Universidad Nacional Autónoma de México

Facultad de Ciencias

Modelación Estadística Interdisciplinaria: Un Caso de Estudio Empleando Modelos Lineales con Efectos Mixtos para el Análisis de Datos de Adicciones

T E S I S

QUE PARA OBTENER EL GRADO DE:

Actuario

PRESENTA:

Ana Paulina Pérez Romero

DIRECTOR DE TESIS:

Dr. Miguel Nakamura Savoy

Ciudad Universitaria, México DF

Septiembre, 2015

Modelación Estadística Interdisciplinaria: Un Caso de Estudio Empleando Modelos Lineales con Efectos Mixtos para el Análisis de Datos de Adicciones

por

Ana Paulina Pérez Romero

Tesis presentada para obtener el grado de

Actuario

en la

Facultad de Ciencias

Universidad Nacional Autónoma de México

Ciudad Universitaria, México DF. Septiembre, 2015

A mi madre...

Madre mía podría escribir libros enteros tan solo para recordar y agradecer todo lo que has hecho y haces por mí; podría escribir y escribir y te aseguro que jamás bastarían las palabras para mostrar el gran amor y admiración que siento por ti. Eres la persona que me enseño a luchar para que los sueños se hagan realidad. Eres un claro ejemplo de amor, bondad y perseverancia, simplemente eres genial. Te quiero hasta el infinito aunque tú solo me quieras hasta la Tierra ¿lo recuerdas?, nunca lo olvides.

A mi hermano...

*Hermano mío la vida ha sido una gran aventura a tu lado. Con el tiempo hemos aprendido que el trabajo en equipo permite alcanzar objetivos individuales y de grupo. En todo momento has sido mi apoyo y parte de mi guía; he aprendido de ti. A tu lado he jugado, reído, llorado, gritado, cantado, caminado... y ¿sabes qué? siempre es y será un gusto platicar contigo. Te agradezco una de las enseñanzas más grandes que he tenido: la vida no se trata de buscar sentirse siempre dichoso ó feliz, sino más bien de vivir en plenitud aunque eso implique momentos de tristeza y dolor. Conocedor del mundo, inteligente, vanidoso, orgulloso, enojón, y últimamente precavido simplemente mi hermano y no te cambiaría por nadie. Eres el mejor hermano del mundo. ¡*Bat girl* y mantecadita, por siempre!*

A mis abuelos...

Mis abuelos: Juana y Cirilo. Abuelita, cuanta paciencia, comprensión y sensibilidad hay en su persona, no cabe duda que la vida es la mejor escuela. Le agradezco por compartir sus experiencias conmigo. Don Cirilo, claro ejemplo de trabajo. ¿Quién dijo que el tiempo hace olvidar? pareciera que fue ayer cuando estabas a mi lado y sin embargo han pasado 16 años, ahora solo vives en mi recuerdo. Ya no estás pero creo que vas cumpliendo tu promesa y yo estoy es camino de cumplir la mía.

Agradecimientos

A la Universidad Nacional Autónoma de México y su Facultad de Ciencias, porque me permitieron formarme como profesionista. Nuevos profesionistas para el país implican propuestas novedosas para el desarrollo del mismo.

Al Centro de Investigación en Matemáticas A.C., por el apoyo económico y por dar oportunidad a que estudiantes de diferentes estados del país puedan adquirir nuevas experiencias en su formación profesional.

Al Dr. Miguel Nakamura, un profesor y una persona extraordinarios. Agradezco enormemente todo el apoyo, tiempo y trabajo dedicados a la realización de este escrito. «En la vida hay situaciones feas y situaciones extraordinarias; con la suficiente sensibilidad reconocerás cada una de ellas.» MN

A los sinodales: Dra. Ruth Fuentes, M. en C. José Zamora, Dr. Ricardo Ramírez y Act. Francisco Sánchez. Agradezco sus comentarios y correcciones al presente trabajo, porque en su conjunto pemitieron mejorarlo.

Al Mtro. Luis Pérez, por el tiempo y trabajo brindados a lo largo de la elaboración del texto. Reconocer la importancia de la formación de un equipo interdisciplinario fue la base del planteamiento de una investigación sólida.

A todos mis profesores, por todas sus enseñanzas y palabras de ánimo para continuar siempre adelante. «Las palabras no son tanto de quien las dice, sino de quien realmente las siente y vive su mensaje.» Sergio Olarte

A la familia Romero Torres, en especial a Rosa, Alfredo, Miguel, Hortencia y Rafael, porque fueron un gran apoyo en mi formación en los años de mi infancia. «No hay nada mejor ni más sano, más sólido y útil para los años postreros, que algunos buenos recuerdos, sobre todo si se relacionan con la infancia, ... Si un hombre acumula a lo largo de su vida muchos de esos recuerdos, estará seguro hasta el final de sus días. Y aunque sólo conservemos un recuerdo grato en el corazón, incluso este puede servir en un momento dado para salvarnos.» Fedor Dostoievski

A mis amigos, por todos aquellos momentos compartidos y por todos los que nos quedan por vivir. «La gloria de la amistad no es la mano tendida, ni la sonrisa bondadosa, ni disfrutar

de compañía; es la inspiración espiritual que sentimos al descubrir que alguien cree en nosotros, y está dispuesto a darnos confianza.» Ralph Waldo Emerson

A todas aquellas personas que he encontrado a lo largo de mi andar por la vida, porque cada una de ellas me ha dejando una enseñanza. «When we take action on the things that truly matter deep in our hearts, when we move in directions that we consider valuable and worthy, when we clarify what we stand for in life and act accordingly, then our lives become rich and full and meaningful, and we experience a powerful sense of vitality. This is not some fleeting feeling, it is a profound sense of a life well lived. And although such a life will undoubtedly give us many pleasurable feelings, it will also give us uncomfortable ones, such as sadness, fear and anger. This is only to be expected. If we live a full life, we will feel the full range of human emotions.» Russ Harris

Índice general

Modelación Estadística Interdisciplinaria: Un Caso de Estudio Empleando Modelos Lineales con Efectos Mixtos para el Análisis de Datos de Adicciones

por

Ana Paulina Pérez Romero

Resumen

El análisis de datos así como el planteamiento de modelos matemáticos que establecen relaciones entre variables para representar fenómenos naturales o bien procedimientos diseñados por el hombre, son los principales intereses de aquellos quienes hacen estadística. Todo esto se realiza con la finalidad de tomar decisiones o entender algo respecto a la esencia del fenómeno.

En el presente trabajo se establece un modelo estadístico que expresa la relación entre las variables involucradas en un fenómeno psicológico. Con dicho modelo se busca conocer si la variación existente en la variable nivel de consumo de sustancias psicoactivas se puede explicar como función de un conjunto de variables independientes. Asimismo se busca dar las características que debe tener la información a recolectar en el trabajo de campo para que al ser analizadas se puedan dar respuestas estadísticamente pertinentes al problema.

Las características del problema sugieren algunos puntos importantes a considerar. Por un lado tenemos que se realizarán mediciones sobre un mismo paciente a lo largo del tiempo que dure el tratamiento, debido a lo cual podemos afirmar que el conjunto de observaciones provenientes de una misma persona no son independientes entre sí. Por otro lado, con base en estudios psicológicos, el experto estableció la relación existente entre las variables involucradas en el estudio.

Un modelo lineal con efectos mixtos o modelo lineal mixto (explicado ampliamente en Demidenko, 2004 y Jiang, 2007) es, en términos generales, un modelo estadístico que permite trabajar con datos agrupados o bien datos longitudinales. Este tipo de modelos considera dos fuentes de variación en los datos: la variación entre los distintos grupos o individuos y la variación entre las observaciones de un mismo grupo o individuo. Los modelos lineales mixtos reciben este nombre debido a que involucran efectos fijos y efectos aleatorios al mismo tiempo. En estos modelos se distinguen dos tipos de coeficientes: los efectos que corresponden a toda la población (poblacionales) y los específicos por sujeto. Los primeros son constantes desconocidas por estimar como en un modelo de regresión lineal múltiple; mientras que los segundos son variables aleatorias de quienes se estiman los parámetros que definen su distribución, aunque también se puede estimar el efecto a través de una esperanza condicional obteniendo así un predictor.

Así, el modelo quedó establecido como un modelo de 3 niveles relacionados entre sí, cada uno de los niveles conformado por uno o tres modelos lineales con efectos mixtos. Este planteamiento requirió tanto de la revisión de las bases teóricas de esta clase particular de modelos estadísticos, como de la interacción continua psicólogo-estadístico (el trabajo multidisciplinario se ve reflejado en cada paso de la formulación del modelo ya señalado) para poder conseguir un modelo adecuado.

Se realizará la simulación de cuatro escenarios diferentes. El primero de ellos fue establecido por el experto. El segundo fue propuesto con base en las condiciones del tratamiento psicológico,

principalmente la disponibilidad de tiempo. Los últimos dos casos se sustentan en la importancia de sugerir una solución estadísticamente justificada. Considerar diferentes situaciones y compararlas entre sí permitirá situarnos lo más cerca del caso más viable de realizarse. Cada juego de datos obtenido de cada escenario es considerado como una muestra aleatoria real. Con la información construida de esta forma se puso en marcha la metodología estadística (con ayuda de la librería $lme4$ en el software estadístico R) que podrá ser empleada de manera idéntica una vez que se tenga la información real. La metodología estadística arrojará como resultados estimadores puntuales e intervalos de confianza para cada uno de los parámetros. Así, hemos llegado a proponer como la opción más viable a aquella que toma en cuenta 18 observaciones por paciente y 34 pacientes para el estudio. Un experimento con estas características permitirá al experto detectar diferencias significativas (si las hay) en los parámetros que son de interés primordial.

Asimismo se presentan las tablas de datos simulados y los gráficos que en su conjunto permiten, de manera visual, conocer el comportamiento de cada una de las variables involucradas en el estudio. Estos dos últimos instrumentos permitieron ilustrar al usuario las modificaciones que ocurren en los datos simulados cuando se cambian los valores asignados a los parámetros que se ingresan al simulador.

Capítulo 1

Introducción

Enfrascados en una complicada pero entretenida conversación, una tarde de verano, dos personas se divisan a los lejos. Uno de ellos pregunta «¿Cuál será el modelo estadístico que mejor represente la relación entre las variables de mi estudio?, ¿Qué características deberán tener los datos que recolecte para que después de ser analizados obtenga resultados estadísticamente significativos?». Bueno, y es que siendo un investigador ¿quién no se preguntaría esto? al estar frente a un comité que exige delimitar: tiempo, recursos económicos, disponibilidad de personal, horas de trabajo de campo, etc. Entonces aquel, el experto en estadística dijo: «Pues bien, en cierta medida, la estadística puede dar respuesta a tus preguntas, así que comencemos a trabajar.»

El análisis de conjuntos de datos (información proveniente de mediciones o encuestas hechas a grandes poblaciones) así como el planteamiento de modelos matemáticos que establecen relaciones entre variables para la representación de fenómenos naturales o bien de procedimientos diseñados por el hombre, estos últimos siempre condimentados con esa pizca natural de aleatoriedad, son los principales intereses e inspiración de aquellos que hacen estadística. He aquí la importancia de esta sutil herramienta tan apreciada por unos cuantos.

Una de las finalidades de la modelación estadística es la toma de decisiones. Ya sea en la ciencias médicas, biológicas ó físicas, en la industria, en instituciones públicas, *etc.*, la estadística siempre estará presente. El análisis de datos, como bien lo menciona Torres (2015) para el periódico El País, se ha llevado a todos los ámbitos de la vida. Desde los bancos hasta la NBA, donde se puede concluir qué jugadores, cómo y cuándo rinden mejor a partir de analizar el número de tiros fallados y acertados así como la distancia que hay entre las posiciones de los jugadores y la altura de los mismos. También es posible anticipar picos de desempleo. «Las administraciones pueden tomar medidas proactivas para hacer más fácil la vida de los ciudadanos, como incrementar la frecuencia del metro o reducir el precio de los billetes» comenta Fernando Meco, director de marketing en SAS.

Toda esta gama de posibilidades para el empleo de la estadística nos lleva al conocido trabajo multidisciplinario. Un equipo de trabajo multidsciplinario es un grupo de personas con diferentes profesiones o experiencias laborales que se plantean resolver un determinado problema. Un grupo con esta estructura permite un adecuado planteamiento de modelos, ya que las hipótesis y suposiciones estarán sustentadas por argumentos que ofrezcan cada uno de los expertos.

Por ejemplo, predecir la explosión de un volcán ubicado en las cercanías de un poblado se podría hacer con una precisión y confiabilidad determinados si en el estudio participan físicos, geólogos y estadísticos. En este caso, la respuesta a la interrogante permitiría tomar las medidas de seguridad pertinentes en el mejor tiempo y forma posibles para mantener a salvo a la población de los alrededores. Como bien podemos darnos cuenta con un adecuado análisis de la información se puede dar una solución a problemas que debido a su naturaleza tienen asociado de forma inherente el bienestar público.

El modelo estadístico que aquí presentaremos es el resultado de horas de trabajo individual y en conjunto de un estadístico y un psicólogo. Una de las finalidades de dicha tarea fue representar un fenómeno psicológico a través de un modelo matemático así como plantear las bases del diseño experimental para la realización del trabajo de campo, es decir, dar las características que debe tener la información a recolectar para poder dar respuestas estadísticamente significativas al problema.

1.1. Motivación

La realización de un trabajo de modelación matemática por parte de un estadístico en conjunto con un experto en un área de conocimiento diferente a la estadística, no resulta ser una opción al alcance de todos los estudiantes. Sin embargo, se puede llegar a presentar la oportunidad. Un experto en el área de salud mental ha visto la importancia de trabajar en conjunto con un estadístico para el análisis de los datos que obtiene al realizar sus estudios, dado el requerimiento de tener que reportar resultados estadísticamente significativos para poder afirmar que efectivamente sus estudios aportan información para la creación de nuevo conocimiento. Si bien es cierto que ambos investigadores (el estadístico y el psicólogo) se inician en sus correspondientes áreas (no tienen un grado alto de experiencia) también es cierto que los dos expertos pueden llegar a formar un buen equipo de trabajo que pudiera conducirlos a futuros estudios.

La formulación de un modelo estadístico desde el inicio hasta el final resulta ser un proceso muy interesante y más aún si se trata de un estudio que aún no se realiza. Esta situación abarca desde saber preguntar y explicar, hasta el tener que revisar nuevos temas de la teoría estadística y usar dicha información en el planteamiento del modelo. Cuando existe comunicación continua entre los expertos se puede preguntar al respecto de cualquier inquietud que surja en el proceso de formulación. Con un adecuado trabajo en equipo se plantearán hipótesis que sean lo más concordantes con la realidad, es decir, no se harán tantas suposiciones fuera de lo que es posible.

Existen eventos como el Taller de Solución de Problemas Industriales organizado por el Centro de Investigación en Matemáticas, A.C. que involucran a los estudiantes con el propósito de dar posibles soluciones a dudas que aquejan tanto a investigadores de áreas de la salud y biológicas como a las empresas privadas y públicas. Resulta motivador para el estudiante formar parte de esta propuesta de trabajo pues se enfrenta a situaciones particulares de modelación.

De la experiencia de participar en este evento podemos resaltar un par de comentarios. El primero nos dice algo acerca de la comunicación; en un principio pareciera que no se pudiese lograr una comunicación efectiva, posiblemente por el desconocimiento de muchos tecnicismos empleados por cada uno de los expertos. Suele ocurrir que algunas ocasiones las personas empleamos ciertas palabras en el lenguaje coloquial y tenemos la falsa creencia de conocer el empleo de la misma en ciertas especialidades o profesiones. Establecer una buena comunicación y tolerancia entre cada una de las partes será un buen cimiento para alcanzar la solución de problemas que sean de interés común.

Un segundo comentario plantea que el establecimiento de las diferentes problemáticas se puede presentar básicamente de dos formas. En la primera, el especialista tiene preguntas por resolver. Se cuenta con información; el problema es que no se tiene una metodología de análisis de datos. Es importante mencionar que la información puede no ser suficiente en cuestión de cantidad y calidad. En la segunda forma, el especialista pregunta por el diseño experimental. Aquí el interés principal es la cantidad y características de la información que se requiere para poder contestar interrogantes específicas. Esto se hace justo antes de iniciar el trabajo de campo para la recolección de datos. Parece muy conveniente el trabajo que el estadístico y el especialista pueden hacer antes de iniciar el estudio correspondiente pues, entre otras cosas, permitirá una mejor distribución de recursos como lo es el tiempo.

1.2. Problema en Psicología del adicto

El consumo de sustancias psicoactivas en México es un gran problema debido a las consecuencias sociales y personales que trae consigo: accidentes automovilísticos, delincuencia, violencia intrafamiliar, desempleo, *etc.* La Encuesta Nacional de Adicciones 2011 indica que de 2002 a 2008 hubo una tendencia a la alza en el consumo de sustancias; sin embargo, para el periodo que va de 2008 a 2011 no se observaron cambios significativos lo cual no necesariamente implica que se deje de prestar atención al problema pues el gobierno mexicano sigue invirtiendo en políticas públicas para erradicar el problema que aqueja a hombres y mujeres, desde adolescentes hasta personas adultas y de todos los estratos sociales.

Actualmente, los tratamientos psicológicos son considerados una buena herramienta para combatir los altos niveles de consumo de sustancias psicoactivas en México. La mayoría de estos programas, como lo menciona Pérez-Romero (2014), son cognitivo-conductuales y a pesar de que han demostrado ser eficaces la tasa de recaída al consumo es elevada por lo cual se ha considerado diseñar nuevos tratamientos o, en su caso, mejorar los existentes.

