrencia de la estructura reforzada que la de las estructuras no reforzadas en este modelo. El intercepto para el j -ésimo sujeto está dado por $0.55 + u_{0j}$, en donde la varianza estimada para los interceptos fue de $\sigma_{u_0}^2 = 2.228$. En la figura 25 se muestra la gráfica de oruga, en la que se pueden observar las desviaciones de los interceptos por rata, ordenados del más chico al más grande y su intervalo de confianza al 95%.

En esta figura es posible ver que permitir variabilidad en el intercepto es necesario. Vemos que las ratas del grupo AAA (en azul fuerte) tienen un intercepto que está por arriba del promedio, mientras que las ratas del grupo AAB (en lila) están por abajo del promedio, es decir, al inicio estas ratas empiezan contestando correctamente por abajo del promedio. Por su parte los interceptos de las ratas de los grupos ABB y ABA están en

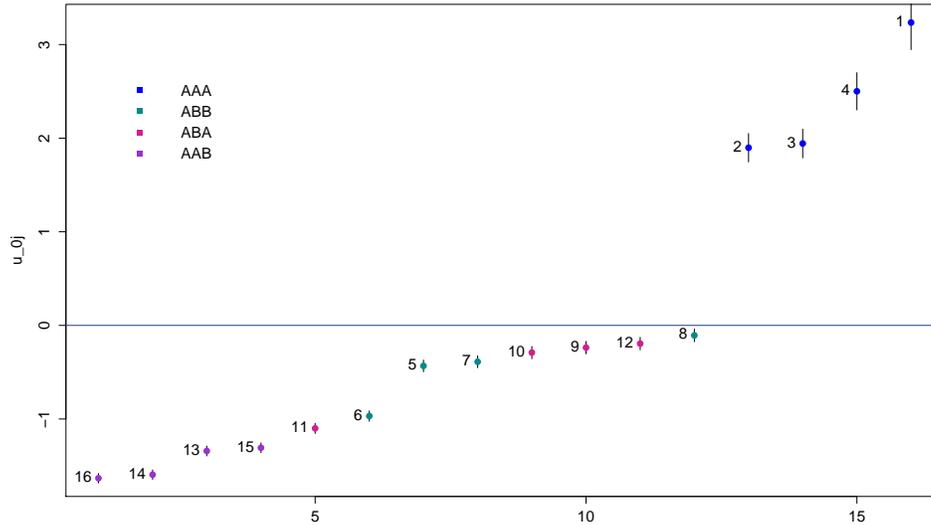


Figura 25: Gráfica de oruga.

medio, muy cercanos a lo que sería el promedio, sin embargo, ninguno de sus intervalos de confianza incluye al 0. En la tabla 11 vemos varios índices sobre el ajuste de este modelo. Es posible observar que la devianza del modelo es de 81,993.4 con 70,134 grados de libertad. El modelo tiene muchos grados de libertad ya que tenemos 70,136 observaciones y sólo ajustamos dos parámetros, el intercepto y su variabilidad. Con un modelo con tantas observaciones va a ser difícil que la devianza indique un buen ajuste.

Tabla 11: Índices de ajuste.

AIC	Devianza	Grados de libertad
81997.4	81993.4	70134

Para comprobar que añadir variabilidad en los interceptos es necesario, es decir que $\sigma_{u_0}^2 > 0$, se corrió el modelo con el intercepto fijo y se compararon las log-verosimilitudes de los dos modelos. Menos dos veces la diferencia de las log-verosimilitudes del modelo con y sin variabilidad en los interceptos, dio 15,001.77 con 1 grado de libertad, lo que da evidencia fuerte de que es necesario incluir en el modelo el intercepto como efecto

aleatorio.

Dado que nos interesaba la relación entre la variable dependiente y sesión, es decir, analizar cómo se dieron las curvas de crecimiento o en este caso de aprendizaje en los sujetos, el segundo modelo que se ajustó fue con sesión como efecto fijo. El modelo resultante fue:

$$\text{logit}(\pi_{ij}) = 0.25 + 0.019\text{sesión} + u_{0j} \quad (6.4.2)$$

Tabla 12: Resultados modelo 2.

Coeficientes (Error. Est)	Varianza de u_{0j}	AIC	Devianza	Grados de libertad
Intercepto: 0.25 (0.39)	2.6	81387.8	81381.8	70133
Sesión: 0.019(0.0007)*				

En la tabla 12 se muestran los resultados del modelo. El coeficiente de sesión resultó positivo y significativo, esto nos indica que la ocurrencia de la estructura reforzada aumentó con el tiempo. Es posible notar que al incluir la variable sesión como efecto fijo la varianza de los interceptos se mantuvo relativamente parecida al modelo con sólo el intercepto aleatorio. Al agregar sesión al modelo, hubo una disminución en la devianza y en el AIC, con la pérdida de sólo un grado de libertad. La introducción del paso del tiempo al modelo parece mejorarlo significativamente. Aunque cabe recordar que para datos desagregados la devianza no es un buen indicador.

En este modelo las líneas ajustadas para cada rata deberían de diferir en el intercepto por una cantidad \hat{u}_{0j} para la j -ésima rata, pero dado que la variable sesión fue introducida como un efecto fijo, la gráfica de los valores predichos para cada rata debería de ser un conjunto de líneas paralelas. Esta gráfica se muestra en la figura 26.

En esta figura vemos 16 rectas, una ajustada por sujeto. Los colores hacen alusión a la estructura reforzada. Esto nos permite ver que, de acuerdo a este modelo, las ratas del grupo AAA tienen un intercepto mucho mayor y son las que menos sesiones estuvieron en el experimento. Por otro lado las ratas del grupo ABA y ABB tienen un intercepto más bajo que AAA y vemos que las líneas ajustadas para estos dos grupos están muy