1.2.1. Estudio preliminar en Psicología del adicto

Los programas cognitivo-conductuales ya existentes podrían ser mejorados integrándoles técnicas basadas en atención plena (*mindfulness*) y aceptación. Ambos procesos enfatizan el estar atento a los estímulos internos (deseos intensos por consumir ó *craving*) y externos, sin juzgarlos, aceptándolos tal y como son. El experto realizó un estudio piloto que consistió en la integración de las técnicas ya mencionadas dentro de un programa cognitivo conductual tradicional ya existente (Prevención Estructurada de Recaídas para Estudiantes Universitarios (PEREU), traducido y adaptado por Quiroga & Vital, 2003) con el propósito de ampliar su eficacia.

Uno de los objetivos de dicho estudio fue explorar si existía relación entre las variables: supresión de pensamientos (medida a través del Cuestionario de Supresión de Pensamientos-Oso Blanco), *craving* (evaluada a través de la Escala de Alcohol de Penn y adaptada a otras drogas) y días transcurridos desde el último consumo. El psicólogo sugirió posible esta relación debido a que la literatura especializada en el tema señala que existe interacción entre las variables supresión de pensamientos y *craving*, y que ambas se encuentran asociadas a la recaída.

La información recabada en dicho estudio proviene de 29 pacientes. Las variables medidas en cada una de las personas son: sustancia psicoactiva de consumo, años de consumo de la sustancia, días sin consumir la sustancia antes de iniciar el tratamiento, supresión de pensamientos y nivel de *craving*.

En ese estudio, asimismo se sugirió un modelo de regresión lineal múltiple para modelar el nivel de la variable *craving* como función de las variables supresión de pensamientos y días sin consumir la sustancia antes de iniciar el tratamiento. Si bien es cierto que este es un modelo posiblemente simple, también es cierto que parece ser una primera buena aproximación. Esta primera propuesta motiva a realizar mejorías sobre la misma o, en dado caso, reestructurar el diseño experimental.

La base de datos generada por el estudio piloto mostró ser de gran utilidad en el presente trabajo pues ayudó a rescatar la distribución de las variables, con la finalidad de simularlas (como veremos en el Capítulo 3) debido a la falta de información al respecto. Estos datos simulados permitirán practicar la estimación de parámetros que se realizaría en caso de contar con datos reales. Para la presente situación prestemos especial atención a las variables: sustancia psicoactiva de consumo, años de consumo de la sustancia y días sin consumir la sustancia antes de iniciar el tratamiento.

1.2.2. Protocolo de investigación en Psicología del adicto

La realización y análisis del estudio piloto motivó a continuar sobre la misma línea de investigación. Esta vez se buscará un modelo que represente la relación existente entre diferentes grupos de variables, y con ayuda de este mismo, comprobar la eficacia de un nuevo tratamiento dirigido a personas consumidoras de sustancias psicoactivas en México. La eficacia de dicha terapia se verá reflejada en la disminución estadísticamente significativa de la cantidad de consumo de drogas.

Los tratamientos psicológicos generalmente consisten en una serie de intervenciones que siguen una secuencia, en otras palabras, el terapeuta entrena al paciente en la práctica de nuevas conductas a lo largo de un determinado periodo de tiempo. Medir la eficacia que algún tratamiento podría tener sobre un paciente implica medir ciertas características de la persona a lo largo del tiempo; de esta forma podríamos observar la existencia de posibles variaciones en los niveles de dichas características, teniendo así dos o más mediciones de la misma variable sobre el mismo individuo.

El tratamiento busca de forma indirecta que los pacientes tengan una mayor aceptación de sus eventos privados (como lo son sensaciones, emociones y pensamientos) lo cual modificará la forma en que las personas se relacionan con su estado de ánimo negativo reflejado en sus niveles de ansiedad, depresión y *craving*. Lo anterior podría llegar a reducir el nivel de consumo de drogas. Aunado a esto, el experto, formulador del tratamiento, ha dicho que existen ciertas características en los pacientes que son importantes considerar al evaluar la efectividad de la terapia: años de consumo de la sustancia psicoactiva, días sin consumo de sustancias psicoactivas antes de iniciar el tratamiento y número de tratamientos previos.

En la Figura (1-1) podemos observar la estructura general del tratamiento. Las diferentes variables involucradas en el estudio se pueden agrupar en cuatro grandes grupos con base en hipótesis de la teoría psicológica. Un primer grupo representa al constructo de aceptación, integrado por las variables aceptación de emociones, sensaciones y pensamientos. El segundo grupo caracteriza al Estado Afectivo Negativo (EAN) de las personas, está compuesto por depresión, ansiedad y *craving*. Un tercer conjunto consiste de las variables que son inherentes al paciente: años de consumo de la sustancia psicoactiva, días sin consumo de sustancias psicoactivas antes de iniciar el tratamiento y número de tratamientos previos. Finalmente consideramos a la

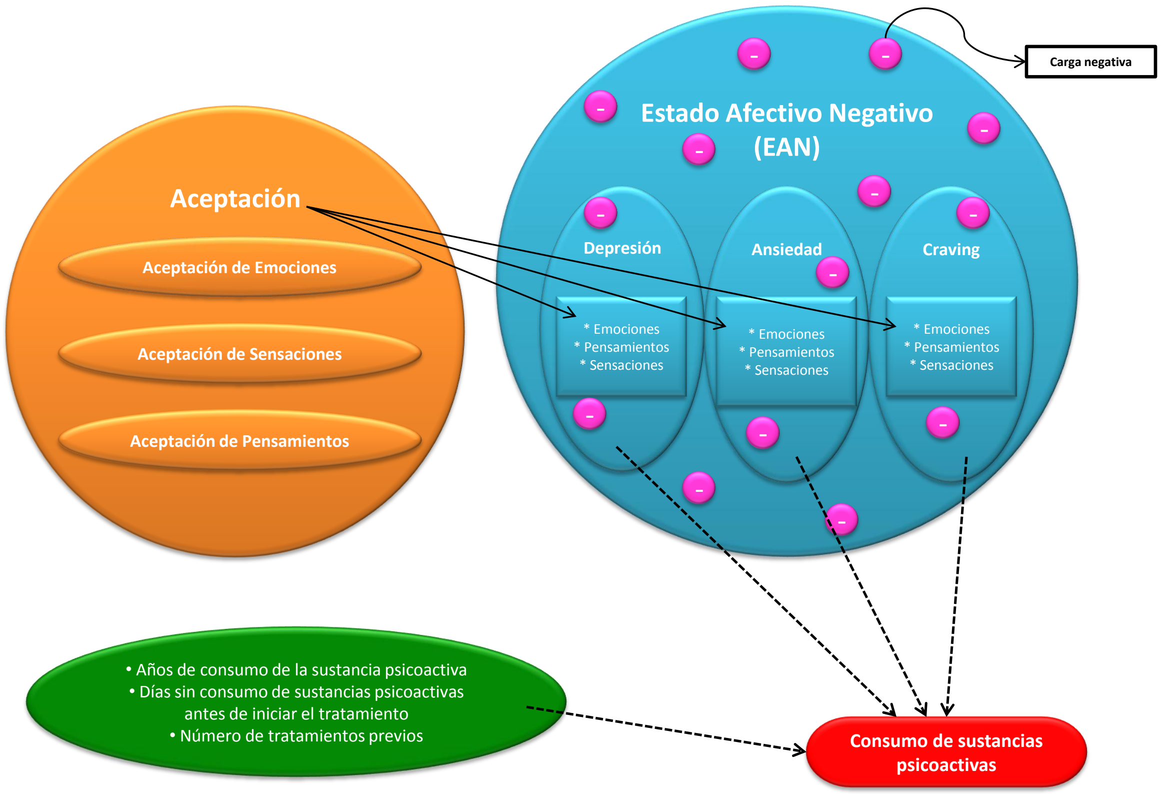

Figura 1-1: Variables y sus Agrupamientos Conceptuales

variable nivel de consumo (en rojo), pues ésta será nuestra principal variable dependiente.

Cada una de las variables mencionadas posee su propio instrumento y escala de medición. Por cuestiones prácticas la mayoría de las escalas se consideraran continuas (ver Tabla 1-1). Observemos que el nivel de consumo semanal puede oscilar dentro de un rango de valores muy amplio (0–300) debido al tipo de sustancia; por ejemplo, el consumo de cocaína podría no alcanzar un máximo de 10 gramos a la semana ya que dosis superiores podrían provocar en algunos casos hasta la muerte, mientras que una persona con problemas de consumo de alcohol fácilmente podría rebasar las 10 copas y llegar a un máximo de 200 o más en una semana. Las sustancias consideradas son alcohol, cocaína, mariguana, piedra y pvc.

Para poder contestar si el nuevo tratamiento realmente es eficaz el experto requiere hacer el

	Variables	***Instrumento medición***	***Escala de medición***	***Tipo de escala***
Aceptación	* Aceptación de Pensamientos * Aceptación de Sensaciones * Aceptación de Emociones	Prueba psicológica (Cuestionario)	9-45	Continua
Estado de ánimo negativo	* Craving * Ansiedad * Depresión	Prueba psicológica (Cuestionario)	* 0-28 * 0-63 * 0-63	Continua
Variables inherentes al paciente	* Tratamientos previos * Días sin consumo * Años de consumo	Datos del Paciente	* 1-3 * 1-90 * 3-50	Discreta
Variable dependiente	Consumo	Paciente	1-300	Continua

Tabla 1-1: Descripción de Variables

trabajo de campo y tomar las mediciones de las características que ha considerado importantes, previamente mencionadas. Surge así la siguiente pregunta «¿Cuántos pacientes se deberán considerar en el estudio?» de tal forma que la cantidad de datos sea suficiente para poder afirmar con cierta confiabilidad que los resultados obtenidos (después de hacer el correspondiente análisis estadístico) son válidos. Una vez dicho que se medirán en repetidas ocasiones los niveles de las diferentes variables sobre un mismo paciente, la segunda pregunta que surge es «¿Cuántas mediciones por paciente tendrían que realizarse?». De la misma forma y para ayudar a responder estas dos últimas interrogantes, el experto pregunta lo siguiente «¿De qué forma se podría establecer una relación entre todas las variables que el estudio considera?» o, en otras palabras, «¿Cuál sería el modelo estadístico que mejor represente la interacción entre las variables del estudio?». Para ofrecer una solución, el estadístico requiere retomar la teoría de los modelos existentes en la estadística que mejor compaginen con las características del estudio psicológico.

Con base en experiencias propias el experto ha propuesto un número de pacientes (10) y un número de observaciones por paciente (7) para la realización del estudio. Sin embargo, desconoce si un trabajo con estas características bridará la información suficiente para afirmar, con cierta precisión, si los posibles cambios observados en los niveles de consumo serán significativos, y más aún, para poder decir si el nivel de consumo se relaciona con las variables que se manipulan en el tratamiento propuesto.

Por su parte, la teoría estadística nos dice que los modelos de regresión permiten analizar la relación existente entre un conjunto de variables. En el caso más simple se modela la relación lineal entre dos variables. La teoría al respecto es muy amplia ya que ha sido desarrollada desde los primeros años del siglo XIX. Estos modelos de regresión implican la estimación de parámetros. Parámetros estadísticamente significativos permitirán comprobar la asociación existente entre variables. Junto con el desarrollo teórico se ha ido perfeccionando el software que permite hacer los correspondientes cálculos de una forma eficiente y rápida, como lo es el caso de las librerías existentes en el software estadístico R. Todo parece indicar que ésta es la dirección que seguiremos para comenzar a buscar una solución a la problemática propuesta.

Capítulo 2

Marco Teórico

En el capítulo anterior un experto en salud mental nos ha planteado un importante problema a abordar. El proporcionar una solución implicaría, entre otras cosas, mostrar evidencia de la funcionalidad de la integración de nuevos componentes en las terapias tradicionales empleadas en el tratamiento que se brinda a personas consumidoras de drogas en México. El propósito primordial de esta tesis es la modelación estadística de los datos a producirse mediante el estudio propuesto por el investigador. Así, el primer paso a dar consiste en identificar nociones fundamentales que apunten a un determinado modelo estadístico. Como se verá, la idea será recurrir a una clase muy versátil de modelos estadísticos.

Las características del tratamiento y de las variables, tal como fueron mostradas en el capítulo anterior comienzan a sugerir algunos puntos importantes a considerar. Por un lado, tenemos que se realizarán mediciones sobre un mismo paciente a lo largo del tiempo que dure el tratamiento, debido a lo cual podemos afirmar que el conjunto de observaciones provenientes de una misma persona no son independientes entre sí. Por otro lado, con base en estudios psicológicos (por ejemplo, Jonathan Bricker *et al.*, 2013) el experto estableció la relación existente entre las variables involucradas en el estudio. Se espera que conforme avance el tiempo la aceptación de pensamientos, sensaciones y emociones de los pacientes vayan en aumento. El incremento o decremento en los niveles de ansiedad (ocurriendo algo similar para el caso de depresión y *craving*) está en función del grado de aceptación de pensamientos, sensaciones y emociones. El nivel de consumo de una persona depende de su grado de ansiedad, depresión y *craving*, así como de los años que llevaba consumiendo drogas, de los días que tenía sin consumir la sustancia justo antes de iniciar el tratamiento y del número de tratamientos previos que ha tenido.

Podemos observar que todas las variables están relacionadas entre sí. En primera instancia podríamos sugerir un único modelo; sin embargo, debido a que el investigador ha expuesto esta organización claramente segmentada en tres partes es que hemos llegado a la formulación de tres modelos diferentes pero relacionados entre sí. Esta forma de organización es conocida estadísticamente como modelación jerárquica.

Un modelo lineal con efectos mixtos o modelo lineal mixto, explicado ampliamente en Demidenko (2004) y Jiang (2007) es, en términos generales, un modelo estadístico que permite trabajar con datos agrupados o bien datos longitudinales. Este tipo de modelos considera dos fuentes de variación en los datos: la variación entre los distintos grupos o individuos y la variación entre las observaciones de un mismo grupo o individuo. El modelo lineal con efectos mixtos es, probablemente, como lo menciona Fitzmaurice, *et al.* (2009), el método más usado para el análisis de datos longitudinales. El empleo de estos modelos en otras ciencias se remota a los 80's en un artículo publicado por Laird and Ware. Los métodos estadísticos para el análisis de datos longitudinales así como la regresión lineal clásica y el método de mínimos cuadrados tienen sus primeros orígenes en el área de la astronomía.

Los modelos lineales mixtos comparten las características del estudio en cuestión, por lo cual se concluyó que un modelo de este tipo sería una adecuada representación del fenómeno psicológico. En una primera instancia se han de considerar modelos lineales debido a la facilidad de implementación de herramientas teóricas y de software ya existentes. Se deja para trabajos futuros el análisis del mismo problema con modelos no lineales, el cual puede ser tratado con ayuda de la librería *nlme* del software estadístico R.

Los modelos lineales con efectos mixtos reciben este nombre debido a que involucran efectos fijos y efectos aleatorios al mismo tiempo. Demidenko (2004) menciona que a partir de esto último es que en estos modelos se distinguen dos tipos de coeficientes: los promedios poblacionales y aquellos que son específicos de cada grupo o sujeto. Los primeros son constantes desconocidas por estimar como en un modelo de regresión lineal múltiple, mientras que los segundos son variables aleatorias de quienes se estiman los parámetros que definen su distribución.

En el presente capítulo conoceremos un poco más de los modelos lineales mixtos. Mostraremos la teoría necesaria para el cálculo de los estimadores puntuales de parámetros y de los intervalos de confianza de manera similar a cuando se estudia el modelo de regresión lineal múltiple.

2.1. Modelo lineal con efectos mixtos

En estudios médicos, sociales, biológicos, *etc.* en algunas ocasiones se presentan mediciones correlacionadas entre sí ya que son observaciones realizadas sobre la misma persona o sobre una determinada subpoblación. Por ejemplo, datos tomados de diferentes camadas de ratón son independientes; mientras que las observaciones hechas en una sola camada son dependientes dado que los miembros pertenecen a una misma madre, de donde los sujetos comparten características similares tanto genéticas como de alimentación.

Si es posible representar a un fenómeno natural con un modelo lineal mixto, el establecimiento de los efectos fijos y aleatorios depende directamente de la situación a analizar. Si por ejemplo consideramos variables como edad, estado civil, sexo, *etc.* normalmente éstas suelen ser clasificadas como efectos fijos. Asimismo a cada sujeto o grupo involucrados en el experimento se le puede asociar un efecto aleatorio cuyo valor no es observable. Dicho efecto puede deberse, por ejemplo, a que cada ente es tomado al azar de la población y al inicio del estudio cuentan con sus niveles particulares de las variables en cuestión; o a que cada individuo o grupo reacciona de forma diferente y única ante ciertos eventos a lo largo de un determinado tiempo; *etc.*

Si consideramos un estudio que cuenta con N individuos o grupos y de cada uno de ellos n_i observaciones, entonces el correspondiente modelo lineal con efectos mixtos queda matemáticamente expresado como lo hicieron Laird y Ware en 1982 de la siguiente forma:

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \varepsilon_i, \qquad i = 1, ..., N, \tag{2-1}$$

donde:

- $\mathbf{y}_i$ vector de respuestas del i-ésimo individuo o grupo, de tamaño $n_i \times 1$.
- $\mathbf{X}_i$ matriz de covariables o efectos fijos, de tamaño $n_i \times m$.
- β vector de parámetros poblacionales ó coeficientes de los efectos fijos, de tamaño $m \times 1$.

- $\mathbf{Z}_i$ matriz de efectos aleatorios, de tamaño $n_i \times k$.
- ε_i vector de términos de error, cada uno de los cuales tiene media cero y varianza σ^2, de tamaño $n_i \times 1$.
- $\mathbf{b}_i$ vector de efectos aleatorios con media cero y matriz de covarianza $\mathbf{D}_* = \sigma^2\mathbf{D}$, de tamaño $k \times 1$. En este caso la dimensión de $\mathbf{D}$ es $k \times k$.

Se asume que todos los vectores aleatorios $\{\mathbf{b}_i,\ \varepsilon_i,\ i = 1, ..., N\}$ son mutuamente independientes. En el presente trabajo se considera que estos vectores tienen distribución normal debido a la simplificación de los cálculos correspondientes para la obtención de los estimadores de los parámetros involucrados en el modelo lineal mixto. Así, con $\mathbf{I} = \mathbf{I}_{n_i}$ la matriz identidad de tamaño $n_i \times n_i$:

$$\varepsilon_i \sim \mathrm{N}\left(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I}\right), \qquad \mathbf{b}_i \sim \mathrm{N}\left(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{D}\right).$$

Dado el supuesto de distribución normal de los vectores aleatorios, mencionado previamente, Demidenko (2004) presenta una forma de expresar al modelo (2-1). Ésta nos dice de la distribución de cada uno de los vectores respuesta de cada sujeto o grupo en estudio:

$$\mathbf{y}_i \sim \mathrm{N}\left(\mathbf{X}_i\beta, \sigma^{\mathbf{2}}\left(\mathbf{I} + \mathbf{Z}_i\mathbf{D}\mathbf{Z'}_i\right)\right), \qquad i = 1, ..., N. \tag{2-2}$$

Observemos que N ecuaciones de la forma (2-1) pueden ser expresadas en una sola ecuación como a continuación se muestra. Para poder ver en forma más detallada la expansión de cada uno de los elementos integrantes de la ecuación se recomienda consultar Demidenko (2004),

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \varepsilon,$$

donde

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}, \qquad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{Z}_N \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{bmatrix}, \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}.$$

Al respecto de la notación, como podemos percatarnos, a lo largo del texto se empleará el formato en minúsculas negritas para denotar vectores y mayúsculas negritas para matrices. Las letras romanas minúsculas indicarán efectos aleatorios.

Identificabilidad del Modelo

Es importante que un modelo estadístico sea identificable para poder realizar una inferencia adecuada. Cuando sucede que un modelo no cuenta con esta característica resulta imposible estimar consistentemente los parámetros a partir de los datos. El límite de un estimador consistente tiene que ser único.

Formalmente, como lo menciona Cheng & Van Ness (1999), si $\mathbf{Z}$ es un vector aleatorio cuya distribución pertenece a una familia $F = \{F_\theta; \theta \in \Theta\}$, entonces el parámetro θ_i, el i-ésimo componente del vector θ es identificable si y sólo si no existen dos $\theta \in \Theta$ tal que si el i-ésimo componente es diferente hagan que $\mathbf{Z}$ tenga la misma distribución. Diremos que el vector de parámetros θ es identificable si y sólo si todos sus componentes son identificables. Además, se dice que el modelo es identificable si θ es identificable.

Con acuerdo en Demidenko (2004) para que el modelo lineal mixto sea identificable para β, la matriz $\sum \mathbf{X}_i'\mathbf{X}_i$ debe ser no singular y $\sum n_i > m$. Lo anterior significa que no debe haber multicolinealidad entre las $\mathbf{X}_i$, además el número de variables debe ser menor que el número de observaciones. Para que el mismo modelo sea identificable para σ^2 y $\mathbf{D}$, al menos una matriz $\mathbf{Z}_i'\mathbf{Z}_i$ debe ser definida positiva y $\sum_{i=1}^{N}(n_i - k) > 0$.

2.2. Inferencia para el modelo lineal con efectos mixtos

La muestra aleatoria es una pieza clave en la estadística inferencial. Consiste de un conjunto de sujetos o grupos representativos de la población en estudio. En estadística, como lo menciona

Hodges & Lehmann (1970), las cantidades a ser estimadas son los parámetros de los modelos probabilísticos o bien alguna función de los mismos. La información disponible, para calcular cada uno de los estimadores, consiste en una serie de datos tomados de la muestra aleatoria.

Casella & Berger (2002) mencionan que las estimaciones bien pueden ser puntuales o por intervalo. Un estimador puntual, como su nombre lo indica, asigna al parámetro correspondiente un único valor y es función de una muestra aleatoria. Se conocen diferentes métodos para la estimación de parámetros puntuales, dentro de los cuales podemos mencionar: método de momentos, método de máxima verosimilitud, método de mínimos cuadrados, *etc.* Por otro lado, los estimadores por intervalo están determinados por una pareja de números. Cuando esta segunda forma de estimación ofrece una cierta confianza de cubrir al parámetro poblacional de interés se conoce como intervalo de confianza. La probabilidad de que efectivamente el intervalo cubra al parámetro real se conoce como nivel de confianza y se expresa con $1 - \alpha$, donde α es conocido como nivel de significancia y representa la probabilidad de que el intervalo no cubra al parámetro. La información obtenida de la muestra estadística es la que se emplea para los cálculos correspondientes, igual que en el caso anterior.

En el caso de los modelos lineales con efectos mixtos, la matriz de varianzas y covarianzas de los efectos aleatorios $\mathbf{D}$ es desconocida, así que establecer un estimador para ella es de gran importancia dentro de la teoría de estos modelos. Si esta matriz es conocida, entonces, es posible calcular un estimador para los efectos fijos empleando mínimos cuadrados generalizados ó ponderados. En esta sección presentaremos el desarrollo de estas dos ideas considerando el método de máxima verosimilitud.

2.2.1. Función de Log-verosimilitud y estimadores puntuales

La función de verosimilitud, como nos explican Casella & Berger (2002), es función de los parámetros de un modelo estadístico y es un estadístico importante que resume toda la información presente en la muestra aleatoria. Este mismo ente se emplea en el método más popular para el cálculo de estimación de parámetros: método por máxima verosimilitud.

En el método de estimación por máxima verosimilitud se calcula el punto en el cual la función de verosimilitud alcanza su valor máximo, estableciendo este valor como el estimador puntual del parámetro poblacional. Algunas veces este cálculo puede llegar a ser complejo debido a lo cual se emplea el logaritmo de la función de verosimilitud. El estimador así obtenido será el parámetro que con mayor probabilidad nos dará la muestra observada.

La función de log-verosimilitud para el modelo lineal con efectos mixtos, sin tomar en cuenta el término constante $C = -(N_T/2)\ln(2\pi)$ con $N_T = \sum_{i=1}^{N} n_i$, tal como es mostrada por Demidenko (2004) queda escrita como sigue:

$$l(\theta) = -\frac{1}{2}\left\{N_T \ \ln\sigma^2 + \sum_{i=1}^{N}\left[\ln\left|\mathbf{I}+\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'\right|\right.\right.$$
$$\left.\left. + \sigma^{-2}(\mathbf{y}_i - \mathbf{X}_i\beta)'(\mathbf{I}+\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i')^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)\right\}\right., \quad \text{(2-3)}$$

donde $\theta = \left(\beta', \sigma^2, \text{vech}'\left(D\right)\right)$, y vech($\mathbf{D}$) es el vector de elementos de la matriz simétrica $\mathbf{D}$ de tamaño $k(k+1)/2$. La dimensión del vector de parámetros θ es $m+1+k(k+1)/2$. El estimador obtenido por máxima verosimilitud maximiza la función l sobre el espacio de parámetros

$$\Theta = \left\{\theta : \beta^m, \sigma^2 > 0, \mathbf{D} \ \text{ es definida no negativa}\right\}. \quad \text{(2-4)}$$

El teorema mencionado a continuación, explicado ampliamente en Demidenko (2004), sugiere que el estimador por máxima verosimilitud en el modelo lineal con efectos mixtos existe con probabilidad 1 si el número total de observaciones es suficientemente grande, particularmente si $\sum(n_i - k) - m > 0$. Sin embargo, la existencia del estimador por máxima verosimilitud no garantiza que la matriz $\hat{\mathbf{D}}_{\mathbf{ML}}$ sea definida positiva.

Teorema 1 *Los estimadores por máxima verosimilitud para el modelo con efectos mixtos (2-2) sobre el conjunto de parámetros (2-4) existen si y sólo si se cumple*

$$S_{min} = \sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}_\infty)'(\mathbf{I} - \mathbf{Z}_i\mathbf{X}_i^+)(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}_\infty) > 0,$$

donde $\hat{\beta}_\infty$ es el estimador por mínimos cuadrados ordinarios de los efectos fijos.

Parametrización de la función de log-verosimilitud

Demidenko (2004) nos presenta tres formas diferentes de parametrizar a la función (2-3): parametrización empleando reducción de dimensión, verosimilitud perfil y parametrización empleando la matriz inversa (o matriz de precisión) de $\mathbf{D}$. La parametrización permitirá, entre otras cosas, que los correspondientes cálculos para la obtención de estimadores sean más sencillos.

Parametrización empleando reducción de dimensión. La función de log-verosimilitud tal como se presenta en la ecuación (2-3) involucra la obtención de la inversa de una matriz de dimensión $n_i \times n_i$ y el cálculo de un determinante. Es posible reducir la dimensión de la matriz a k con el debido manejo de las fórmulas de reducción de dimensión (ver Demidenko, 2004). Recordando que n_i es el número de observaciones por sujeto y k es el número de efectos aleatorios.

Empleando las fórmulas de reducción de dimensión, la función (2-3) puede escribirse como sigue:
$l(\theta) = -\frac{1}{2}\left\{N_T \ \ln\sigma^2 + \sum_{i=1}^{N} \ln|\mathbf{I} + \mathbf{D}\mathbf{Z}_i'\mathbf{Z}_i| + \sigma^{-2}\sum_{i=1}^{N}\left[S_i - (\mathbf{Z}_i'\mathbf{e}_i)'(\mathbf{I} + \mathbf{D}\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{D}(\mathbf{Z}_i'\mathbf{e}_i)\right]\right\}$,
donde $\mathbf{e}_i = \mathbf{e}_i(\beta) = \mathbf{y}_i - \mathbf{X}_i\beta$ es el vector de residuos de tamaño $n_i \times 1$ para $i = 1, ..., N$; y $S_i = S_i(\beta) = \mathbf{e}_i'\mathbf{e}_i = \|\mathbf{y}_i - \mathbf{X}_i\beta\|^2$ es la suma de los residuos al cuadrado del i-ésimo sujeto o grupo e $\mathbf{I}$ es la matriz identidad de tamaño $k \times k$.

Si en el modelo (2-2) se mantiene constante a la matriz $\mathbf{D}$ sucede que l alcanza su máximo en el estimador por mínimos cuadrados generalizados (GLS), mostrado a continuación:

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{I} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i')^{-1}\mathbf{X}_i\right]^{-1}\left[\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{I} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i')^{-1}\mathbf{y}_i\right]. \tag{2-5}$$

Verosimilitud perfil y parametrización empleando la matriz inversa (o matriz de precisión) de D. El vector óptimo de efectos fijos β y la varianza σ^2 se pueden expresar empleando la matriz $\mathbf{D}$, de esta forma pueden ser eliminados de la función de log-verosimilitud. Consideraremos dos funciones perfil: perfil varianza y perfil total. Por otra parte, en la parametrización que involucra reducción de dimensión observamos que la función de log-verosimilitud puede expresarse a través de $\mathbf{D}^{-1}$. Computacionalmente la mejor opción de parametrización será la que involucra a la matriz inversa de $\mathbf{D}$ y a la función perfil total.

Si calculamos la derivada de (2-3) respecto a σ^2, la función l alcanza su máximo en:

$$\hat{\sigma}^2 = \frac{1}{N_T}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{X}_i\beta)'(\mathbf{I} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i')^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta). \tag{2-6}$$

Sustituyendo (2-6) en la fórmula (2-3) tenemos la función de log-verosimilitud perfil varianza, obteniendo así un problema de maximización equivalente pero sin que aparezca σ^2, tal como se muestra en la siguiente fórmula, donde p indica que la función es perfil,

$$l_p(\beta, \mathbf{D}) = -\frac{1}{2}\left\{N_T \ln \sum_i (\mathbf{y}_i - \mathbf{X}_i\beta)'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta) + \sum_i \ln|\mathbf{V}_i|\right\}, \tag{2-7}$$

donde $\mathbf{V}_i = \mathbf{V}_i(\mathbf{D}) = \mathbf{I} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'$.

A partir de la fórmula (2-7) y de las hipótesis de la parametrización que emplea reducción de dimensión es posible llegar a la siguiente parametrización de la función perfil (ver Demidenko, 2004), con $\mathbf{r}_i = \mathbf{Z}_i'\mathbf{e}_i$ y $\mathbf{M}_i = \mathbf{Z}_i'\mathbf{Z}_i$,

$$l_p(\beta, \mathbf{D}) = -\frac{1}{2}\left\{N_T \ln\left(\sum_i [S_i - \mathbf{r}_i'(\mathbf{D}^{-1} + \mathbf{M}_i)^{-1}\mathbf{r}_i]\right) + \sum_i \ln|\mathbf{I} + \mathbf{D}\mathbf{M}_i|\right\}. \tag{2-8}$$

Ahora bien, si consideramos $\mathbf{D}_{\text{-}} = \mathbf{D}^{-1}$ obtenemos la parametrización que emplea a la matriz de precisión:

$$l_p(\beta, \mathbf{D}_{\text{-}}) = -\frac{1}{2}\left\{N_T \ln\left(\sum_i [S_i - \mathbf{r}_i'(\mathbf{D}_{\text{-}} + \mathbf{M}_i)^{-1}\mathbf{r}_i]\right) + \sum_i \ln|\mathbf{D}_{\text{-}} + \mathbf{M}_i| - N\ln|\mathbf{D}_{\text{-}}|\right\}. \tag{2-9}$$

Otra posibilidad consiste en excluir a β de la función de log-verosimilitud empleando el estimador expresado en (2-5). Si $\mathbf{V}_i$ es conocida, definida como $\mathbf{V}_i = \mathbf{V}_i(\mathbf{D}) = \mathbf{I} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'$, matriz de covarianzas del vector asociado a la variable dependiente $\mathbf{y}_i$, obtenemos el estimador que se muestra abajo. Esta es otra forma de plantear mínimos cuadrados ponderados.

$$\hat{\sigma}^2_{GLS} = {}^{min}_{\beta}\frac{1}{N_T}\sum(\mathbf{y}_i - \mathbf{X}_i\beta)'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta) = \frac{q}{N_T}.$$

De esta forma, con $q = {}^{min}_{\beta}\sum(\mathbf{y}_i - \mathbf{X}_i\beta)'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)$, la función de log-verosimilitud perfil total está dada por:

$$l_P(\mathbf{D}_{\text{-}}) = \frac{1}{2}\left[N_T \ln q(\mathbf{D}_{\text{-}}) + \sum_i \ln|\mathbf{D}_{\text{-}} + \mathbf{M}_i| - N\ln|\mathbf{D}_{\text{-}}|\right],$$

siendo esta última expresión de la función de log-verosimilitud, posiblemente, la mejor opción a emplear para la realización de cálculos numéricos. La parametrización perfil-verosimilitud es utilizada en la construcción de intervalos de confianza.

Algoritmos de maximización

Para obtener el máximo de la función de log-verosimilitud y poder asignar el correspondiente valor numérico a los parámetros en cuestión, lo primero que debemos calcular son las derivadas parciales respecto a los parámetros de la función (2-2). Con acuerdo en Demidenko (2004) para maximizar la función de log-verosimilitud existen tres algoritmos comúnmente empleados en estadística: EM por sus siglas en inglés (*Expectation-Maximization*), puntajes de Fisher y Newton-Raphson. Estos algoritmos tienen la siguiente formulación general

$$\mathbf{t}_{s+1} = \mathbf{t}_s + \lambda_s \delta_s, \qquad s = 0, 1, ...,$$

donde $\mathbf{t}_s$ es el estimador máximo verosímil en la s-ésima iteración, $0 < \lambda_s \leqslant 1$ es la longitud de avance y δ_s es el vector dirección, calculado como $\delta_s = \mathbf{H}_s^{-1}\mathbf{g}_s$, donde $\mathbf{g}_s = \left.\frac{\partial l}{\partial \mathbf{t}}\right|_{\mathbf{t}=\mathbf{t}_s}$ es el gradiente de la función de log-verosimilitud l, y $\mathbf{H}_s$ es una matriz definida positiva.

La diferencia entre los tres algoritmos mencionados anteriormente es la matriz $\mathbf{H}$ (ver Demidenko, 2004). Para el algoritmo Newton-Raphson, $\mathbf{H}$ es la matriz Hessiana negativa (la matriz negativa de las segundas derivadas $-\frac{\partial^2 l}{\theta^2}$); para el algoritmo puntajes de Fisher, $\mathbf{H}$ es la matriz esperada del Hessiano $-\frac{\partial^2 l}{\theta^2}$. Para asegurar la convergencia, la matriz $\mathbf{H}$ debe ser definida positiva. Para el algoritmo Newton-Raphson, $\mathbf{H}$ deberá ser definida positiva en una vecindad del máximo. Para el algoritmo puntajes de Fisher esta matriz es definida positiva, siempre que el modelo estadístico sea especificado correctamente. Para el algoritmo EM, la matriz $\mathbf{H}$ es también definida positiva y $\lambda_s = 1$ es fijo. Para este último algoritmo la función de log-verosimilitud incrementa al pasar de una iteración a otra, $l(\mathbf{t}_{s+1}) > l(\mathbf{t}_s)$ si el gradiente en $\mathbf{t}_s$ es diferente de cero, por lo que resulta redundante conocer λ_s. Lo anterior no ocurre para los otros dos algoritmos así que se requiere encontrar λ_s, empezando con $\lambda = 1$. En cada iteración existe λ_s positiva que permite un incremento en la log-verosimilitud si $\mathbf{g}_s \neq 0$. Comúnmente para el caso de los modelos con efectos mixtos, la no convergencia está asociada con el hecho de que algunas entradas en la diagonal de la matriz $\mathbf{D}$ son muy cercanas a cero, es decir, la matriz $\mathbf{D}$ no es definida positiva.

2.2.2. Intervalos de confianza

Para los modelos lineales mixtos es posible obtener intervalos de confianza tanto para los parámetros asociados a los efectos fijos como para las diferentes varianzas involucradas en el modelo (ver Jiang, 2007). En esta sección mostraremos los intervalos de confianza cuya construcción toma en cuenta los supuestos de normalidad que ya hemos mencionado al inicio de este capítulo.