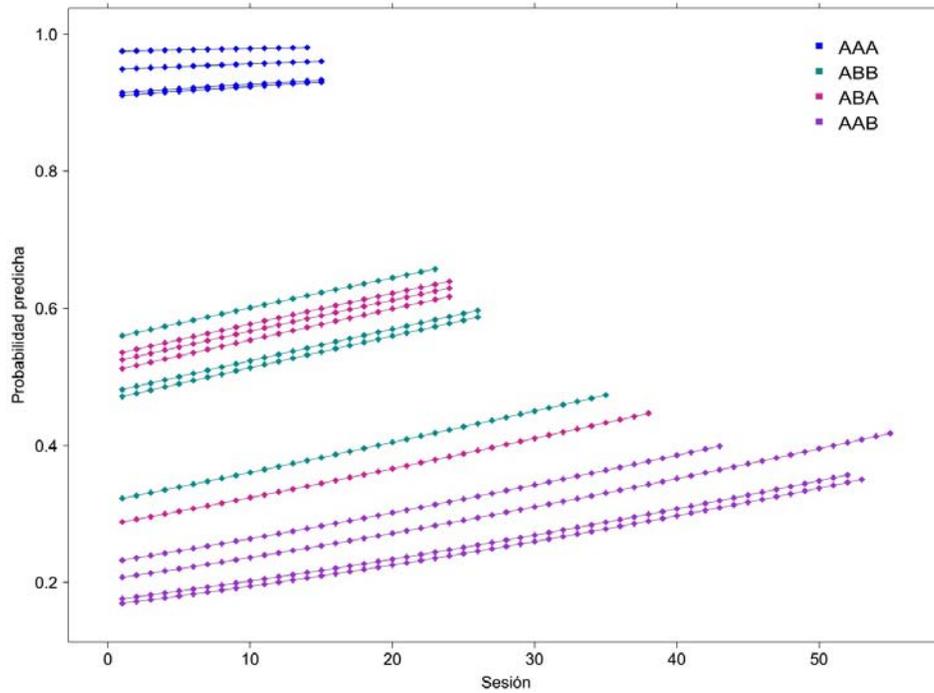


Figura 26: **Probabilidades predichas por el modelo 2.**

cerca. Sin embargo, como ya se había visto en los otros análisis, dos ratas de estos grupos tuvieron un comportamiento diferente, con interceptos por debajo de las demás ratas de su grupo y tardaron más en completar la fase experimental. Finalmente las ratas del grupo AAB son las que tienen el intercepto más bajo y son las que más tardan en terminar el experimento. Todo esto es consistente con los análisis previos. Esto concuerda a grandes rasgos con la figura 24, sin embargo, en los datos observados parece que las pendientes difieren por rata.

Entonces, el siguiente paso en este análisis fue permitir que tanto la pendiente de sesión como el intercepto fueran efectos aleatorios en el modelo, quedando el tercer modelo ajustado de la siguiente manera:

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 \text{sesión} \quad (6.4.3)$$

$$\beta_0 = -0.39 + u_{0j}$$

$$\beta_1 = 0.15 + u_{1j}$$

Cabe mencionar que en este tercer modelo se permitió que el intercepto y la pendiente covariaran. Los resultados del modelo se muestra en la tabla 13.

Tabla 13: Resultados modelo 3.

Efectos fijos				
Coefficientes	Estimado	Error Est.	z	Pr(> z)
Intercepto	-0.38572	0.22183	-1.739	0.08207
Sesión	0.14697	0.05542	2.652	0.00801 **
Efectos aleatorios				
Grupos	Coefficientes	Varianza	Desv. Est.	Corr
sujeto	Intercepto	0.83153	0.9119	
	Sesión	0.04896	0.2213	0.36
Devianza: 79825.3 con 70131 grados de libertad				
AIC:79881.1				

El coeficiente de sesión fue de nuevo positivo y altamente significativo. Si exponenciamos este coeficiente obtenemos $\exp(0.14697) = 1.1583$, lo que nos dice que al pasar de una sesión a otra el aumento en las posibilidades de ejecutar la estructura reforzada fue de 16 %. Asimismo, vemos que la varianza del intercepto disminuyó, de 2.6 a 0.83. Esto nos podría estar indicando que la varianza de los interceptos está relacionada con la varianza de las pendientes.

La correlación entre la pendiente y el intercepto fue de 0.36. Por el signo positivo de esta correlación podemos inferir que las ratas con un intercepto mayor mostraron un crecimiento o aprendizaje más rápido que las ratas que comenzaron con un intercepto menor. Al agregar la pendiente como efecto aleatorio, el AIC disminuyó de 81,387.8 para el modelo con sesión como efecto fijo a 79,881.1 para este modelo, es decir hubo una disminución de 1,506.7 y sólo perdimos dos grados de libertad.

En la figura 27 se muestran las probabilidades predichas para cada rata del modelo 3. Este modelo de curvas de crecimiento parece explicar bien los datos, sin embargo, son tan-

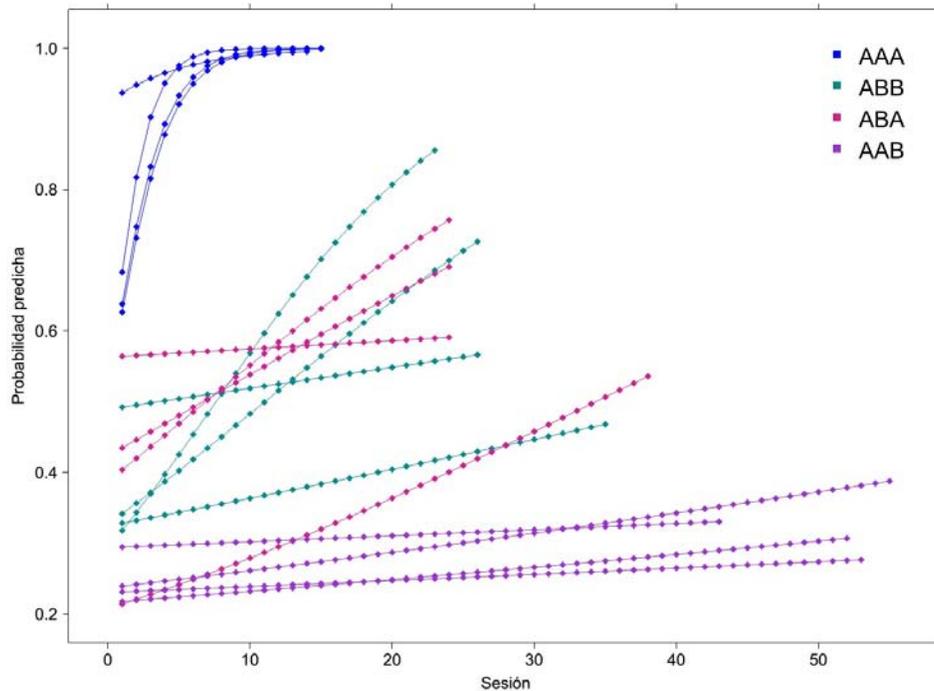


Figura 27: **Probabilidades predichas por el modelo 3.**

tos datos y están desagrupados que resulta difícil hacer la evaluación con la devianza y sus grados de libertad. Sin embargo, parece haber evidencia de que este modelo describe bien el proceso de aprendizaje general de las ratas. Si comparamos las proporciones observadas de la figura 24 con las ajustadas en la figura 27, vemos que las ratas mostraron mucha variabilidad, pero el patrón general de adquisición de la estructura reforzada parece estar correctamente capturado en las predicciones del modelo.

Se tenían otras variables explicativas, las cuales se probaron ajustando otros modelos que las incluyeran. En la tabla 14 se muestra un resumen de los modelos ajustados. Los modelos 1, 2 y 3 son los modelos que acaban de ser discutidos. Incluir un intercepto y pendiente aleatorios resultó en decrementos significativos en el AIC. El cuarto modelo que se probó incluyó a la variable status_1. Este modelo se mantuvo prácticamente igual al modelo 3. Esta variable no resultó significativa y si comparamos los valores del índice AIC, de hecho el modelo que incluye a status_1 tiene un AIC mayor que el modelo que no la incluye. Asimismo, la diferencia entre la devianza del modelo 3 y 4 es de apenas 0.5,

por lo que parece que no tiene sentido incluir esta variable.