Es importante mencionar que no existe un método analítico que permita obtener intervalos de confianza exactos para la varianza de los efectos aleatorios, sin embargo, es posible encontrar intervalos de confianza aproximados. Satterthwaite en 1946 propuso un método donde el objetivo era construir un intervalo de confianza para $\zeta = \sum_{i=1}^{h} c_i\lambda_i$, donde $\lambda_i = \mathrm{E}(S_i^2)$ y S_i^2 es la suma de cuadrados medios correspondientes al i-ésimo factor (fijo o aleatorio) en el modelo. Este método trata de encontrar el número de grados de libertad, d, tal que los primeros dos momentos de la variable aleatoria $d\sum_{i=1}^{h} c_i S_i^2/\zeta$ se igualen con los de una variable aleatoria χ_d^2. Graybill y Wang en 1980 propusieron un método para mejorar el de Satterthwaite. Este nuevo procedimiento permite calcular el intervalo de confianza aproximado para una combinación lineal no negativa de λ_i's.

2.3. Casos especiales del modelo lineal con efectos mixtos

Dos casos especiales de los modelos lineales con efectos mixtos son: los modelos lineales mixtos con pendientes aleatorias o coeficientes de regresión aleatorios y los modelos lineales mixtos con intercepto aleatorio. Bajo las debidas hipótesis ambos modelos pueden llegar a ser más particulares, conocidos como modelos balanceados. Estas situaciones tan específicas son la base del modelo presentado en el Capítulo 3.

Un modelo lineal con efectos mixtos se dice balanceado si n_i el número de observación por sujeto, es un valor constante para todo i y las matrices de efectos aleatorios $\mathbf{Z}_i$ son iguales para todos los sujetos o grupos, es decir, $\mathbf{Z}_i = \mathbf{Z}$. Considerar modelos de este tipo es de gran importancia debido a la simplicidad que se puede llegar a tener en los cálculos de los estimadores de parámetros por máxima verosimilitud.

En esta sección presentaremos los dos modelos ya indicados, serán expresados matemáticamente en la misma forma que lo hace Demidenko (2004). En primera instancia consideraremos un par de ejemplos a manera de ilustración; en ellos la relación lógica entre variables se adoptó por razones de conveniencia.

Consideremos el modelo peso *vs.* talla. En un primer experimento tenemos cinco crías de ratón, cada una de ellas de camadas diferentes. Las crías se eligen de tal forma que su peso y talla sea muy similar (casi la misma). Así al inicio del experimento todos los ratones parten de un mismo punto (peso, talla). A lo largo de seis semanas a partir del nacimiento se tomarán mediciones de ambas variables sobre cada uno de los cinco ratones, una medición por semana. El efecto aleatorio que asociamos a cada cría será debido a que son pertenecientes a familias diferentes, es decir, su desarrollo, reflejado en el incremento de peso y talla será distinto debido a que cada camada se desenvuelve bajo circunstancias diferentes. Este experimento queda expresado matemáticamente en cinco rectas que tienen el mismo intercepto pero diferentes pendientes, como se observa en la gráfica de la izquierda de la Figura 2-1. Estadísticamente este modelo es llamado modelo lineal mixto con pendientes aleatorias.

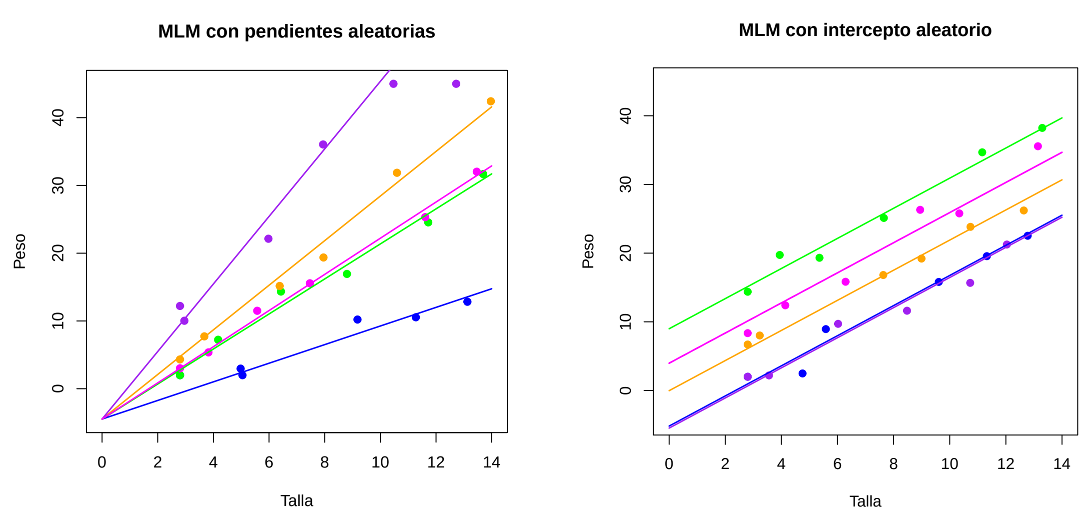

Figura 2-1: Casos especiales del Modelo Lineal Mixto.

Por otra parte, el segundo experimento considera cinco crías pero de una misma camada.

De igual forma que en el caso anterior se tienen seis parejas de observaciones por cada ratón a partir del momento del nacimiento. Al inicio cada ratón tendrá un peso y una talla diferentes debido a que pertenecen a un mismo grupo. Asimismo, al pertenecer a la misma camada se desarrollarán de forma muy parecida por similares condiciones de cuidado y alimento con lo cual su incremento de talla y peso será semejante para cada uno de los cinco animales. Lo anterior se verá reflejado en cinco rectas con diferentes interceptos pero misma pendiente, como se observa en la gráfica de la derecha de la Figura 2-1. Estadísticamente este modelo es llamado modelo lineal mixto con intercepto aleatorio.

Modelo lineal mixto con coeficientes de regresión aleatorios

El modelo lineal mixto con coeficientes de regresión aleatorios se presenta al hacer $\mathbf{Z}_i = \mathbf{X}_i$ en (2-1). Es expresado por Demidenko (2004) tal como se muestra a continuación:

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{a}_i + \varepsilon_i, \qquad \mathbf{a}_i = \beta + \mathbf{b}_i, \qquad i = 1, ..., N.$$

Un caso muy particular es el modelo balanceado con pendientes aleatorias. En esta situación, todos los grupos de datos tienen el mismo número de observaciones, es decir, $n_i = n$, además $\mathbf{Z} = \mathbf{X}_i = \mathbf{Z}_i \ \forall i = 1, ..., N$. Una de las ventajas de trabajar con este modelo es que las hipótesis permiten que los cálculos para la obtención de los estimadores máximo-verosímiles sean más sencillos.

Modelo lineal mixto con intercepto aleatorio

Otro caso especial dentro de la teoría de los modelos lineales con efectos mixtos es aquel que tiene intercepto aleatorio. Este modelo puede ser expresado como sigue:

$$y_{ij} = a_i + X_{ij}\beta + \varepsilon_{ij}, \qquad j = 1, ..., n_i, \qquad i = 1, ..., N,$$

donde y_{ij} es la j-ésima observación del i-ésimo individuo. El intercepto individual es la suma del parámetro (intercepto) poblacional α y un efecto aleatorio, $a_i = \alpha + b_i$. Se asume que $\varepsilon_{ij} \sim \mathrm{N}\left(0, \sigma^2\right)$ y $b_i \sim \mathrm{N}\left(0, \sigma^2 d\right)$ son independientes, donde σ^2 es la varianza entre las observaciones de cada sujeto y d es la varianza escalada del efecto aleatorio.

En notación matricial y dadas las distribuciones ya dichas para ε_{ij} y b_i, en el modelo lineal mixto con intercepto aleatorio la distribución de la respuesta puede ser escrita como sigue:

$$\mathbf{y}_i \sim \mathrm{N}\left(\mathbf{X}_i\beta, \boldsymbol{\sigma^2}\left(\mathbf{I}_i + d\mathbf{1}_i\mathbf{1'}_i\right)\right), \qquad i = 1, ..., N,$$

donde $\mathbf{X}_i$ es la matriz diseño de tamaño $n_i \times m$, β es el vector de efectos fijos de tamaño $m \times 1$, $\mathbf{1}_i$ es un vector de unos de tamaño $n_i \times 1$, $\boldsymbol{\sigma^2}$ es la varianza entre las observaciones de cada individuo o grupo de datos y $\boldsymbol{\sigma^2}d$ es la varianza de los efectos aleatorios.

Un caso muy particular de este modelo se presenta cuando $n_i = n$ y $\mathbf{X}_i = \mathbf{X}$, es decir cuando tenemos el mismo número de observaciones por individuo y las mismas covariables medidas sobre cada uno de los sujetos. Este último es conocido como modelo lineal mixto balanceado con interceptos aleatorios.

2.4. Métodos computacionales en R para el modelo lineal con efectos mixtos

Los cálculos requeridos para la obtención de los estimadores de los parámetros involucrados en un modelo lineal con efectos mixtos, en algunas ocasiones pueden llegar a ser complicados, por lo cual se han buscado procedimientos lo más sencillos y óptimos posibles que permitan llegar al mismo resultado. Como ya lo menciona Bates (2011), uno de estos métodos computacionales resume la problemática a encontrar la solución de un problema de mínimos cuadrados ponderados.

Una de las librerías en el software estadístico R que emplea el procedimiento señalado es *lme*4 (ver Bates *et al.*, 2014). En esta librería encontramos funciones que permiten el análisis de datos cuyo comportamiento podría ser representado a través de modelos lineales con efectos mixtos, modelos lineales mixtos generalizados o bien modelos mixtos no lineales.

El presente trabajo nos exige revisar cómo especificar la estructura de un modelo lineal mixto con ayuda de la librería indicada. La función que revisaremos a continuación y que encontraremos en *lme*4 es *lmer()*. La composición de esta función permite concentrar toda la información contenida en un modelo lineal mixto, es decir, permite poder diferenciar claramente cuáles son los efectos fijos y cuáles los efectos aleatorios asociados al modelo.

Los resultados proporcionados por la función *lmer()* permiten conocer tanto estimadores puntuales como lo son interceptos, coeficientes asociados a los efectos fijos, varianzas de los efectos aleatorios, *etc.*; así como intervalos de confianza. En adición, la librería *lattice* permite hacer gráficos que muestran en una sola imagen, por ejemplo, el comportamiento de cada uno de los grupos de datos provenientes de cada sujeto involucrado en el estudio. Estas gráficas nos pueden ayudar a la determinación de los efectos aleatorios.

Una consideración importante antes de plantear el modelo con la función *lmer()* es tener bien definido si el modelo tiene efectos aleatorios anidados o bien efectos aleatorios cruzados. Faraway (2005) menciona que los primeros se presentan cuando los niveles de algún factor son únicos en los niveles de otro factor. Por otra parte, los efectos aleatorios cruzados se presentan cuando no son anidados, es decir, cuando todos los niveles de un factor se presentan en todos los niveles de otro factor. Es importante tener clara esta situación para poder marcarla con la debida notación empleada en la función.

2.4.1. Sintaxis de la función *lmer()*

En el texto de Galecki & Burzykowski (2013) se explica de forma general y clara la organización de la estructura de la función *lmer()*; a continuación mostraremos la información básica al respecto. La expresión completa (2-10) se compone de diversas partes separadas por una coma, entre las cuales podemos mencionar como principales a la fórmula y la base de datos. El resto de los componentes son más específicos y se pueden consultar en la sección de ayuda del software estadístico R.

```
lmer(formula, data=NULL, REML=TRUE, control=lmerControl(), start = NULL,
      verbose = 0L, subset, weights, na.action, offset, contrasts,...)  (2-10)
```

Por su parte, la fórmula consta de dos partes separadas por una tilde "∼". En el lado izquierdo tenemos un único término: la variable dependiente. El lado derecho puede consistir en más de un elemento, los cuales serán separados por un signo "+". En este segundo extremo tenemos dos clases de términos: los primeros son aquellos que permiten especificar los efectos fijos asociados al modelo y se escriben tal cual aparece el nombre de la variable en la base de datos. Los segundos términos especifican a los efectos aleatorios.

Precisemos al respecto de la notación empleada para establecer a los efectos aleatorios asociados al modelo lineal mixto. Cada elemento se escribirá entre paréntesis y estará dividido en dos partes separadas por una barra vertical "|". La expresión a la izquierda de la barra es la matriz modelo para los efectos aleatorios. El término a la derecha de la barra es una variable factor. Veamos algunos ejemplos, que combinándolos permiten expresar modelos lineales mixtos con características más específicas:

- `(1|g1)`. Indica interceptos aleatorios. Es un intercepto para cada uno de los niveles de la variable factor g1.
- `(z1|g1)`. Indica interceptos y pendientes aleatorias correlacionados y con varianzas diferentes.
- `(1|g1) + (0+z1|g1)`. El primer término indica interceptos aleatorios y el segundo pendientes aleatorias. Así indicamos interceptos y pendientes aleatorios independientes.
- `(1|g1) + (1|g1:g2)`. Indica un modelo con interceptos aleatorios anidados. En este caso `g2` anidado en `g1`.
- `(1|g1) + (1|g2)`. Indica un modelo con efectos aleatorios cruzados. Los niveles de la variable `g2` están cruzados con los niveles de la variable `g1`.

Una vez que se expresa a un modelo lineal mixto en términos de la notación empleada en la función *lmer()*, lo que sigue es conocer cómo extraer resultados específicos, por ejemplo, estimadores puntuales o intervalos de confianza. Conocer esta información nos permitirá poder concluir al respecto de nuestro estudio en cuestión y en dado caso hacer inferencia estadística.

Si asignamos a `modelo` el resultado de la función *lmer()*, es decir, `modelo <- lmer(...)`, entonces:

- `summary(modelo)`. Proporciona los estimadores puntuales de los parámetros asociados a los efectos fijos y a los efectos aletorios del modelo.
- `coef(modelo)`. Proporciona estimadores puntuales para cada uno de los modelos que representa el comportamiento de los datos pertenecientes a un mismo individuo o grupo.

- `ranef(modelo)`. Proporciona los efectos aleatorios estimados para cada uno de los modelos que representa el comportamiento de los datos pertenecientes a un mismo individuo o grupo.

- `confint(modelo, method="boot")`. Proporciona intervalos de confianza para los parámetros asociados a los efectos fijos y a los efectos aleatorios del modelo. En este caso se especifica que el método empleado en la construcción de los intervalos de confianza es *bootstrap*.

Capítulo 3

Modelo estadístico, simulador y procedimientos de estimación

El presente capítulo está dividido en tres partes. La primera abarca la formulación del modelo matemático; en su conjunto expresa cómo establecer de forma estadística la relación existente entre las variables involucradas en el fenómeno psicológico presentado en el Capítulo 1. La segunda parte describe el simulador de datos de las variables involucradas en el estudio, el cual será una herramienta importante pues con el análisis de la información simulada se podrán comprobar las hipótesis de modelación discernidas por el experto. Finalmente, la última sección muestra la rutina de estimación de parámetros involucrados en el modelo mediante el uso de la librería *lme4* disponible en el software estadístico R; estos procedimientos serán empleados de manera literal por el usuario una vez que cuente con los datos reales.

La información expuesta en el Capítulo 2 y la comunicación continua entre el estadístico y el experto permitieron construir el modelo estadístico que representa el comportamiento de las variables en consideración. Así, se logró caracterizar a cada una de las respectivas partes que conforman al modelo. Se concluyó que diferentes modelos lineales mixtos relacionados entre sí de una forma muy particular, captarían la esencia de los fenómenos psicológicos del estudio.

Con la ayuda de un simulador se puede llegar a conocer el alcance de la realización de un experimento concreto con determinadas características. Por ejemplo, en este caso, con el debido análisis de los datos simulados, el experto podrá comprobar qué pasaría si el trabajo de campo se realizará con el número de pacientes y número de observaciones por persona que él considera pudiese llegar a tener en una primera instancia. También podrá observar qué repercusiones, sobre las conclusiones finales, podría tener la modificación del comportamiento hipotético preestablecido de alguna variable.

La simulación de diferentes escenarios y la práctica de la inferencia estadística (procedimientos de estimación de parámetros e intervalos de confianza que se realizarán una vez que se tengan los datos experimentales) para cada uno de ellos nos permitirán llegar a un escenario aceptable, el cual después de ser analizado permitirá al experto establecer ciertas conclusiones en su área de estudio. Con base en este caso elegido se podrá construir una tabla de datos que presente los valores correspondientes para cada una de las variables previamente mencionadas en el Capítulo 1; se considerará a este conjunto de datos como una posible muestra real después de haber sido aprobada por el experto. Los valores así obtenidos se podrán presentar gráficamente permitiendo conocer de manera visual la evolución de cada una de las variables conforme el tiempo avanza y el tratamiento progresa.

La información que proporcionen los diferentes escenarios simulados dará paso a conocer con cierta precisión y confiabilidad el número de pacientes (N) que se requiere tomen el tratamiento y el número de observaciones (n) de cada una de las variables por individuo. Podremos concluir al respecto de esto último gracias a los ensayos de la inferencia estadística con los correspondientes datos simulados. Para este caso particular de estudio notemos que tendremos la misma cantidad de información por cada uno de los pacientes.

Las diferentes actividades realizadas en cada una de las tres secciones de este capítulo, en conjunto, han permitido al estadístico explicar al psicólogo al respecto de estimación puntual de parámetros, intervalos de confianza, efectos de la modificación de las características de los datos sobre los parámetros, *etc.* La retroalimentación continua de información entre ambos expertos ha sido un elemento clave en la formulación del modelo. Podría ser posible que no existiera este flujo de ideas. Sin embargo, como se podrá apreciar, el trabajo multidisciplinario permite construir el modelo estadístico de un fenómeno natural con una clara explicación del por qué de cada una de las partes que lo componen.