Tabla 14: Comparación de modelos para datos desagregados.

Modelo	Efectos fijos	Efectos Aleatorios	AIC	Devianza	g.l.
Modelo 1	Intercepto: 0.55(0.31)	$\text{Var}(u_{0j})=2.23$	81997.4	81993.4	70134
Modelo 2	Intercepto: 0.25(0.38) Sesión: 0.018(0.0007)*	$\text{Var}(u_{0j})=2.6$	81387.8	81381.8	70133
Modelo 3	Intercepto: -0.39 (0.22) Sesión: 0.15 (0.05)*	$\text{Var}(u_{0j})=0.83153$ $\text{Var}(u_{1j})=0.04896$ $\text{Corr}(u_{0j},u_{1j})= 0.36$	79835.3	79825.3	70131
Modelo 4	Intercepto: -0.38 (0.23) Sesión: 0.15 (0.05)* status_1: -0.014 (0.018)	$\text{Var}(u_{0j})=0.83632$ $\text{Var}(u_{1j})=0.04916$ $\text{Corr}(u_{0j},u_{1j})= 0.36$	79836.8	79824.8	70130
Modelo 5	Intercepto: 2.58(0.43) Sesión: 0.14(0.05)* grupoABB:-3.68(0.46)* grupoABA: -3.63(0.46)* grupo AAB: -4.48(0.47)*	$\text{Var}(u_{0j})=1.62467$ $\text{Var}(u_{1j})=0.04603$ $\text{Corr}(u_{0j},u_{1j})= -0.95$	79816.1	79800.1	70128

En el quinto modelo se incluyó la variable grupo, en este caso los coeficientes fueron todos significativos y al comparar su devianza con la del modelo 3 vemos una disminución de 25.2 y en cuanto al AIC vemos que hay una disminución de 19.2 con una diferencia de 3 grados de libertad. La variable grupo parece ser importante en cierto grado, no obstante, la disminución en los índices de ajuste no es muy grande y al revisar las probabilidades predichas, resultó que éstas fueron muy parecidas a las del modelo 3, el cual no incluye la variable grupo. Al final, pareció que el modelo más sencillo que describe adecuadamente el patrón general de las curvas de aprendizaje de cada individuo es el modelo 3.

Dado que la única variable explicativa que se tenía a nivel 1 no resultó significativa, se decidió agrupar los datos de manera que se modelara la proporción de éxitos por sesión, lo que permitió hacer el análisis con las devianzas de manera más precisa. En este caso el

nivel 1 fueron las sesiones y el nivel 2 las ratas, de manera que seguimos con 16 grupos. Se volvieron a correr los modelos (con excepción del modelo 4) pero ahora con los datos agrupados y sus resultados se pueden observar en la tabla 15. Vemos que los valores estimados para los efectos fijos y aleatorios para los modelos con los datos agrupados y desagrupados son muy similares, con diferencias de décimas o milésimas. Evidentemente, los grados de libertad, los AIC y las devianzas son más pequeñas para los datos agrupados, ya que en este caso la base de datos tuvo una longitud de 482. Sin embargo, los cambios en los índices de ajuste son parecidos a los cambios observados para los datos desagrupados. En este caso también vemos una disminución significativa al agregar la variable sesión al modelo, hay una disminución en la devianza de 614.6 con la pérdida de un grado de libertad. Asimismo, al agregar sesión como un efecto aleatorio hay una disminución muy importante en la devianza de 1,624 con la pérdida de 2 grados de libertad únicamente. Finalmente así como se encontró con los modelos con los datos desagrupados, vemos que la variable grupo disminuye la devianza en apenas 20.9 perdiéndose 3 grados de libertad.

Tabla 15: Modelos para datos agregados por sesión.

Modelo	Efectos fijos	Efectos Aleatorios	AIC	Devianza	g.l.
Modelo 1	Intercepto: 0.56(0.37)	$\text{Var}(u_{0j})=2.286$	6381	6377	480
Modelo 2	Intercepto: 0.26(0.40) Sesión: 0.018(0.0007)*	$\text{Var}(u_{0j})=2.66$	5768	5762.4	479
Modelo 3	Intercepto: -0.42 (0.21) Sesión: 0.16 (0.06)*	$\text{Var}(u_{0j})=0.71207$ $\text{Var}(u_{1j})=0.05912$ $\text{Corr}(u_{0j},u_{1j})= 0.41$	4148.3	4138.3	477
Modelo 5	Intercepto: 2.53(0.622) Sesión: 0.16(0.059)* grupoABB: -3.67(0.69)* grupoABA: -3.61(0.70)* grupo AAB: -4.45(0.74)*	$\text{Var}(u_{0j})=1.70904$ $\text{Var}(u_{1j})=0.05587$ $\text{Corr}(u_{0j},u_{1j})= -0.95$	4133.4	4117.4	474

Básicamente, llegamos a las mismas conclusiones de cuál es el mejor modelo con los

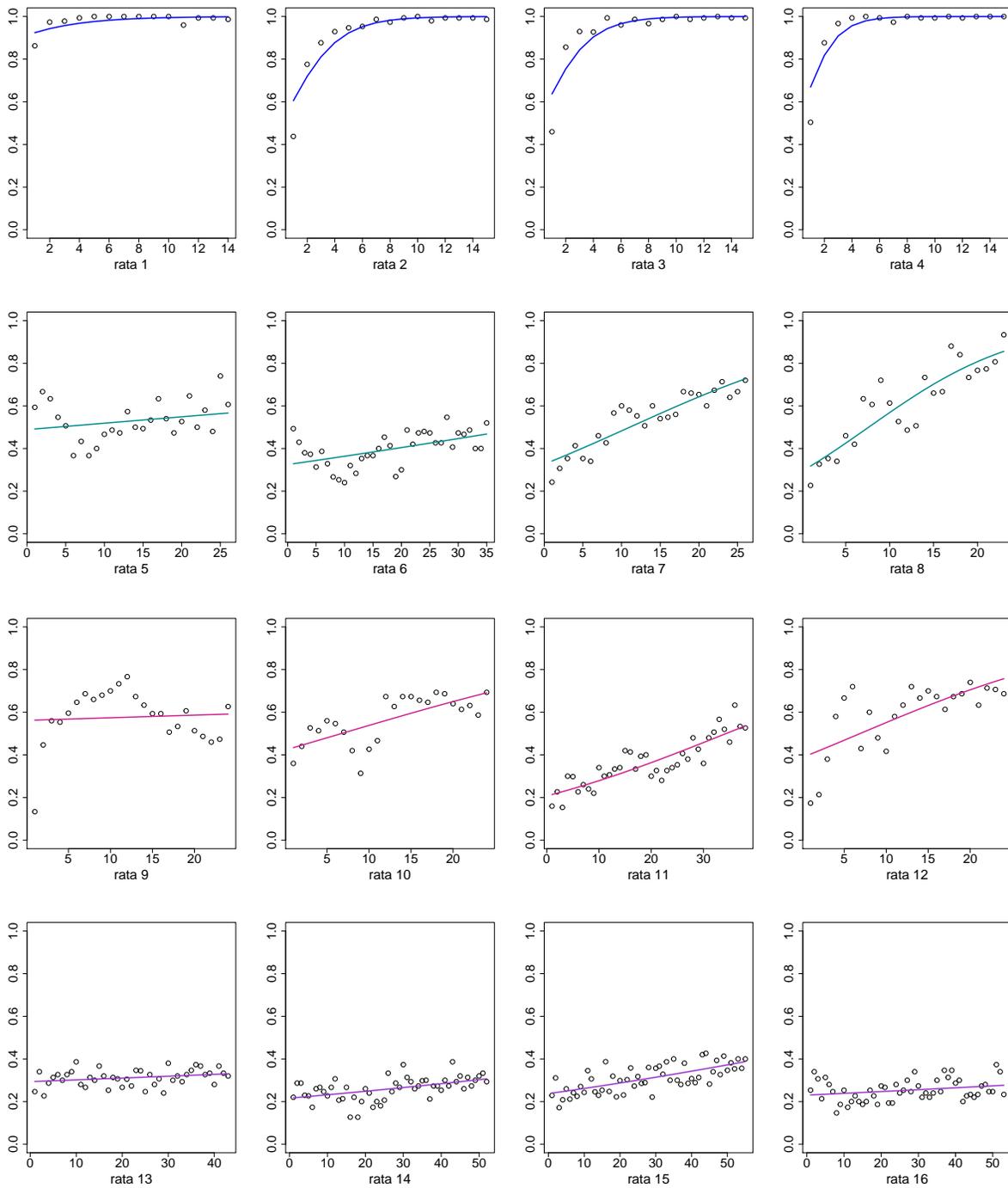
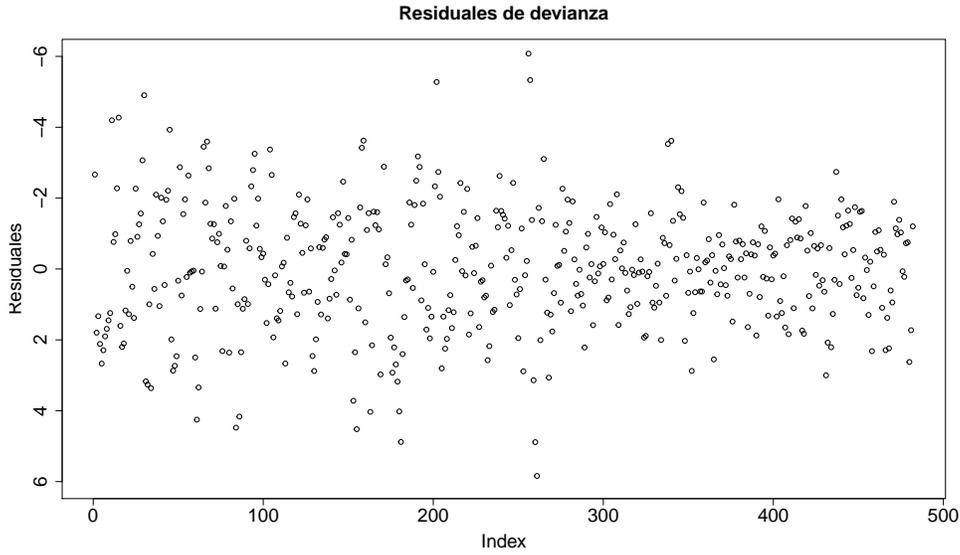


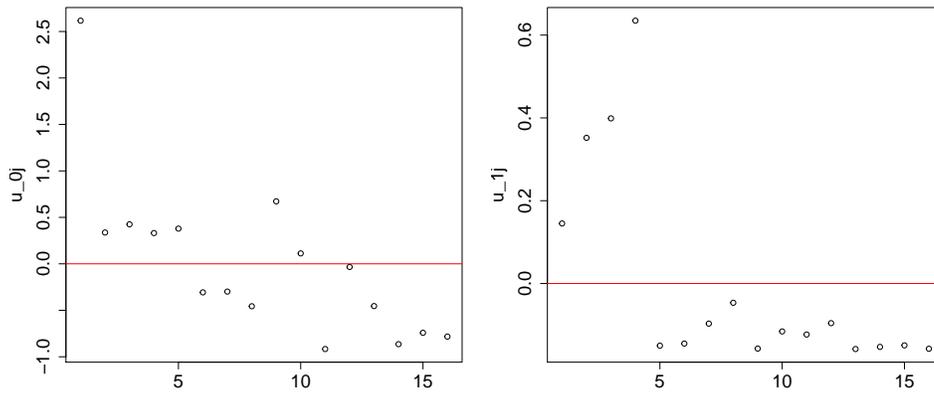
Figura 28: Valores predichos y observados por rata del modelo 3.

datos agrupados y desagrupados. Nos quedamos con el modelo 3 de los datos agrupados, y la interpretación es que un aumento de una sesión lleva a un incremento de 17% en las

posibilidades de ejecutar la estructura reforzada. Si pensamos en bloques de 5 sesiones, entonces tenemos que $e^{5 \cdot 0.16} = e^{0.8} = 2.22$, al pasar 5 sesiones las posibilidades de realizar la conducta correcta se duplican.



(a) Residuales de devianza a nivel 1.



(b) Residuales a nivel 2.

Figura 29: **Residuales del modelo 3 para los datos agrupados.**

Para tener una idea más clara del ajuste del modelo se graficaron los datos observados de cada individuo junto con lo predicho por el modelo 3 para los datos agrupados. En la figura 28 se muestran 16 gráficas, cada una correspondiente a un sujeto. En esta figura

vemos que las ratas mostraron mucha variación en la ejecución de la estructura reforzada a lo largo del experimento; sin embargo, las líneas ajustadas parecen predecir bien la tendencia general del aprendizaje de cada sujeto.

En la figura 29 se muestran los residuales de ambos niveles del modelo 3. En la gráfica (a) vemos los residuales de devianza del nivel 1, en donde parece no haber patrones claros, lo cual es bueno. Sin embargo, hay algunos residuales que son muy grandes, lo que podría indicar que hace falta alguna variable explicativa en el modelo. En la gráfica (b) vemos los residuales a nivel 2 tanto para el intercepto (gráfica de la izquierda) como para la pendiente (gráfica de la derecha). Dado que son muy pocos residuales de nivel 2, ya que sólo se tuvieron 16 grupos, es difícil interpretarlos. No obstante, parece que en el caso de los residuales de las pendientes, la variabilidad en ellos se debe principalmente a los primeros 4 residuales, los cuales pertenecen a las ratas del grupo AAA. Estas ratas aprendieron mucho más rápido que el resto de los sujetos del experimento, cuyos residuales son muy parecidos. Asimismo, en los residuales de los interceptos se puede ver que el primero residual correspondiente a la primera rata es muy diferente a los demás. Para ver si podría ser una observación influyente, se volvió a correr el tercer modelo pero sin los datos de la primera rata. Dicho modelo quedó de la siguiente manera:

Tabla 16: Modelo sin los datos de la primera rata.