3.1. Modelo Estadístico

Cuando se tiene un problema a resolver y se ha detectado alguna teoría estadística que permita explicar el comportamiento del fenómeno en cuestión, hay que empezar primero por revisar el modelo más básico de dicha metodología. Algunas ocasiones los mejores modelos no necesariamente son los más complejos. Esta sugerencia nos guiará a modelos más sofisticados si es que así se requiere.

En esta sección abordaremos una parte esencial del presente trabajo: el planteamiento en lenguaje matemático del problema expuesto por el psicólogo. Recordando del Capítulo 1, al experto le interesa establecer una relación entre tres grupos de variables. El primero conformado por aquellas que manipula con su tratamiento: aceptación de pensamientos, sensaciones y emociones; ansiedad; depresión y *craving*. En el segundo grupo están las variables que son inherentes a cada uno de los pacientes: años de consumo, días sin haber consumido antes de iniciar el tratamiento y número de tratamientos previos. Finalmente, el tercer conjunto está conformado por la variable: nivel de consumo de sustancias psicoactivas. Una finalidad de establecer dicha relación es conocer si hay evidencia de que el consumo de sustancias adictivas puede reducir si el tratamiento controla variables como las ya mencionadas.

El modelo ha sido formulado en tres grandes partes, las cuales se explicarán con detalle en la presente sección. Cada una de estas partes está anidada en sentido creciente, es decir, la primera está anidada en la segunda y la segunda esta anidada en la tercera. La anidación se postuló a partir de lo dicho por el experto, quien señaló lo siguiente: el tratamiento manipula los niveles de aceptación de pensamientos, sensaciones y emociones (Nivel 1). Se conoce de la literatura en psicología que cuando la aceptación de los estados afectivos negativos es mayor se alteran o disminuyen sus funciones aversivas (cargas afectivas negativas de un evento psicológico). En este caso, el estado afectivo negativo es asociado a los síntomas representativos de la ansiedad, depresión y *craving* (Nivel 2). Finalmente, cuando el paciente ha cambiado su reacción frente a las funciones aversivas de la sintomatología asociada a la ansiedad, a la depresión y a el *craving* su consumo de drogas disminuirá de tal forma que podría llegar a no consumir al final del tratamiento (Nivel 3).

3.1.1. Nivel 1. Modelación de las variables que manipula el tratamiento

A lo largo del tiempo que dure el tratamiento el psicólogo entrenará al paciente en estrategias que le permitan incrementar sus niveles de aceptación de pensamientos, sensaciones y emociones. El experto espera que el nivel de cada una de estas variables tenga un incremento al final del tratamiento. Se afirma que, dada la heterogeneidad en la población, cada uno de los pacientes posee su propio nivel de aceptación de pensamientos, sensaciones y emociones al inicio del tratamiento y además cada una de las personas reacciona de manera independiente a la terapia. Estas dos últimas afirmaciones se ven reflejadas en el intercepto y pendiente aleatorios del modelo, respectivamente. Dada esta información damos paso a la siguiente formulación:

Ajustamos un modelo lineal mixto con dos niveles. Consideremos aceptación de pensamientos (A) como variable dependiente y al tiempo (t) como variable independiente.

$$A_{ij} = a_i^A + b_i^A t_j + \varepsilon_{ij}, \tag{3-1}$$

con

$$\begin{array}{l} a_i^A = \alpha + \mathrm{a}_i^A \\ b_i^A = \beta + b_i^A \end{array}, \qquad \text{y} \qquad \begin{array}{l} i = 1, ..., N \\ j = 1, ..., n \end{array},$$

donde,

- A_{ij} es interpretada como la observación j-ésima del i-ésimo paciente de la variable aceptación de pensamientos.
- a_i^A es el intercepto de la recta asociada al i-ésimo individuo. Es la suma del parámetro poblacional α y un efecto aleatorio a_i^A.
- b_i^A es el coeficiente de regresión aleatorio asociado al i-ésimo individuo, $\beta = \mathbf{E}(b_i^A)$.
- t_j es la covariable tiempo.
- Asumimos que $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_\varepsilon^2)$, $\mathrm{a}_i^A \sim \mathrm{N}(0, \sigma^2_{\mathrm{a}^A})$ y $b_i^A \sim \mathrm{N}(0, \sigma^2_{b^B})$ son independientes. En términos de la expresión (2-1) estamos tomando $\mathbf{D}$ como una matriz diagonal con $\sigma^2_{\mathrm{a}^A}$ y $\sigma^2_{b^B}$ en la diagonal.

Haciendo las sustituciones correspondientes podemos escribir a (3-1) de una forma más explícita, como se muestra a abajo. Esta expresión será de gran utilidad, por su fácil lectura, para la simulación de datos de esta variable.

$$A_{ij} = (\alpha + \mathrm{a}_i^A) + (\beta + \mathrm{b}_i^A)t_j + \varepsilon_{ij}, \qquad \text{con} \quad \begin{array}{l} i = 1, ..., N. \\ j = 1, ..., n. \end{array}.$$

De forma análoga tenemos los respectivos modelos para las variables aceptación de emociones (B) y aceptación de sensaciones (C):

$$\begin{aligned} B_{ij} &= a_i^B + b_i^B t_j + \varepsilon_{ij} \\ &= (\alpha + \mathrm{a}_i^B) + (\beta + \mathrm{b}_i^B)t_j + \varepsilon_{ij}, \end{aligned} \tag{3-2}$$

y

$$\begin{aligned} C_{ij} &= a_i^C + b_i^C t_j + \varepsilon_{ij} \\ &= (\alpha + \mathrm{a}_i^C) + (\beta + \mathrm{b}_i^C)t_j + \varepsilon_{ij}, \end{aligned} \tag{3-3}$$

donde,

$$\begin{array}{l} a_i^B = \alpha + \mathrm{a}_i^B \\ b_i^B = \beta + \mathrm{b}_i^B \end{array}, \quad \begin{array}{l} a_i^C = \alpha + \mathrm{a}_i^C \\ b_i^C = \beta + \mathrm{b}_i^C \end{array}, \quad \text{con} \quad \begin{array}{l} i = 1, ..., N \\ j = 1, ..., n \end{array}.$$

Todas las personas tienen un cierto grado de aceptación, debido a lo cual podemos considerar un parámetro poblacional α. Cada persona reacciona a su ambiente de manera muy particular, es decir, la rapidez y constancia con la cual manejan su grado de aceptación cambia de sujeto a sujeto, reflejándose en una perturbación al parámetro fijo β.

Observemos que los dos parámetros: α y β permanecen sin ninguna modificación y en los tres modelos (3-1), (3-2) y (3-3). Se concibieron los mismos parámetros α y β para la modelación de cada una de las tres variables A, B y C debido a que el experto mencionó que las variables de aceptación de pensamientos, sensaciones y emociones son parte de un constructo más general llamado aceptación y son medidas con un solo cuestionario que tiene escalas independientes pero iguales para cada una de éstas. La simulación de dichas variables se facilitará gracias a estas consideraciones.

3.1.2. Nivel 2. Modelación del cambio de reacción (conducta) frente a los estados afectivos negativos

El Estado Afectivo Negativo de las personas, en el presente estudio psicológico, es representado por el conjunto de variables: *craving* (D), ansiedad (E) y depresión (F). Estas variables son tomadas en cuenta en el estudio porque el Estado Afectivo Negativo es muy frecuentemente asociado con la recaída en el consumo de drogas. Las intervenciones psicológicas tradicionales enfocadas al tratamiento de los síntomas asociados a la depresión, ansiedad y *craving*, en México, se basan en un cambio directo de pensamientos, sensaciones y emociones negativos involucrados. En el presente estudio el psicólogo plantea un tratamiento donde el cambio se hace de forma indirecta, es decir, a través de la aceptación de dichos síntomas, esperando así mejorar los resultados obtenidos hasta la fecha con los tratamientos actuales.

Cuando una persona tiene mayor aceptación de pensamientos, sensaciones y emociones sucede que también presenta modificación de las cargas afectivas negativas de los síntomas asociados a la depresión, ansiedad y *craving*. El experto espera que los síntomas asociados a D, E y F sean sobrellevados por el paciente de una forma diferente debido al tratamiento, tal que finalmente los niveles asociados a estas variables tiendan a disminuir. Lo anterior permite establecer una relación entre las seis variables ya mencionadas. Se ha construido un modelo lineal mixto para cada una de las variables dependientes D, E y F.

Se conoce en psicología que todas las personas presentan síntomas asociados a depresión, ansiedad y *craving* lo cual permite establecer la existencia de un parámetro poblacional. Cada uno de los pacientes llega a tratamiento con sus propios niveles de las variables D, E, y F debido a la heterogeneidad de la población. El experto afirma que cada persona modificará los niveles asociados a cada una de estas últimas tres variables de manera individual y única, pues cada persona evoluciona de forma independiente durante el tratamiento. Con base en estas dos últimas observaciones es que se plantearon modelos lineales mixtos con intercepto y pendientes aleatorios.

El modelo lineal mixto que considera a la variable *craving* (D) como dependiente y a aceptación de pensamientos (A), emociones (B) y sensaciones (C) como covariables queda expresado como sigue:

$$D_{ij} = c_i^D + d_i^D A_{ij} + e_i^D B_{ij} + f_i^D C_{ij} + \zeta_{ij}, \tag{3-4}$$

con

$$\begin{aligned} c_i^D &= \gamma^D + \mathrm{c}_i^D \\ d_i^D &= \delta^D + \mathrm{d}_i^D \\ e_i^D &= \eta^D + \mathrm{e}_i^D \\ f_i^D &= \theta^D + \mathrm{f}_i^D \end{aligned}, \qquad \text{y} \qquad \begin{aligned} i &= 1, ..., N \\ j &= 1, ..., n \end{aligned},$$

donde,

- D_{ij} es interpretada como la observación j-ésima del i-ésimo paciente de la variable *craving*.

- c_i^D es el intercepto asociado al i-ésimo individuo. Es la suma del parámetro poblacional γ y un efecto aleatorio c_i^D.

- d_i^D, e_i^D y f_i^D son coeficientes de regresión aleatorios asociados al i-ésimo individuo.

- A_{ij}, B_{ij} y C_{ij} son las covariables aceptación de pensamientos, emociones y sensaciones respectivamente.

- Asumimos que $\zeta_{ij} \sim \mathrm{N}(0, \sigma^2_\zeta)$, $\mathrm{c}_i^D \sim \mathrm{N}(0, \sigma^2_{c^D})$, $\mathrm{d}_i^D \sim \mathrm{N}(0, \sigma^2_{d^D})$, $\mathrm{e}_i^D \sim \mathrm{N}(0, \sigma^2_{e^D})$ y $\mathrm{f}_i^D \sim \mathrm{N}(0, \sigma^2_{f^D})$ son independientes.

De forma análoga tenemos los respectivos modelos para las variables ansiedad (E) y depresión (F):

$$E_{ij} = c_i^E + d_i^E A_{ij} + e_i^E B_{ij} + f_i^E C_{ij} + \zeta_{ij}, \tag{3-5}$$

y

$$F_{ij} = c_i^F + d_i^F A_{ij} + e_i^F B_{ij} + f_i^F C_{ij} + \zeta_{ij}, \tag{3-6}$$

donde,

$$\begin{array}{lcl} c_i^E = \gamma^E + c_i^E & & c_i^F = \gamma^E + c_i^F \\ d_i^E = \delta^E + d_i^E & & d_i^F = \delta^E + d_i^F \\ e_i^E = \eta^E + e_i^E & , & e_i^F = \eta^E + e_i^F \\ f_i^E = \theta^E + f_i^E & & f_i^F = \theta^E + f_i^F \end{array} , \qquad \text{con} \qquad \begin{array}{l} i = 1, ..., N \\ j = 1, ..., n \end{array},$$

especificando $\gamma^E, \delta^E, \eta^E$ y θ^E son efectos fijos; y c_i^E, d_i^E, e_i^E, f_i^E, c_i^F, d_i^F, e_i^F y f_i^F son efectos aleatorios.

La ansiedad y depresión se miden a través de un respectivo cuestionario con misma escala de medición. El experto supone que la media poblacional del nivel de cada una de estas variables es muy similar debido a lo cual se puede observar que el parámetro γ^E es el mismo para los modelos (3-5) y (3-6). Los parámetros δ^E, η^E y θ^E son considerados los mismos, dado que las variables E y F se espera evolucionen de forma parecida. Este supuesto de igualdad entre algunos parámetros permite reducir la cantidad de estimadores a calcular y con ello reducir la variablidad de las estimaciones para un mismo tamaño de muestra.

3.1.3. Nivel 3. Modelación de la variable nivel de consumo

Un tratamiento psicológico es eficaz si permite al paciente hacer cambios cuantitativos en su conducta en la dirección esperada. En el presente trabajo se desea comprobar la eficacia del tratamiento propuesto por el psicólogo, es decir, verificar si al final del tratamiento el consumo de drogas ha disminuido o en su caso ha sido nulo para cada uno de los pacientes. Cabe señalar que además de las variables que se manipulan con el tratamiento, también hay otras que se considera contribuyen a la disminución del consumo y que son inherentes al paciente: años de consumo de la sustancia psicoactiva (K), días sin haber consumido antes de iniciar el tratamiento (L) y número de tratamientos previos (M).

De manera similar a los dos niveles anteriores se plantea un modelo lineal con efectos mixtos. Los efectos aleatorios están asociados al intercepto y a los coeficientes de las variables D, E y F. Los efectos fijos están asociados a las otras tres covariables K, L y M. Propusimos esta clasificación ya que las primeras tres variables se modifican de forma particular conforme el tiempo transcurre y el segundo grupo K, L y M son variables que mantienen su valor fijo a lo largo del tiempo que tarda el estudio.

El modelo lineal mixto que considera a la variable nivel de consumo (H) como dependiente y como covariables a D, E, F, K, L y M es el siguiente:

$$H_{ij} = g_i + h_i D_{ij} + m_i E_{ij} + v_i F_{ij} + pK_i + qL_i + sM_i + \xi_{ij}, \tag{3-7}$$

con

$$\begin{aligned} g_i &= \kappa + g_i \\ h_i &= \lambda + h_i \\ m_i &= \varphi + m_i \\ v_i &= \psi + v_i \end{aligned}, \qquad \text{y} \qquad \begin{aligned} i &= 1, ..., N \\ j &= 1, ..., n \end{aligned},$$

donde,

- H_{ij} es interpretada como la observación j-ésima del i-ésimo paciente de la variable nivel de consumo.
- g_i es el intercepto asociado al i-ésimo individuo. Es la suma del parámetro poblacional κ y un efecto aleatorio g_i.
- h_i, m_i y v_i son coeficientes de regresión aleatorios asociados al i-ésimo individuo.
- K_i, L_i y M_i son las covariables años de consumo de la sustancia psicoactiva, días sin consumo de la sustancia previos al tratamiento y número de tratamientos previos, respectivamente.
- p, q y s son los efectos fijos del modelo.
- Asumimos que $\xi_{ij} \sim \mathrm{N}(0, \sigma_\xi^2)$, $g_i \sim \mathrm{N}(0, \sigma_g^2)$, $h_i \sim \mathrm{N}(0, \sigma_h^2)$, $m_i \sim \mathrm{N}(0, \sigma_m^2)$ y $v_i \sim \mathrm{N}(0, \sigma_v^2)$ son independientes.

3.1.4. Consideraciones al respecto del modelo

Dos importantes suposiciones en el presente trabajo son: afirmar que toda persona que inicia el tratamiento lo terminará y que las mediciones están completas y son tomadas en el mismo momento para todos los sujetos. Tales afirmaciones fueron adoptadas debido, en parte, a la falta de información, pues el experto no conoce con gran certeza qué porcentaje de personas abandona la terapia y en qué momento del tratamiento deciden hacerlo. La facilidad de planteamiento del modelo y la facilidad con la cual se obtienen los respectivos cálculos involucrados son otros incentivos para suponer dichas hipótesis. El modelo tendría que ser reestructurado si no se toman en cuenta estos hechos.

Cada uno de los modelos lineales con efectos mixtos acabados de presentar en este capítulo, son la composición de dos modelos: un modelo lineal mixto con intercepto aleatorio y uno con pendientes aleatorias. Para cada uno de los pacientes tenemos el mismo número de observaciones y para cada uno se medirán todas las variables involucradas en el estudio, es decir, se han planteado modelos balanceados.

3.2. Simulador

Un simulador puede ser un programa de computadora que permita reproducir la estructura de los datos reales de un experimento. Para cada conjunto de parámetros iniciales que alimentan al simulador (propuesto bajo ciertas hipótesis) se obtendrán diferentes resultados que serán considerados datos observados en el experimento real.

Para la presente situación, la finalidad primordial de la elaboración de un simulador es poder conocer qué pasaría si el estudio de campo se realizará bajo ciertas circunstancias, por ejemplo: conocer las posibles conclusiones si se cuenta con un número pequeño de pacientes; observar el efecto sobre el nivel final de consumo de sustancias psicoactivas si el cambio en los niveles de las covariables se comporta de forma diferente al esperado; *etc.* El experto podrá tener una idea anticipada de los posibles resultados que podría alcanzar una vez que conozca las condiciones reales bajo las cuales se llevará a cabo el estudio. Posiblemente con esto, el experto, adquirirá un cierto grado de conciencia que le ayudará a intentar buscar tener las condiciones óptimas para la realización de su experimento. Aunado a lo anterior podrá concebir una mejor organización de recursos y tiempo.

El proceso de simulación de las variables involucradas en el modelo se ha realizado en tres etapas; cada etapa corresponde, respectivamente, a cada uno de los niveles mencionados en la sección anterior. Cada vez que se realice una prueba con el simulador el resultado será una tabla que presumiblemente será equiparable a una muestra real de datos. Esta muestra será utilizada para practicar la inferencia estadística propuesta y a partir de este análisis se concluirá sobre la posible eficacia del tratamiento.

A continuación mostraremos un esbozo de la estructura general del simulador (para mayores detalles ver Apéndice 1), para que en caso de que el lector así lo desee manipule los datos de entrada del mismo para obtener sus propios resultados, o bien, compruebe los mostrados en el presente trabajo en el capítulo correspondiente.

Lo primero que se requiere es asignar valores numéricos a N (número de pacientes) y n (número de observaciones por paciente). Enseguida, ubicándonos en el modelo (3-1), asignamos datos a los parámetros: α, β, σ^2_ε, $\sigma^2_{\mathrm{a}^A}$ y $\sigma^2_{b^B}$. En primera instancia se obtendrán N números aleatorios de $\mathrm{a}_i^A \sim \mathrm{N}(0, \sigma^2_{\mathrm{a}^A})$ y $b_i^A \sim \mathrm{N}(0, \sigma^2_{b^B})$ uno por cada paciente. Posteriormente se hace,

$$A_{ij} = \alpha + \mathrm{a}_i^A + \beta t_j + b_i^A t_j + \varepsilon_{ij}, \qquad \text{con} \quad \begin{matrix} i = 1, ..., N \\ j = 1, ..., n \end{matrix},$$

obteniendo así n observaciones de la variable aceptación de pensamientos para cada uno de los N individuos. Se hace lo mismo para las otras dos variables del Nivel 1: aceptación de emociones (B) y aceptación de sensaciones (C).

La segunda etapa de la simulación es muy parecida a la primera, pero esta vez nos situamos en el modelo (3-4) y asignamos valores a los parámetros: γ^D, δ^D, η^D, θ^D, σ^2_ζ, $\sigma^2_{\mathrm{c}^D}$, $\sigma^2_{\mathrm{d}^D}$, $\sigma^2_{\mathrm{e}^D}$ y $\sigma^2_{\mathrm{f}^D}$. Primero se obtendrán N números aleatorios de $\mathrm{c}_i^D \sim \mathrm{N}(0, \sigma^2_{\mathrm{c}^D})$, $\mathrm{d}_i^D \sim \mathrm{N}(0, \sigma^2_{\mathrm{d}^D})$, $\mathrm{e}_i^D \sim \mathrm{N}(0, \sigma^2_{\mathrm{e}^D})$ y $\mathrm{f}_i^D \sim \mathrm{N}(0, \sigma^2_{\mathrm{f}^D})$ uno por cada paciente. Posteriormente, empleando los correspondientes datos obtenidos en el Nivel 1 para A, B y C, se hace:

$$D_{ij} = (\gamma^D + \mathrm{c}_i^D) + (\delta^D + \mathrm{d}_i^D)A_{ij} + (\eta^D + \mathrm{e}_i^D)B_{ij} + (\theta^D + \mathrm{f}_i^D)C_{ij} + \zeta_{ij}, \qquad \text{con} \quad \begin{matrix} i = 1, ..., N \\ j = 1, ..., n \end{matrix},$$

obteniendo así n observaciones de la variable *craving* para cada uno de los N individuos. Se hace lo mismo para las otras dos variables del Nivel 2: ansiedad (E) y depresión (F).

En la tercera y última parte del simulador seguiremos el mismo procedimiento que en las dos etapas anteriores pero esta vez para simular datos de la variable nivel de consumo de sustancias psicoactivas (H), a través de hacer:

$$H_{ij} = (\kappa + g_i) + (\lambda + h_i)D_{ij} + (\varphi + m_i)E_{ij} + (\psi + n_i)F_{ij} + pK_i + qL_i + sM_i + \xi_{ij}, \quad \begin{array}{l} i = 1, ..., N \\ j = 1, ..., n \end{array}.$$

Una vez que tenemos todos los datos simulados y el psicólogo ha opinado que efectivamente podría ser una muestra de datos reales dado su comportamiento cualitativo, los organizamos en una tabla (Base 2). Con lo anterior procedemos a hacer los análisis correspondientes. En un primer momento se puede graficar (con ayuda de la librería *lattice*) cada una de las variables para el número, N, de pacientes y de esta forma observar la evolución de dichas variables a lo largo del tiempo. También se obtienen las estimaciones puntuales de los parámetros y sus correspondientes intervalos de confianza.

3.2.1. Consideraciones al respecto de la simulación

Existen consideraciones que se debieron tomar en cuenta para poder llegar a caracterizar por completo a los datos obtenidos con ayuda del simulador. Estos detalles son similares para los tres niveles del modelo y existen algunas otros adicionales que son particulares del último nivel. Empecemos por mencionar los primeros:

Redondeo de decimales

Las escalas de las pruebas psicológicas con las cuales se mide a las variables involucradas en el estudio (aceptación de pensamientos, sensaciones y emociones, *craving*, ansiedad y depresión) son discretas, siempre mayores o iguales a cero según sea el caso y con un respectivo límite superior. En los Niveles 1 y 2 del modelo la simulación de cada una de estas variables tiene por contradominio a los números reales, no precisamente enteros, debido a lo cual a cada uno de los resultados simulados se redondea al entero más próximo.

Valores Negativos y valores que rebasan el límite superior de la escala de la prueba psicológica

La simulación de cada una de las variables dependientes de los Niveles 1 y 2 arroja valores que, por la naturaleza de los supuestos de normalidad hechos sobre el modelo, algunas ocasiones pueden llegar a tomar valores extremos. Es decir, pueden darse valores negativos o bien valores muy grandes. A todo número negativo se le asigna el valor mínimo de la escala de medición correspondiente a la variable y a todo número que sobrepasa el límite superior de la escala correspondiente se le asigna precisamente este valor. La presencia de valores negativos también se observa en el Nivel 3, ante lo cual procedemos de la forma ya mencionada.

En el mismo sentido, para la simulación de las variables involucradas en el último nivel se tomó en cuenta lo siguiente:

Muestreo empírico de covariables

Al inicio del tratamiento cada uno de los pacientes proporcionará la siguiente información: años de consumo de la sustancia psicoactiva (K), días sin consumo de la sustancia antes de iniciar el tratamiento (L), número de tratamientos previos (M), tipo de droga que consume (P), *etc.* El modelo (3-7) involucra estas variables para la explicación del nivel de consumo final de drogas (H). La simulación de esta variable H requiere simular cada una de las covariables K, L, M y P. Para el caso de las variables K, L y P se busca mantener la estructura de la base de datos de un estudio previo del experto (Base 1). Para conocer más al respecto de las características de esta base se puede consultar Pérez-Romero (2014). Tomamos un número aleatorio en el rango de los números enteros $[1-29]$ y la información correspondiente a ese número de paciente en la Base 1 es asignada al i-ésimo individuo con $i = 1, \ldots, N$ de la Base 2.

Se consideró como una primera aproximación simular los mismos datos con distribuciones uniformes; sin embargo, esta suposición no es realista pues los fenómenos naturales muy rara vez presentan este comportamiento. Otra posibilidad consistía en ajustar una densidad conocida a los datos de cada una de las variables K, L y N pero para esto se consideró que la cantidad de datos disponibles es insuficiente. Así, se concluyó que la opción más viable era la del muestreo de casos sucedidos en el estudio piloto.

Determinación de densidades

La Base 1 no muestra información al respecto de la variable número de tratamientos previos (M), debido a lo cual no podemos aplicar el mismo procedimiento de simulación empleado para

Sustancia	***Consumo mínimo*** (m^*)	***Consumo máximo*** (M^*)	***Unidad de medida***
Acohol	0	220	Copas
Cocaína	0.44	3.44	Gramos
Marihuana	1.52	3.85	Cigarros
Piedra	0.44	3.44	Gramos
Pvc	0	125	Mililitros

Tabla 3-1: Rangos del nivel de consumo de las sustancias psicoactivas.

las variables K, L y P. El experto ha dicho, con base en su experiencia, que el rango de esta covariable puede ir de 1 hasta 3. Con sólo esta información lo que se decide hacer para la simulación de M es afirmar que sigue una distribución uniforme discreta en el rango dictado por el psicólogo.

Coeficiente de variación constante

Los pacientes que serán atendidos en el estudio se espera que principalmente sean consumidores de: alcohol, cocaína, marihuana, piedra o pvc. El psicólogo, con base en información propia, estableció los rangos de consumo semanal que pueden llegar a tener los pacientes consumidores de cada una de las sustancias (ver Tabla 3-1). Cabe mencionar que estos datos no fueron obtenidos de una muestra aleatoria, sin embargo, se consideran como una aproximación.

Como podemos observar en la Tabla (3-1) los rangos de consumo varían de una sustancia a otra, por ejemplo, una persona puede llegar a consumir un poco más de 200 copas de alcohol mientras que otra persona consumidora de cocaína podría llegar a consumir como máximo 3.44 gramos. Dada esta diferencia tan pronunciada en los rangos de consumo no es posible simular de manera directa a la variable nivel de consumo (H).

Con la información contenida en la Tabla (3-1) es posible hacer estimaciones aunque sean un poco burdas de la media $\left(\hat{\mu} = \dfrac{m^* + M^*}{2}\right)$ y desviación estándar $\left(\hat{\sigma} = \dfrac{M^* - m^*}{6}\right)$ del nivel de consumo para cada una de las sustancias, información mostrada en la Tabla (3-2). Ubicándonos en la última columna de esta segunda tabla observamos que en general se cumple $\dfrac{\hat{\mu}}{\hat{\sigma}} \approx 3$, lo cual es estadísticamente conocido como coeficiente de variación constante. En la simulación se

Sustancia	***Media*** ($\hat{\mu}$)	***Desviación Estándar*** ($\hat{\sigma}$)	$\frac{\hat{\mu}}{\hat{\sigma}}$
Acohol	110	36.67	3.00
Cocaína	1.94	0.50	3.88
Marihuana	2.06	0.65	3.19
Piedra	1.94	0.50	3.88
Pvc	62.50	20.83	3.00

Tabla 3-2: Estimaciones para la media ($\hat{\mu}$) y desviasión estándar ($\hat{\sigma}$) del nivel de consumo de las sustancias psicoactivas.

buscará que H^*_{ij} (consumo reescalado) con $i = 1, .., N$ y $j = 1, ..., n$ adquiera el valor aproximado de 3 disminuyendo conforme el tiempo avance. El producto $H^*_{ij}\sigma$ nos proporcionará el consumo final (H_{ij}) en la unidad de medida correspondiente al tipo de sustancia.

Veamos un ejemplo que permita explicar un poco más la idea de coeficiente de variación constante. Supongamos dos sustancias, 1 y 2. Supongamos que X es la variable independiente, y que los consumos de cada una de las sustancias son de las siguientes formas:

$$C_1 = \alpha_1 + \beta_1 X + \sigma_1 \varepsilon_1,$$
$$C_2 = \alpha_2 + \beta_2 X + \sigma_2 \varepsilon_2.$$

ε_1 y ε_2 tienen media cero y desviación estándar uno. Observemos entonces que C_i para $i = 1, 2.$, tiene media $\alpha_i + \beta_i X$ y desviación estándar σ_i. Cada consumo tiene su propia media y escala.

Ahora supongamos que se cumple que media/desviación es constante para todas las sustancias. Es decir, $(\alpha_1 + \beta_1 X)/\sigma_1 = (\alpha_2 + \beta_2 X)/\sigma_2$. Entonces se cumple que,

$$C_1/\sigma_1 = (\alpha_1 + \beta_1 X)/\sigma_1 + \varepsilon_1,$$
$$C_2/\sigma_2 = (\alpha_2 + \beta_2 X)/\sigma_2 + \varepsilon_2.$$

Luego, los consumos reescalados C_1/σ_1 y C_2/σ_2 tendrían la misma media, y varianza uno. Esto equivale a asumir que la relación entre consumo reescalado y X es una misma recta. Esto quiere decir que, reescalando cada consumo usando su σ correspondiente, la misma recta gobierna a todas las sustancias, y que el efecto de X recae sobre la media del consumo reescalado.

3.3. Procedimientos de estimación

Una vez que el estudio de campo se haya realizado y el experto cuente con la base de datos reales se procederá a la realización de gráficos y estimación de parámetros e intervalos de confianza. Actualmente contamos con la tabla de datos simulados con la cual practicaremos la inferencia estadística propuesta.

El proceso de estimación se realizará en tres etapas en la misma forma que el planteamiento del modelo y el simulador. En cada uno de los niveles se obtendrán los estimadores puntuales y los intervalos de confianza de cada uno de los parámetros involucrados en cada modelo.

Como ya mencionabamos en la última sección del Capítulo 2, debemos de prestar atención en el uso de la notación empleada en la función *lmer()* al momento de especificar cada uno de los componentes del modelo. Veamos a continuación cómo se escribe el modelo (3-1) en términos de la función ya mencionada; los estimadores de los parámetros a calcular son: $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_{\varepsilon}^2$, $\hat{\sigma}_{\mathrm{a}}^2$ y $\hat{\sigma}_b^2$:

```
m_t1 <- lmer(t_1 ~ tiem + (1|ind) + (0+tiem|ind))
```

donde:

- `t_1`. Corresponde a la variable dependiente aceptación de pensamientos (A).
- `tiem`. Corresponde a la covariable tiempo (t).
- `(1|ind)`. Corresponde al efecto aleatorio asociado al intercepto.
- `(0+tiem|ind)`. Corresponde al efecto aleatorio asociado a la pendiente. El cero nos indica que las pendientes aleatorias son independientes de los interceptos aleatorios.

Las estimaciones puntuales que obtenemos son: el parámetro poblacional ($\hat{\alpha}$), el coeficiente asociado al efecto fijo del tiempo ($\hat{\beta}$), las varianzas asociadas a cada uno de los efectos aleatorios ($\hat{\sigma}_{\mathrm{a}}^2$ y $\hat{\sigma}_{\mathrm{b}}^2$) y la varianza de la perturbación asociada al modelo ($\hat{\sigma}_{\varepsilon}^2$). Por otra parte, los intervalos de confianza se calcularán con el 95 % de confianza y por el método *bootstrap*. Este método requiere especificar el número de remuestreos a realizarse para el cálculo del intervalo de confianza. Para todos los casos se han propuesto 2500 réplicas.

Se realizará un procedimiento de estimación similar al presentado arriba para los modelos (3-2) y (3-3); asimismo se obtendrán los estimadores puntuales para los parámetros e intervalos de confianza del Nivel 2, para los modelos (3-4), (3-5) y (3-6). Para el caso del Nivel 3 es exactamente lo mismo, como lo mostraremos enseguida:

```
m3 <- lmer(consumo ~ Cr + An + De + años_cons + dias_sin_cons + tratam_previos
                    + (1|ind) + (0+Cr|ind) + (0+An|ind) + (0+De|ind)
```

- `consumo`. Corresponde a la variable dependiente nivel de consumo de sustancias psicoactivas (H).
- `Cr` , `An` , `De` , `años_cons` , `dias_sin_cons` , `tratam_previos`. Corresponde a la variable asociada a cada una de las mencionadas: *craving* (D), ansiedad (E), depresión (F), años de consumo de la droga (K), días sin consumo antes de iniciar el tratamiento (L) y número de tratamientos previos (M) respectivamente.
- `(1|ind)`. Corresponde al efecto aleatorio asociado al intercepto.
- `(0+Cr|ind)` , `(0+An|ind)` , `(0+De|ind)`. Corresponde al efecto aleatorio asociado a las respectivas variables. El cero nos indica que las pendientes aleatorias son independientes de los interceptos aleatorios.

Capítulo 4

Resultados

En la Sección 3.2 del capítulo anterior se ha revisado el proceso de simulación de datos correspondientes a cada una de las variables involucradas en el estudio. Ahora lo primero que se realizará es calibrar los parámetros involucrados en el modelo y que son requeridos por el simulador. Se recurre a este concepto de calibración en virtud de que no hay datos observados todavía; si los hubiera, este proceso sería simplemente un concepto de estimación de parámetros con base en datos observados.

La calibración consiste en buscar parámetros que permitan obtener el comportamiento esperado de la variable dependiente involucrada en los modelos, (3-1) a (3-7), respectivamente. El resultado del procedimiento anterior son muestras de datos que presumiblemente pudieran ser consideradas datos reales del estudio. Con esta información se calcularán las estimaciones correspondientes que, después de ser analizadas, permitirán establecer el requerimiento de información que permitirá obtener con mayor seguridad resultados estadísticamente significativos.

Se realizará la simulación de cuatro diferentes escenarios. El primero de ellos fue establecido por el experto. El segundo fue propuesto con base en las condiciones del tratamiento psicológico; principalmente la disponibilidad de tiempo. Los últimos dos casos se sustentan en la importancia de sugerir una solución estadísticamente significativa. Considerar diferentes situaciones y compararlas entre ellas permitirá situarnos lo más cerca del caso más conveniente.

En el presente capítulo mostraremos los resultados de cada uno de estos escenarios, es decir, los estimadores puntuales y sus intervalos de confianza para cada uno de los parámetros del modelo. Asimismo, presentaremos las tablas de datos simulados y los gráficos que en su conjunto permiten, de manera visual, conocer el comportamiento de cada una de las variables involucradas en el estudio. Estos dos últimos instrumentos permitieron ilustrar al usuario las modificaciones que ocurren en los datos simulados cuando se cambian los valores asignados a los parámetros que se ingresan al simulador. El análisis de los resultados ya mencionados será presentado en el siguiente capítulo.

4.1. Realizaciones de la simulación

La Tabla (4-1) resume los diferentes escenarios considerados; son cuatro grandes casos y para cada uno de ellos tenemos diferentes subcasos debidos a modificaciones sobre la variable número de pacientes. En total revisaremos los resultados de 9 escenarios diferentes; cada uno de éstos tiene motivos específicos para ser considerado, los cuales se explicarán enseguida.