Efectos fijos	Efectos Aleatorios	AIC	Devianza	g.l.
Intercepto: -0.60 (0.13)	$\text{Var}(u_{0j})=0.25685$	4037.9	4027.9	463
Sesión: 0.15 (0.06)*	$\text{Var}(u_{1j})=0.06287$			
	$\text{Corr}(u_{0j}, u_{1j})= 0.47$			

Si comparamos este modelo con el que sí incluye a la primera rata, el ajuste es aproximadamente el mismo. Lo que cambió fue la varianza estimada para los interceptos, la cual disminuyó considerablemente. Sin embargo, todo lo demás se mantuvo más o menos igual. Finalmente, cabe mencionar que los modelos también se corrieron con quasi-verosimilitud, encontrándose que los valores estimados para los parámetros fueron muy similares a los encontrados con máxima verosimilitud, por lo tanto, se decidió no incluirlos en el análisis.

Capítulo 7

Discusión

En el presente estudio se buscaba analizar la adquisición de secuencias conductuales tomando en cuenta su estructura espacio-temporal, dimensión de la conducta que ha sido poco estudiada, en parte, porque la unidad conductual comúnmente utilizada en el AEC ha sido una respuesta discreta. El uso de esta unidad de análisis ha aportado mucha información, no obstante, esta concepción de la unidad conductual tiene más de 100 años y el estudio de unidades analíticas más complejas puede llevar al análisis de otros procesos interesantes. En el presente trabajo se utilizaron diversos modelos lineales generalizados para analizar los datos recabados. Estos modelos nos proporcionaron una flexibilidad que permitió manejar diversas variables dependientes con diferentes distribuciones de probabilidad de la familia exponencial.

Generalmente, en el Análisis Experimental de la Conducta la definición de unidad de análisis se hace en términos metodológicos, quedando definida como aquella respuesta de la cual depende la entrega del reforzador. Dicha respuesta suele ser la única que se mide, aun cuando respuestas con diferentes propiedades ocurran. Por ejemplo, si se estuvieran reforzando respuestas ejecutadas con una fuerza específica, respuestas emitidas con otras fuerzas no serían consideradas como parte de la unidad conductual, aún cuando hayan sido producto del procedimiento establecido (Catania, 1973).

Consideramos que esta definición de unidad analítica llevó al AEC al menos a dos limitaciones. Primero, al definir la unidad de análisis como una respuesta discreta podemos

estar perdiendo cambios ordenados en otras conductas. Baum (2012) ha sugerido que si pensamos en la conducta como un continuo y en que existe un tiempo limitado en el que podemos dividir nuestro comportamiento, una unidad de análisis más significativa es la distribución de la conducta entre diversas actividades. De manera que registrar únicamente una respuesta lleva a un análisis incompleto del efecto de diversos eventos ambientales sobre el comportamiento. Asimismo, una sola respuesta podría ser parte de una unidad que está extendida en el tiempo.

La segunda limitación a la que llevó la definición de la unidad de análisis de Skinner fue que dado que se realizó en términos funcionales, la estructura espacio-temporal de las respuestas pasó a un segundo plano. Esto se debió en parte a que al considerar como la unidad conductual a una sola respuesta, los estudios de la estructura conductual se limitan a la topografía de la respuesta.

Si en cambio, tomamos a una secuencia de respuestas ordenadas como unidad, es posible manipular directamente la estructura, como se hizo en el presente trabajo. Asimismo, el uso de secuencias de respuestas permitió la definición de todas las posibles secuencias realizables por las ratas. De manera que medir su ocurrencia y analizar la distribución de la conducta de los animales entre las posibles secuencias fue un paso natural. Es así, que es posible y tiene ventajas extender la definición de la unidad conductual al estudio de la distribución de la conducta, en vez de hacer referencia a una sola respuesta.

El objetivo general del experimento analizado fue estudiar el efecto de la estructura espacio-temporal en diversos aspectos de la conducta de los animales. En primera instancia, se buscaba una forma de analizar las diferentes maneras en que los animales fueron ejecutando la estructura reforzada. Mediante el análisis de supervivencia fue posible tener una primera aproximación a la ejecución de las secuencias reforzadas, donde se observó que la ocurrencia fue muy diferente de acuerdo a la estructura. Para empezar, hubo una clara diferencia en la rapidez con que las conductas ocurrieron. Para el grupo AAA, las 2,000 conductas reforzadas ocurrieron muy rápido, mientras que para el grupo AAB fueron ocurriendo de manera más paulatina. Por su parte, el grupo ABA y ABB se agruparon quedando en medio. Aunque hubo mucha variabilidad, vista en el hecho de

que las pruebas log-rank resultaron altamente significativas, parece que si hubo diferencias debido a la estructura.

Para resumir las curvas de supervivencia de cada rata se utilizó la mediana. Este análisis nos acercó a tener un gradiente de dificultad para las diferentes estructuras, sugiriendo que el gradiente va desde la estructura AAA que es la más sencilla ya que su mediana es la más pequeña, seguida de las estructuras ABA y ABB cuyas medianas son más grandes, y finalmente AAB sería la estructura más complicada de ejecutar y la que lleva más tiempo. Neuringer *et al.* (2001) esbozó un gradiente de dificultad de acuerdo a la frecuencia con que diversas secuencias fueron ejecutadas, con resultados muy parecidos a los aquí encontrados. Así como en lo encontrado en este trabajo, la estructura más sencilla en los resultados de Neuringer fue la AAA mientras que estructuras con alternaciones fueron consideradas más complicadas. Neuringer hizo uso de secuencias de 3 respuestas a 3 operandos, lo que le dio un intervalo de estructuras más diverso.

Sin embargo, queda la pregunta de qué tanto más difícil es una estructura que la otra. Si consideramos a las medianas de las curvas de supervivencia como un indicador de la dificultad de ejecutar una estructura en particular, el análisis indicó que en promedio de las medianas de las 4 ratas del grupo AAA fue de 1,168, al comparar este valor con el promedio del grupo ABB, el cual fue de 2,343, tenemos que ABB es prácticamente el doble de difícil que AAA. Asimismo, se encontró que en promedio, la mediana de supervivencia del grupo ABA también es el doble de la mediana del grupo AAA. Para el grupo AAB, en promedio las ratas tuvieron una mediana de 3,833 lo cual equivale a 3.2 veces la mediana del grupo AAA y a 1.6 veces la mediana del grupo ABB y ABA. Parece haber una relación lineal en el aumento de las medianas de supervivencia de acuerdo a la estructura y podrían servir de indicador de la dificultad de ejecución.