Caso A

Como ya comentamos en el Capítulo 1, antes de iniciar el presente análisis el psicólogo había desarrollado un plan de trabajo. El experto mencionaba, entre otras cosas, que el estudio posiblemente se realizaría con 10 pacientes y 7 mediciones de cada una de las variables por persona. Pues bien, consideremos a esta situación como el primer caso a revisar, el Caso A.

El Caso A se sustenta en la experiencia empírica del experto, pues cuando tomó la decisión no contaba con ningún estudio estadístico que le exigiera cierto requerimiento de información. La

Ejemplo	***Número de observaciones por paciente***	***Número de pacientes***
Caso A	7	10
Caso I	12	12
		30
Caso II	18	12
		25
		34
		42
Caso III	30	12
		42

Tabla 4-1: Casos de estudio.

escasez de tiempo, personal capacitado y espacios disponibles para dar tratamiento a un número extenso de pacientes, se vio reflejada en la decisión del usuario al establecer los parámetros ya mencionados.

Caso I

El tiempo disponible para la realización del estudio es de aproximadamente dos años. La duración del tratamiento es de 3 meses y consiste de una sesión por semana, es decir, el tratamiento completo son 12 sesiones. Algunos recursos necesarios como consultorios y terapeutas capacitados en el área son limitados. Con base en esta información fue que se propuso el Caso I.

El número de sesiones impacta directamente sobre el número de mediciones que se pueden tomar. Para este caso se trataría de poder aplicar los cuestionarios cada sesión del tratamiento, logrando así el máximo de 12 mediciones. El experto afirma que por cuestión de tiempo el paciente no podría responder cada sesión todos los cuestionarios, lo cual reduce la posibilidad de que este escenario se realice así como se ha establecido.

El número de pacientes es otra variable de gran relevancia en el estudio. En la Tabla (4-1) observamos que para este escenario se han propuesto dos situaciones: una con 12 pacientes y otra con 30. En el primer subcaso, al ser pocos pacientes, muy seguramente no se obtendrán resultados estadísticamente significativos a pesar de que posiblemente se lleguen a observar cambios considerables en el puntaje de la variable nivel de consumo al finalizar las sesiones de terapia. El segundo subcaso considera 30 pacientes, se propuso este número debido a que el experto supone que un número máximo de personas que pudiese llegar a tratar estaría alrededor de 35.

Caso II

El Caso II se ha propuesto como consecuencia de revisar los resultados de la estimación de parámetros e intervalos de confianza para los Casos A y I. Veremos más adelante que la cantidad de información en estos dos escenarios no es suficiente para poder obtener resultados estadísticamente significativos. Para establecer el Caso II aumentamos en 6 unidades, respecto al Caso I, la cantidad de observaciones por paciente llegando a un total de 18 mediciones por persona.

Observemos que un tratamiento con 12 sesiones no permitiría tener 18 mediciones. Este incremento estaría sugiriendo al psicólogo una extensión del tratamiento o, en su caso, un seguimiento de los pacientes posiblemente durante un año para así poder distribuir las 18 mediciones a lo largo de este tiempo.

En el Caso II consideramos 18 observaciones por paciente con cuatro diferentes subcasos que toman en cuenta: 12, 25, 34 y 42 personas respectivamente. Posiblemente ya podríamos inferir que 12 pacientes son una cantidad muy reducida; sin embargo, el resultado podría darnos una idea de qué tanto se requeriría incrementar esta variable para poder obtener modificaciones en los resultados. Los subcasos con 25 y 34 pacientes parecerían ser los más viables por el momento ya que están dentro del rango del número de personas que el experto considera puede atender; al ser cantidades considerablemente mayores posiblemente obtengamos resultados realmente significativos. La hipótesis de 42 pacientes se espera mejore aún más la significancia de los resultados aunque en la práctica esta situación al parecer no resultará muy viable.

Caso III

Una cantidad mayor de mediciones de cada una de las variables por persona, digamos 30, implicaría para el experto un tratamiento o un seguimiento de los pacientes durante más de un año. Esta situación no es del todo posible dada la disponibilidad de tiempo y además buscar un seguimiento largo de los pacientes es una tarea que no siempre es posible conseguir. Hemos considerado una buena idea darle un vistazo a este escenario para observar qué tanto se podrían mejorar, si es que mejoran, las estimaciones estadísticas en comparación con los otros 3 escenarios.

En esta situación consideramos únicamente dos hipótesis diferentes para el número de pacientes: 12 y 42. Uno de los problemas con los tratamientos largos es que muy poca gente llegará exitosamente al final de todas las sesiones, así que considerar a muchas personas no asegura su permanencia en el estudio. La tasa de abandono del tratamiento no se conoce con mucha certeza; el experto comenta que en un tratamiento largo es de aproximadamente 30 %; esta situación depende de muchas otras variables, por ejemplo, la relación terapeuta-paciente. Conseguir un estudio con las características indicadas sería muy poco viable principalmente por la disposición de tiempo. Sin embargo, por cuestiones de simulación y académicas comentemos al respecto de los correspondientes resultados. Lo que esperaríamos son estimaciones más próximas a las reales y significativas.

4.1.1. Calibración de parámetros

Una vez con los cuatro escenarios en mente obtengamos los datos que jugarán el rol de muestras aleatorias para cada uno de los casos. Para lograr esto requerimos seguir el procedimiento establecido en la Sección 3.2 del capítulo anterior. Allí se pide asignar valores numéricos a ciertos parámetros; ahora la pregunta que surge es «¿qué valores se deben ingresar?», pues bien, demos respuesta a esta interrogante.

Para cada uno de los cuatro escenarios los parámetros serán diferentes. Esto es debido a la cantidad de tiempo (duración del tratamiento o seguimiento de los pacientes) que considera cada caso. Por ejemplo, en el primer nivel, al considerar a la variable aceptación de pensamientos como función del tiempo, la recta que representa a este fenómeno tendrá pendientes cada vez menos inclinadas conforme pasamos de un escenario a otro.

El experto ha mencionado los rangos sobre los cuales espera que se muevan los puntajes obtenidos en los correspondientes cuestionarios. Al inicio del tratamiento estarán sobre un rango y al final de las sesiones de terapia, o en su caso, al final del seguimiento de los pacientes, se espera que se ubiquen en otro rango. Estas afirmaciones fueron hechas debido a que no cuenta con información previa del posible comportamiento a través de tiempo de algunas de las variables involucradas en el estudio. Por ejemplo, para el caso de las variables aceptación de pensamientos, sensaciones y emociones, el experto ha hecho suposiciones al respecto de los puntajes obtenidos por los pacientes en el respectivo cuestionario: son variables que se espera incrementen su valor conforme el tiempo transcurre iniciando con un nivel que se encuentra en el rango [9,18] y terminando con una calificación en el rango [36,45]. La Tabla (4-2) detalla esta información para las variables del estudio que así lo requieren.

Establezcamos ciertos valores iniciales para los parámetros. Con la información presentada en la Tabla (4-2) podemos, por ejemplo, hacer una aproximación para el valor de los parámetros del Nivel 1 del modelo. La idea consiste en obtener la recta que pasa por los dos puntos cuya primer entrada corresponde a un momento en el tiempo y la segunda a los puntos medios de

Variable	***Rango al iniciar el tratamiento***	***Rango al concluir el tratamiento***
* Aceptación de Pensamientos * Aceptación de Sensaciones * Aceptación de Emociones	[9, 18]	[36, 45]
Craving	[16, 24]	[0, 10]
* Depresión * Ansiedad	[21, 84]	[1, 20]

Tabla 4-2: Rangos.

los intervalos de los rangos inicial y final respectivamente, es decir, por los puntos (1,13.5) y (12,40.5). De esta forma obtenemos la pendiente y la ordenada al origen que definen a la recta que representa el comportamiento de la variable aceptación de pensamientos y de forma similar a las variables aceptación de sensaciones y emociones ya que comparten la misma escala y ciertas características psicológicas.

Para proponer la desviación estándar del efecto aleatorio asociado al intercepto de la misma recta se hizo la aproximación (18-9)/6; el resultado de este cociente es la desviación estándar asociada al correspondiente intervalo de la Tabla (4-2). Para el caso de asignar un valor a la desviación estándar del efecto aleatorio asociado a la pendiente de la recta y para la respectiva desviación del error del modelo, es importante reconocer que se realizó de una forma muy intuitiva pues se hacían algunos ensayos y se le mostraban al experto los gráficos de dispersión de la variable, preguntándole si el comportamiento observado era el que se esperaba y, de no ser así, se hacían las respectivas modificaciones de los valores asignados a estos parámetros. La asignación de los parámetros en los niveles 2 y 3 se hizo de manera similar al Nivel 1.

Los gráficos de dispersión son como los mostrados en la Figura (4-1); cada uno es una composición de N sub-gráficas, una por cada individuo. Con ayuda de estos instrumentos podemos observar el comportamiento general, es decir, el cambio a través del tiempo de las puntuaciones obtenidas en los correspondientes cuestionarios realizados por los pacientes.

Para tener una idea general de cómo la modificación de los valores asignados a los parámetros influyen sobre el comportamiento de los datos veamos un ejemplo. Consideremos el caso de la respuesta aceptación de pensamientos. A la izquierda de la Figura (4-1) se muestra el comportamiento de dicha variable cuando asignamos a los parámetros los siguientes valores:

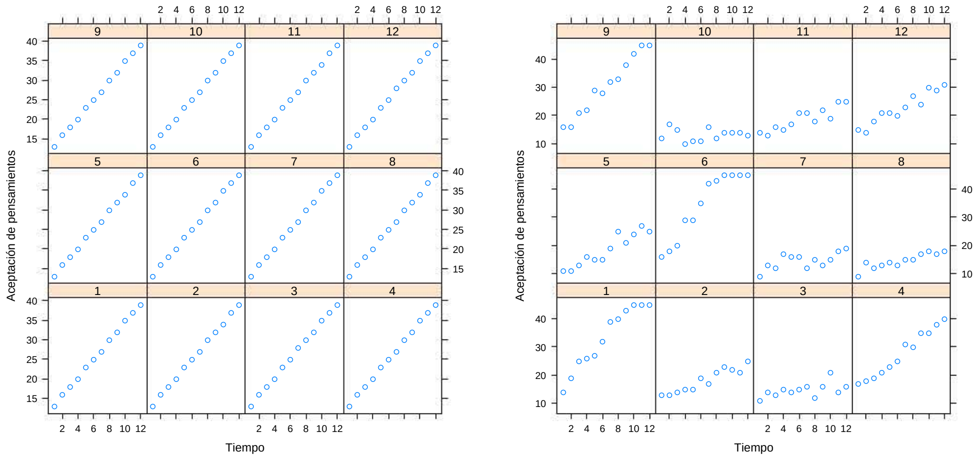

Figura 4-1: Gráficos de dispersión.

$\alpha = 11.03$, $\beta = 2.35$, $\sigma_\varepsilon = 0.05$, $\sigma_{a_A} = 0.001$ y $\sigma_{b_A} = 0.001$, mientras que la gráfica de la derecha es el resultado de una modificación en los parámetros: $\sigma_\varepsilon = 1.85$, $\sigma_{a_A} = 1.5$ y $\sigma_{b_A} = 1.18$. Como es de esperarse los resultados son contrastantes. Podríamos decir que en el primer caso al ser tan pequeñas las desviaciones estándar no existen efectos aleatorios. En un primer momento mostramos al psicólogo el primer gráfico mencionado y observándolo, el experto, alertó al respecto de dos características: la primera fue la variación existente entre las observaciones de cada individuo; mencionó que no era posible que la variable tuviera un comportamiento tan estable. Si bien es cierto que presenta una tendencia creciente también es cierto que se esperarían disminuciones un poco más pronunciadas en ciertos momentos. La segunda consideración fue que no todos los pacientes evolucionan con la misma rapidez. Por estas razones se modificaron los parámetros y se llegó a la gráfica de la derecha. Con base en este par de comentarios pudimos estar más convencidos de que el presente fenómeno psicológico se podía representar con un modelo lineal mixto.

De esta forma fue posible llegar a los resultados finales de asignación de valores a los parámetros del modelo mostrados en las Tablas (4-3) a (4-5). En las Tablas (4-3) y (4-4) se expone la información de cada uno de los tres modelos involucrados en los Niveles 1 y 2, respectivamente, es una correspondencia uno a uno en los renglones donde hay más de un elemento. Tenemos los parámetros para cada uno de los cuatro diferentes casos considerados a analizar.

Parámetro	***Caso A***	***Caso I***	***Caso II***	***Caso III***
α	9	11.03	11.01	11
β	4.5	2.35	1.08	0.87
σ_ε	2.4	1.85	1.85	1.67
σ_{a^A}, σ_{a^B}, σ_{a^C}	1.5 , 1.51, 1.48	1.5 , 1.51, 1.48	1.47, 1.5, 1.48	1.45 , 1.38, 1.41
σ_{b^A}, σ_{b^B}, σ_{b^C}	1.38, 1.36, 1.39	1.18, 1.19, 1.17	0.72, 0.71, 0.74	0.48 , 0.45 , 0.44

Tabla 4-3: Parámetros sugeridos para el Nivel 1.

Parámetro	***Caso A***	***Caso I***	***Caso II***	***Caso III***
γ^D, γ^E	22, 72.5	21, 72.5	20.52, 72.5	20, 58.5
δ^D, δ^E	-0.09, -0.39	-0.09, -0.26	-0.12, -0.32	-0.19, -0.21
η^D, η^E	-0.12, -0.31	-0.15, -0.34	-0.14, -0.31	-0.15, -0.29
θ^D, θ^E	-0.11, -0.34	-0.13, -0.31	-0.13, -0.32	-0.13, -0.26
σ_ζ	1.68	1.38	1.58	1.3
σ_{c^D}, σ_{c^E}, σ_{c^F}	1.53, 6.22, 6.12	1.53, 6.22, 6.12	1.56, 6.24, 6.12	1.59, 6.27, 6.25
σ_{d^D}, σ_{d^E}, σ_{d^F}	0.11, 0.18, 0.16	0.11, 0.16, 0.15	0.11, 0.15, 0.15	0.12, 0.14, 0.15
σ_{e^D}, σ_{e^E}, σ_{e^F}	0.12, 0.43, 0.39	0.12, 0.42, 0.39	0.11, 0.4, 0.39	0.11, 0.38, 0.29
σ_{f^D}, σ_{f^E}, σ_{f^F}	0.11, 0.5, 0.52	0.11, 0.49, 0.51	0.12, 0.44, 0.43	0.13, 0.42, 0.32

Tabla 4-4: Parámetros sugeridos para el Nivel 2.

Parámetro	***Caso A***	***Caso I***	***Caso II***	***Caso III***
κ	1.72	1.72	1.72	1.7
λ	0.0195	0.0215	0.025	0.0325
φ	0.012	0.017	0.0192	0.02125
ψ	0.0111	0.0161	0.0181	0.02375
p	0.008	0.009	0.009	0.009
q	0.0072	0.0082	0.0081	0.00825
s	0.0076	0.0076	0.0072	0.0075
σ_ξ	0.49	0.55	0.55	0.52
σ_{g}	0.01	0.01	0.01	0.011
σ_{h}	0.017	0.017	0.017	0.013
σ_{m}	0.015	0.015	0.015	0.012
σ_{n}	0.013	0.013	0.013	0.011

Tabla 4-5: Parámetros sugeridos para el Nivel 3.

4.2. Tabla de datos simulados y gráficas de dispersión

Para cada pareja nueva de parámetros N y n ingresados al simulador obtendremos un juego diferente de gráficas de dispersión y una tabla de datos simulados. Las gráficas permiten al usuario observar características de las variables, por ejemplo: dispersión y tendencia entre los datos un mismo paciente, velocidad con la cual las personas incrementan o disminuyen sus niveles de cada una de las variables, comparación de la información entre los diferentes pacientes, *etc.*

Por otra parte, la tabla de datos simulados presenta valores específicos. Esta información permite conocer con más precisión los cambios existentes en las puntuaciones obtenidas por las personas en las diferentes pruebas psicológicas, así como conocer la tendencia en los puntajes de la variable de principal interés: nivel de consumo de sustancias psicoactivas.

En términos generales, los dos instrumentos mencionados en esta sección fueron de gran utilidad en el proceso de asignación de valores a los correspondientes parámetros, requeridos para la simulación de datos. Mostrar al experto los datos en una presentación diferente, no conocida por él anteriormente, permitió explicaciones más claras y específicas al respecto de la idea de efectos aleatorios, por ejemplo: pendientes específicas y únicas por paciente, la idea de varianza entre los datos de un mismo paciente, *etc.*

Ahora mostraremos los correspondientes gráficos de dispersión y la tabla de datos simulados para el Caso A y el Caso II con 34 pacientes. Se consideran exclusivamente estos dos escenarios debido a que el Caso A es la propuesta del experto y resultaría interesante poder decirle si con el análisis de la información considerada en esta situación se puede dar respuesta a todas sus interrogantes. El Caso II con 34 pacientes parece ser la opción con resultados estadísticamente significativos (como ya mostraremos más adelante) más posible de realizarse.

Caso A

El juego de gráficos mostrados en la Figura (4-2) ilustra el comportamiento de las variables del primer nivel. Como el experto mencionó: los datos tienen tendencia creciente con ligeras disminuciones en algunos momentos del tiempo. Cada grupo de datos evoluciona de forma particular lo cual se aprecia en las diferentes pendientes (gráfico superior derecho).