Otro análisis que nos permitió conocer cómo fue que los sujetos experimentales fueron ejecutando la estructura reforzada fue la regresión logística multinivel. Este análisis nos permitió estudiar de manera directa las curvas de adquisición de la conducta de cada rata. El modelo final con el que se quedó nos permite argumentar que en promedio, las ratas en efecto estaban aprendiendo la conducta reforzada, ya que el coeficiente para

sesión resultó positivo. Esto nos lleva a considerar que es posible afirmar que el reforzador actuó sobre las secuencias definidas como unidades conductuales, incrementando su probabilidad, lo cual concuerda con la teoría clásica de reforzamiento (Zeiler, 1983). Sin embargo, fue necesario meter al modelo el intercepto y la variable sesión como efectos aleatorios, de manera que la pendiente y el intercepto variaron de sujeto a sujeto. Fue posible obtener la información de que a mayor intercepto mayor pendiente, ya que la correlación entre estos dos fue de 0.41. Esto tiene sentido, ya que las ratas que desde un inicio contestaron bien, aprendieron a ejecutar la estructura reforzada de manera más rápida, en comparación con las ratas que empezaron con una proporción de respuestas correctas más baja.

La estructura parece haber tenido un efecto sobre la variabilidad de los interceptos y las pendientes, sin embargo, debido a la gran variabilidad observada entre los sujetos se hizo un ajuste por individuo. El modelo capturó adecuadamente que las ratas del grupo AAA tuvieron interceptos mucho más altos que las demás ratas, mientras que las ratas del grupo AAB tuvieron interceptos por debajo de la ejecución de las otras ratas. Por su parte, las ratas del grupo ABB y ABA nuevamente quedaron en medio. Esto concuerda con el orden que se encontró con el análisis de supervivencia y en términos generales con los resultados de Polidora (1963).

Una mirada a los residuales u_{1j} de la regresión logística permitió notar que las pendientes más pronunciadas fueron para las ratas del grupo AAA, mientras que las pendientes de las demás ratas fueron mucho más pequeñas. Las pendientes del grupo AAB estuvieron cercanas a 0, lo que nos podría hacer dudar sobre si realmente aprendieron a ejecutar la estructura. Una ejecución azarosa entre las 4 estructuras llevaría a un reforzamiento de alrededor de 25 % en cada sesión. Las ratas de este grupo consiguieron alrededor de 30 % de los reforzadores disponibles por sesión, lo que no está muy lejos del porcentaje que se llevarían si hubieran estado contestando al azar.

En este modelo, también se probaron otras variables para ver si explicaban las curvas de adquisición de las estructuras reforzadas. De ellas, una de interés era la conducta en el ensayo previo, variable que se trabajó con el nombre de `status_1`. Para empezar,

cabe mencionar que dentro del AEC se ha hecho una distinción de acuerdo al nivel de explicación al que se recurre, algunos investigadores se identifican como moleculares y otros como molares. En investigaciones moleculares la relación más importante entre eventos es su cercanía momentánea en el tiempo. Por su parte, la aproximación molar argumenta que el comportamiento no sólo depende de los eventos presentes inmediatos, sino también de mucho eventos en el pasado, y ve al comportamiento como un agregado y no como ocurrencias momentáneas individuales. En este caso, la variable de la conducta en el ensayo anterior podría ser clasificada como un predictor de índole molecular, mientras que la variable sesión haciendo referencia al paso del tiempo sería molar. El hecho de que la variable `status_1` no fuera significativa, nos indica que una perspectiva molar, como la que es tomada por Baum (2012), nos llevó a un análisis general de la conducta extendida en el tiempo, considerando que predictores moleculares no fueron tan relevantes al describir las curvas de adquisición.

Hasta este punto la discusión se ha enfocado al análisis de la adquisición de las estructuras reforzadas, ya que el paradigma Skinneriano hace énfasis en cómo el reforzamiento “fortalece” la respuesta instrumental condicionada. Este énfasis suele llevar a la creencia de que el condicionamiento llevará irremediamente a la repetición de la misma respuesta, generando patrones conductuales estereotipados (Domjan, 2010). El análisis de únicamente la estructura reforzada dio mucha información, sin embargo, parte de la manera en que se definió la unidad conductual en el presente estudio, fue para poder realizar el análisis de las demás estructuras que los sujetos podían realizar. La medición de estas otras estructuras nos proporcionó otro índice para decidir si las ratas aprendieron algo o simplemente variaron aleatoriamente su conducta. En primera instancia, para afirmar que el reforzador tuvo un efecto sobre la conducta se buscaba que hubiera una discriminación de la estructura reforzada con respecto a las demás. En segunda instancia, se buscaba que la variabilidad observada de todas las estructuras no fuera aleatoria.

Se ha encontrado en muchos estudios que en situaciones de reforzamiento sostenido durante mucho tiempo, la variabilidad de las respuestas emitidas disminuye y ocurre la aparición de patrones de respuesta estereotipados. El fenómeno de variabilidad ha sido

asociado más bien a situaciones de extinción, en donde el reforzador es suspendido por completo. Ya desde (1951) Antonitis encontró un decremento de la variabilidad con reforzamiento y un aumento en el intervalo de las respuestas emitidas durante condiciones de extinción. No obstante, esto no implica que el reforzamiento no pueda llevar a variabilidad en la respuesta. De hecho Page y Neuringer (1985) demostraron en una serie de estudios que la variabilidad en efecto puede ser reforzada como una unidad, poniendo en duda que el único efecto del reforzamiento sea la estereotipia. En estos estudios se ha considerado que la variabilidad se da de manera aleatoria, sin un orden claro.

En nuestro estudio se hipotetizó que la variabilidad conductual entre las 4 posibles estructuras podría estar relacionada a la estructura reforzada. Se buscaba conocer si la ejecución de las demás estructuras había ocurrido completamente al azar o si se había conformado un patrón en ellas explicable en parte por la hipótesis de contigüidad. En este punto, el ajuste del modelo log-lineal fue de mucha importancia. Mediante la variable que llamamos *contigüidad* se pudo poner a prueba la hipótesis de si los errores habían sido producto del azar o si asumir una distribución de los errores de acuerdo a contigüidad era más plausible. Asimismo, esta variable nos permitió distinguir el proceso de aprendizaje de la estructura reforzada (la diagonal de las tablas de contingencia) de la ejecución de las demás estructuras. Los resultados mostraron que es una mejor explicación asumir que los errores siguen un orden de acuerdo a contigüidad. Se encontró que las posibilidades de ejecutar la estructura reforzada (la diagonal) fueron 11.3 veces las de cometer errores no relacionados a contigüidad y 3.6 las de cometer errores por contigüidad.

Estos resultados sugieren que las estructuras reforzadas fueron más ejecutadas que las demás estructuras no reforzadas, lo que da evidencia a favor de que el reforzamiento programado actuó a nivel de las secuencias con cierta estructura. Mientras que la contigüidad con el reforzador determinó, en parte, el orden de ejecución de las otras estructuras. Esto va de acuerdo al argumento propuesto por Bachá-Méndez *et al.* (2007), estos autores encontraron en varios experimentos llevados a cabo con secuencias de respuestas, que tanto procesos a nivel de la secuencia como a nivel de la respuesta operaron simultáneamente en unidades conductuales claramente establecidas. En el presente estudio, se confirmó la

presencia de estos dos procesos y fue posible cuantificarlo.

Si pensamos en estos resultados desde una perspectiva de adaptación, es claro que la variabilidad conductual tiene ventajas, ya que ante un cambio en el ambiente, permite que haya otras conductas disponibles en el repertorio que podrían ser exitosas. Asimismo, el hecho de que la variabilidad haya estado determinada en parte por un efecto de contigüidad también puede verse como adaptativo, ya que parece pertinente pensar que es beneficioso para el sujeto tener un sistema que genere variantes parecidas a la conducta que en el presente es exitosa.

Una posible explicación alternativa a los resultados, es que el orden encontrado en las estructuras se haya debido a la manera en que se midió y dividió la conducta para realizar el análisis. En el experimento, las ratas ejecutaban respuestas en dos palancas, una a la izquierda y una a la derecha, en ensayos en los que mediante estímulos discriminativos se demarcaban secuencias de 3 respuestas. Puede ser que las ratas realmente nunca agruparan su conducta en secuencias, sino que respondieran aleatoriamente entre las dos palancas y el orden observado se debiera a que en el análisis se dividió “a posteriori” su conducta en secuencias de tres respuestas y se agrupó por estructura. Para intentar descartar de cierta manera esta explicación alternativa, se decidió simular la ejecución de respuestas a dos opciones de manera aleatoria como lo hicieron Page y Neuringer (1985). De manera que se asignó una probabilidad de 0.5 a generar la respuesta *izquierda* y 0.5 a generar la respuesta *derecha*. Después, dichas respuestas aleatorias se agruparon en secuencias de tres respuestas y fueron clasificadas por estructura. En la figura 30 podemos ver cómo se vería la ejecución de las ratas si estuvieran respondiendo aleatoriamente entre las dos palancas.

Esta distribución de la conducta entre las 4 estructuras no se parece a los datos obtenidos, ya que en una sesión una estructura es ejecutada por arriba de todas, y en la siguiente sesión otra estructura es la más ejecutada, sin una distribución estable. Lo que se observó en los datos recolectados fue que hubo una consistencia en cuanto a la distribución de la conducta entre las posibles estructuras, aún en el grupo AAB que fue el que menos proporción de reforzadores logró obtener. De manera que parece haber

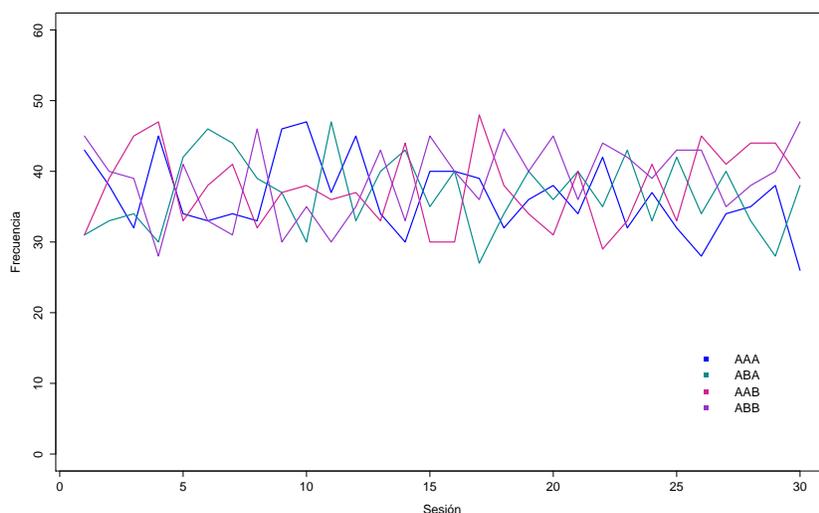


Figura 30: **Simulación de respuestas aleatorias a las dos palancas.**

evidencia que sugiere que el reforzamiento produjo variabilidad ordenada.

Algo que se observó cuando se graficaron las curvas de supervivencia para todas las conductas por rata fue que la variabilidad entre las estructuras fue mucho mayor para el grupo AAB que para los demás grupos. En este grupo se observó que todas las estructuras se ejecutaron hasta el final del experimento y las curvas de supervivencia no se diferenciaron. Los grupos ABB y ABA mostraron un punto intermedio de variabilidad y el grupo AAA fue el que menos variabilidad mostró, al final del experimento la única estructura que ejecutaban los sujetos de este grupo era la reforzada.

Puede ser que la mayor variabilidad observada en las diferentes condiciones de reforzamiento se haya debido a la diferencia en la intermitencia con que los animales obtuvieron el reforzador. Es sabido en el área que el reforzamiento intermitente suele estar acompañado de una mayor variabilidad que el reforzamiento continuo (Eckerman y Lanson, 1969). En nuestro caso, la dificultad de la ejecución de las diferentes estructuras, provocó que el reforzamiento no se diera con la misma frecuencia en los diferentes grupos. Aunque a todas las ratas se les sometió al mismo programa de reforzamiento continuo (i.e. cada vez que una rata realizaba la conducta correcta el reforzador estaba disponible), la frecuencia con que las ratas recibieron el reforzador varió, desde reforzamiento en prácticamente cada

ensayo para el grupo AAA, hasta reforzamiento cada 4 ensayos en promedio para el grupo AAB. Entonces, es posible que las diferencias en cantidad de variabilidad se hayan debido a un co-producto de la estructura reforzada y la intermitencia con que el reforzamiento ocurrió.

Parece claro que la estructura espacio-temporal es una dimensión de la conducta que resulta relevante en el aprendizaje. Las curvas de aprendizaje fueron muy diferentes para los sujetos de los diferentes grupos. El grupo AAA tuvo una ejecución perfecta, discriminando la estructura reforzada de las demás. Los grupos ABB y ABA lograron discriminar la estructura reforzada de las demás estructuras, pero siguieron ejecutando las otras conductas en un orden específico explicable por contigüidad. Finalmente el grupo AAB, nunca logró discriminar la estructura reforzada de las demás. Sin embargo, al ver la distribución completa de los sujetos de este grupo, parece haber emergido un orden, pero no resultó claro que lograran aprender completamente las secuencias reforzadas.

La relación entre comportamiento y reforzamiento fue significativa al ser analizada en términos de las secuencias y las estructuras reforzadas y no en términos de las respuestas individuales a cada palanca. El reforzamiento tuvo un efecto a nivel de la distribución de las conductas, lo cual fue observado en los cambios ordenados del repertorio conductual de los animales ante cambios en la estructura causal de su ambiente. En el marco conceptual esbozado por Baum (2012) podemos pensar de manera más fácil en cómo eventos biológicamente importantes llevan al incremento y decremento de varias respuestas extendidas en el tiempo, induciendo cambios en la distribución completa de la conducta. El estudio de las secuencias nos permitió dar el paso hacia un análisis de este tipo de manera directa y natural.

Referencias

- Agresti, A. (2007). *An introduction to categorical data analysis* (2da ed.). Hoboken, New Jersey: Wiley & Sons, Inc.
- Antonitis, J. J. (1951). Response variability in the white rat during conditioning, extinction and reconditioning. *Journal of the Experimental Psychology*, 42(4), 273-281.
- Bachá-Méndez, G., Reid, A. K., y Mendoza-Soylovna, A. (2007). Resurgence of integrated behavioral units. *Journal of the Experimental Analysis of Behavior*, 87, 5-24.
- Baum, W. M. (2005). *Understanding behaviorism. behavior, culture and evolution* (2da ed.). Oxford, U.K.: Blackwell Publishing.
- Baum, W. M. (2012). Rethinking reinforcement: allocation, induction and contingency. *Journal of the Experimental Analysis of Behavior*, 97(1), 101-124.
- Boring, E. G. (2006). *Historia de la psicología experimental* (3ra ed.). D.F., México: Editorial Trillas.
- Branch, M. N. (1977). On the role of memory in the analysis of behavior. *Journal of the Experimental Analysis of Behavior*, 28, 171-179.
- Brown, H., y Prescott, R. (2006). *Applied mixed models in medicine* (2da ed.). West Sussex, England: Wiley & Sons, Inc.
- Catania, A. (1971). Reinforcement schedules: the role of responses preceding the one that produces the reinforcer. *Journal of the Experimental Analysis of Behavior*, 15, 271-287.
- Catania, A. (1973). The concept of the operant in the analysis of behavior. *Behaviorism*, 1(2), 103-115.
- Collett, D. (2003a). *Modelling binary data* (2da ed.). Boca Ratón, Florida: Chapman &

- Hall/CRC. (Texts in statistical science series).
- Collett, D. (2003b). *Modelling survival data in medical research* (2da ed.). Boca Ratón, Florida: Chapman & Hall/CRC. (Texts in statistical science series).
- Dobson, A. J. (2002). *An introduction to generalized linear models* (2da ed.). Boca Ratón, Florida: Chapman & Hall/CRC. (Texts in statistical science series).
- Domjan, M. (2010). *Principios de aprendizaje y conducta* (6ta ed.). D.F., México: Wadsworth, Cengage learning.
- Dunteman, G. H., y Ho, M. R. (2006). *An introduction to generalized linear models* (1ra ed.). Thousand Oaks, California: Sage Publications, Inc. (Quantitative Applications in the social science).
- Eckerman, D. A., y Lanson, R. N. (1969). Variability of response location for pigeons responding under continuous reinforcement, intermittent reinforcement and extinction. *Journal of the Experimental Analysis of Behavior*, 12(1), 73-80.
- Faraway, J. (2006). *Extending the linear model with r. generalized linear, mixed effects and nonparametric regression models* (1ra ed.). Boca Ratón, Florida: Chapman & Hall/CRC. (Texts in statistical science series).
- Goldstein, H. (2011). *Multilevel statistical methods* (4ta ed.). West Sussex, U.K.: Wiley & Sons, Inc. (Wiley series in probability and statistics).
- Gosling, S., John, O. P., Craik, K. H., y Robins, R. W. (1998). Do people know how they behave? self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74, 1337-1349.
- Grayson, R. J., y Wasserman, E. A. (1979). Conditioning of two-response patterns of key pecking in pigeons. *Journal of the Experimental Analysis of Behavior*, 31, 23-29.
- Kleinbaum, D. G., y Klein, M. (2012). *Survival analysis. a self-learning texts* (3ra ed.). New York, U.S.A.: Springer. (Statistics for biology and health).
- Lee, E. T., y Wang, J. W. (2003). *Statistical methods for survival data analysis* (3ra ed.). Hoboken, New Jersey: Wiley & Sons, Inc (Wiley series in probability and statistics).
- Lieberman, A. D. (2012). *Human learning and memory* (1ra ed.). New York, U.S.A.: Cambridge University Press.

- Lindsey, J. K. (1997). *Applying generalized linear models* (1ra ed.). New York, U.S.A.: Springer (Springer texts in statistics).
- Mazur, J. E. (1998). *Learning and behavior* (4ta ed.). Upper Saddle River, New Jersey: Prentice Hall, Inc.
- Myers, R. H., Montgomery, D. C., Vining, G. G., y Robinson, T. J. (2010). *Generalized linear models. with applications in engineering and the sciences* (2da ed.). Hoboken, New Jersey: Wiley & Sons, Inc (Wiley series in probability and statistics).
- Neuringer, A., Kornell, N., y Olufs, M. (2001). Stability and variability in extinction. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*, 79-94.
- Page, S., y Neuringer, A. (1985). Variability is an operant. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*(1), 429-452.
- Polidora, V. J. (1963). A sequential response method of studying complex behavior in animals and its application to the measurement of drug effects. *Journal of the Experimental Analysis of Behavior*, *6*, 271-277.
- R Core Team. (2013). R: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria. Descargado de <http://www.R-project.org/>
- Reid, A. K., Chadwick, C. Z., Dunham, M., y Miller, A. (2001). The development of functional response units: the role of demarcating stimuli. *Journal of the Experimental Analysis of Behavior*, *76*, 303-320.
- Reid, A. K., Dixon, R. A., y Gray, S. (2008). Variation and selection in response structures. En N. K. Innis (Ed.), *Reflections on adaptive behavior: Essays in honor of j. e. r. staddon*. U.S.A.: Cambridge, MA: MIT Press.
- Schick, K. (1971). Operants. *Journal of the Experimental Analysis of Behavior*, *15*, 413-423.
- Schwartz, B. (1980). Development of complex, stereotyped behavior in pigeons. *Journal of the Experimental Analysis of Behavior*, *33*, 153-166.
- Shimp, C. P. (1976). Organization in memory and behavior. *Journal of the Experimental Analysis of Behavior*, *26*, 113-130.

- Snijders, T. A. B., y Bosker, R. J. (1999). *Multilevel analysis. an introduction to basic and advanced multilevel modeling* (1ra ed.). London, U.K.: Sage Publications Inc.
- Timberlake, W. (1993). Behavior systems and reinforcement: an integrative approach. *Journal of the Experimental Analysis of Behavior*, 60, 105-128.
- Twisk, J. W. R. (2006). *Applied multilevel analysis* (1ra ed.). New York, U.S.A.: Cambridge University Press (Practical guides to biostatistics and epidemiology).
- VonEye, A., y Mun, E. Y. (2013). *Log-linear modeling: Concepts, interpretation, and application* (1ra ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- West, B. T., Welch, K. B., y Galecki, A. T. (2007). *Linear mixed models. a practical guide using statistical software* (1ra ed.). Boca Ratón, Florida: Chapman & Hall/CRC.
- Zeiler, M. (1983). Programas de reforzamiento. variables controladoras. En W. K. H. y J. E. R. Staddon (Ed.), *Manual de conducta operante*. D. F., México: Editorial Trillas.