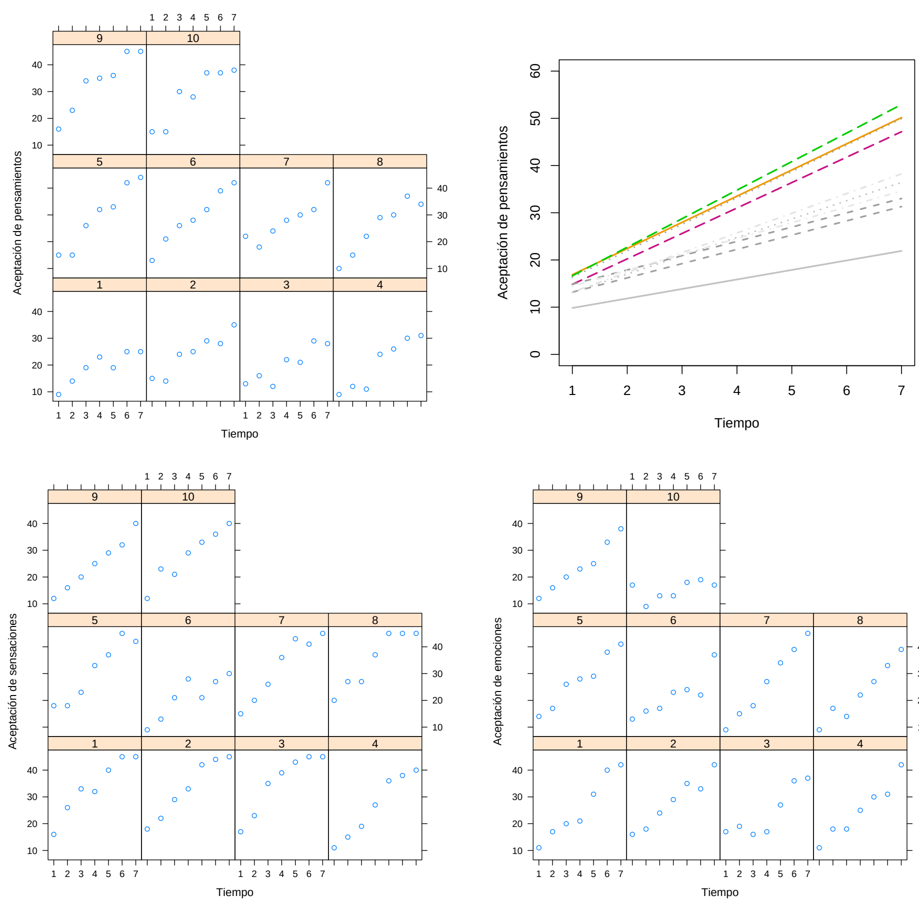

Figura 4-2: Gráficos del Nivel 1.

El juego de gráficos mostrados en la Figura (4-3) ilustra el comportamiento de las variables del segundo nivel. Los niveles de *craving*, ansiedad y depresión presentan una tendencia a disminuir conforme el tiempo transcurre, es decir, conforme la aceptación de pensamientos, sensaciones y emociones incrementa. Cada paciente aprende y emplea las destrezas enseñadas por el terapeuta con cierta particularidad lo cual se refleja en la velocidad con la cual las puntuaciones tienden a disminuir.

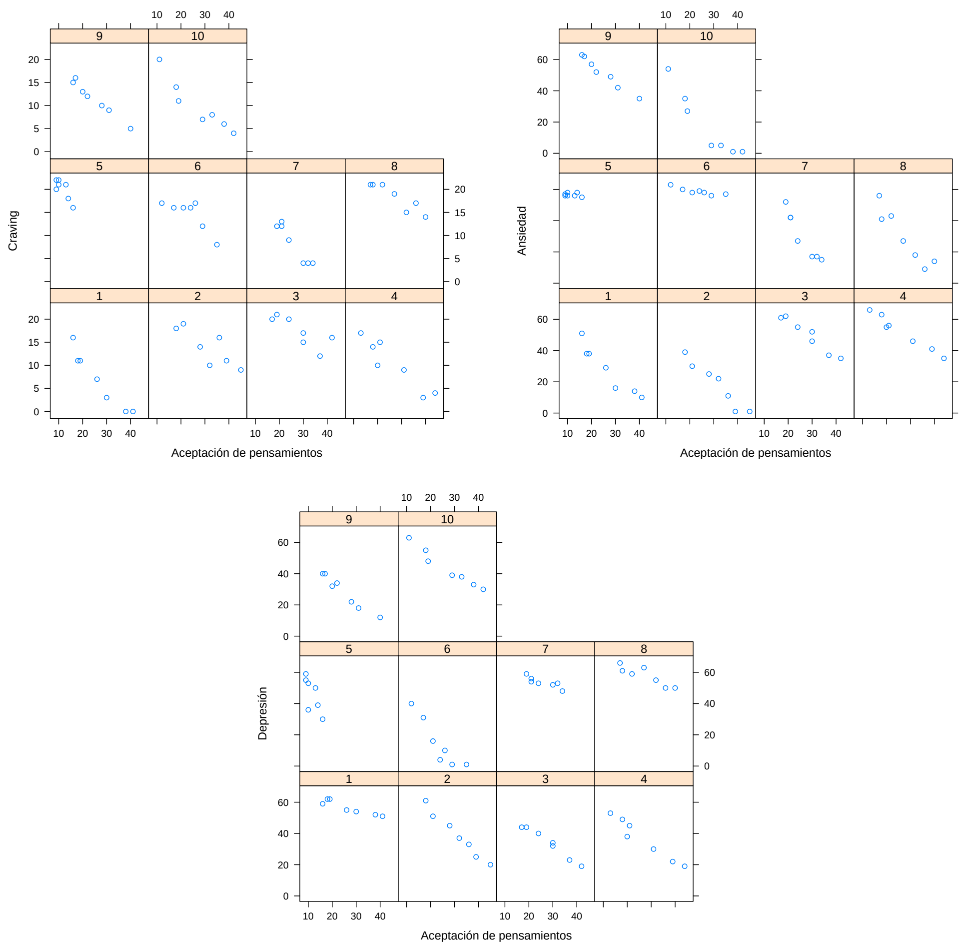

Figura 4-3: Gráficos del Nivel 2.

La gráfica (4-4) presenta el comportamiento de la variable consumo. Como podemos ver se aprecia una ligera tendencia a la baja.

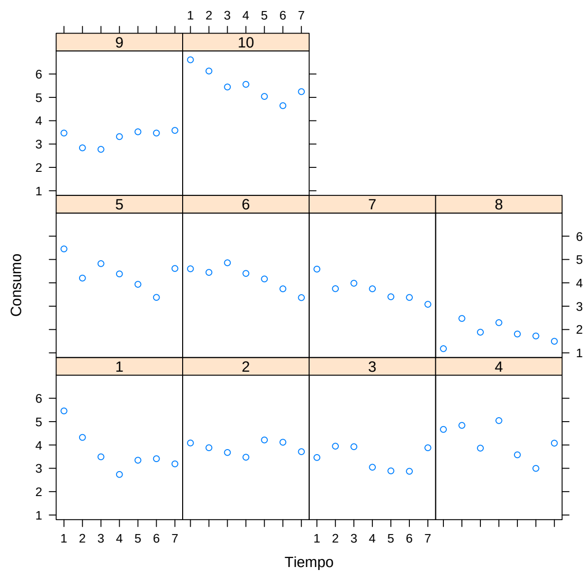

Figura 4-4: Gráfico del Nivel 3.

En la Tabla (4-6) podemos ver los valores numéricos ya representados en los gráficos de las Figuras (4-2) a la (4-4). Así comprobamos las tendencias de cada uno de los grupos de datos.

Sustancia	*Individuo*	*Tiempo*	*Ace Pens*	*Ace Sens*	*Ace Emoc*	*Craving*	*Ansiedad*	*Depresión*	*Años Cons*	*Días S/Co*	*Tratam Prev*	*Unidades Cons*
1	1	1	16	14	12	20	67	55	15	15	1	163
1	1	2	18	17	16	16	64	51	15	15	1	177.4
1	1	3	27	21	23	25	57	39	15	15	1	147.87
1	1	4	29	30	34	13	48	28	15	15	1	118.49
1	1	5	33	33	32	14	48	29	15	15	1	151.76
1	1	6	36	36	37	13	45	21	15	15	1	155.07
1	1	7	42	38	42	11	40	15	15	15	1	125.44
4	3	1	10	17	13	21	62	52	15	30	2	2.26
4	3	2	21	20	17	16	61	44	15	30	2	1.82
4	3	3	29	24	20	15	55	42	15	30	2	1.97
4	3	4	35	28	24	18	51	33	15	30	2	2.11
4	3	5	38	35	21	13	49	35	15	30	2	1.87
4	3	6	45	38	30	9	46	30	15	30	2	2.04
4	3	7	45	45	28	14	40	26	15	30	2	1.69
2	4	1	12	14	13	13	41	55	8	28	1	2.89
2	4	2	12	19	22	11	32	46	8	28	1	1.79
2	4	3	18	32	30	5	14	37	8	28	1	1.89
2	4	4	20	31	30	5	18	38	8	28	1	2.01
2	4	5	19	39	34	6	7	36	8	28	1	1.67
2	4	6	29	45	41	0	1	25	8	28	1	0.89
2	4	7	30	45	43	0	1	24	8	28	1	1.59
4	5	1	11	11	18	12	51	61	13	7	2	2.33
4	5	2	17	18	21	13	44	56	13	7	2	2.89
4	5	3	25	20	25	11	33	47	13	7	2	1.95
4	5	4	29	22	25	11	32	44	13	7	2	1.8
4	5	5	31	28	34	4	20	38	13	7	2	1.13
4	5	6	34	30	40	5	17	33	13	7	2	1.38
4	5	7	39	38	42	0	7	26	13	7	2	1.56
4	6	1	17	14	12	16	62	56	4	20	1	1.14
...	...	...	...	...	...	...	...	...	...	...	...	...

Tabla 4-6: Tabla de datos simulados para el Caso A.

Caso II

El juego de gráficos mostrados en la Figura (4-5) ilustra el comportamiento de las variables del primer nivel. Como el experto lo mencionó: los datos tienen tendencia creciente con ligeras disminuciones en algunos momentos del tiempo. Asimismo cada grupo de datos evoluciona de forma particular lo cual se aprecia en las diferentes pendientes (gráfico superior derecho).

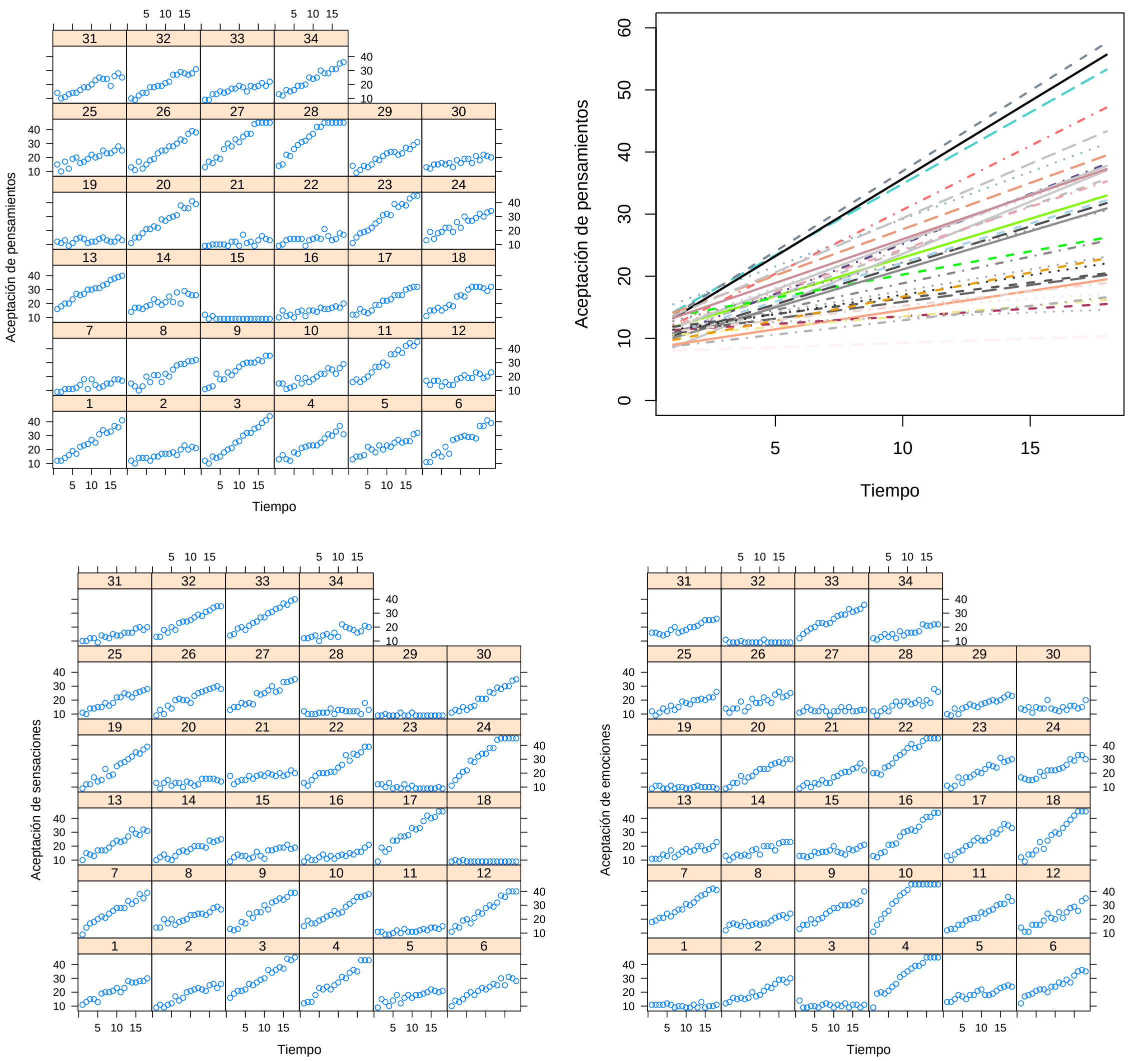

Figura 4-5: Gráficos del Nivel 1.

El juego de gráficos mostrados en la Figura (4-6) ilustra el comportamiento de las variables del primer nivel. Los niveles de *craving*, ansiedad y depresión presentan una tendencia a disminuir conforme el tiempo transcurre, es decir, conforme la aceptación de pensamientos, sensaciones y emociones incrementa. Cada paciente aprende y emplea las destrezas enseñadas por el terapeuta con cierta particularidad lo cual se refleja en la velocidad con la cual las puntuaciones disminuyen.

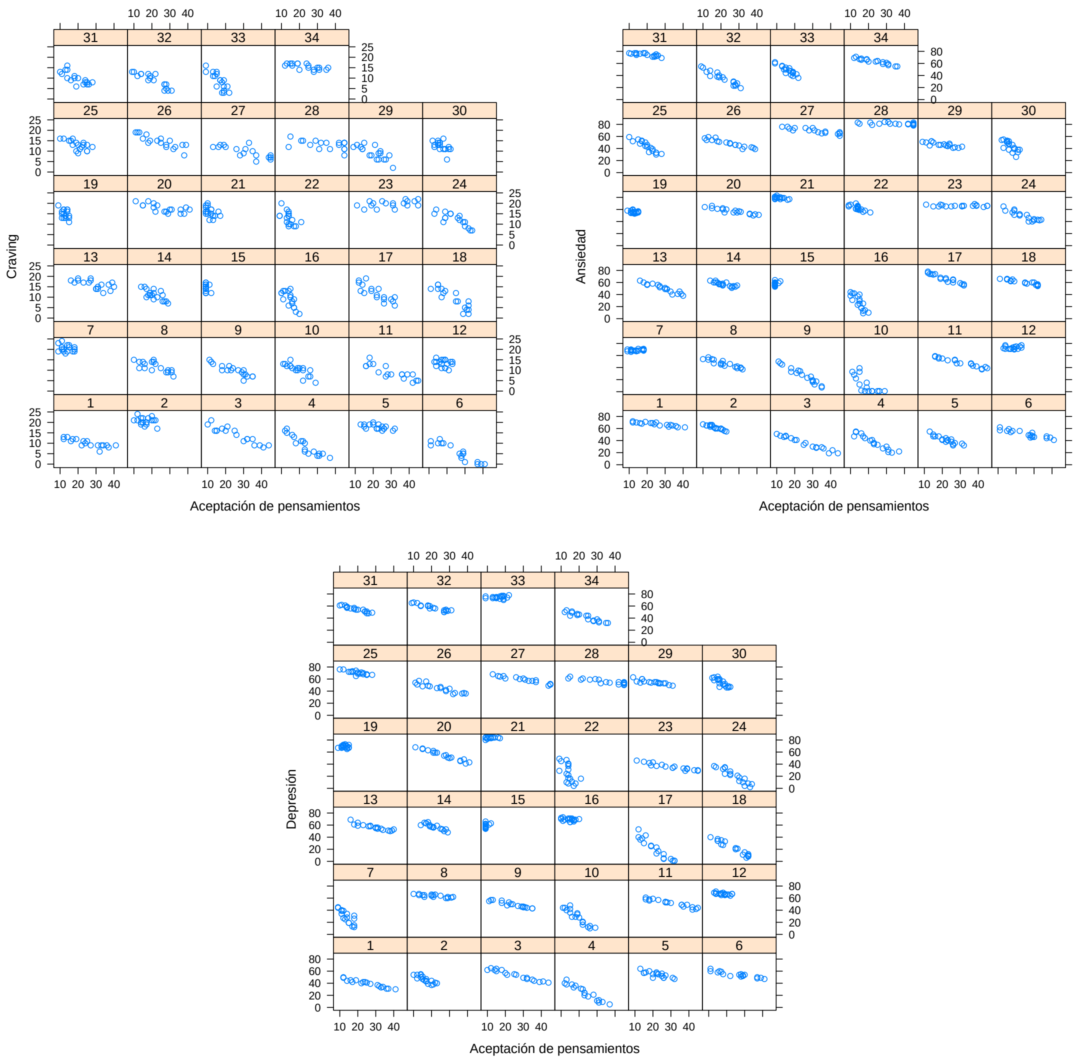

Figura 4-6: Gráficos del Nivel 2.

La gráfica (4-7) presenta el comportamiento de la variable consumo. Como podemos ver se aprecia una ligera tendencia a la baja.

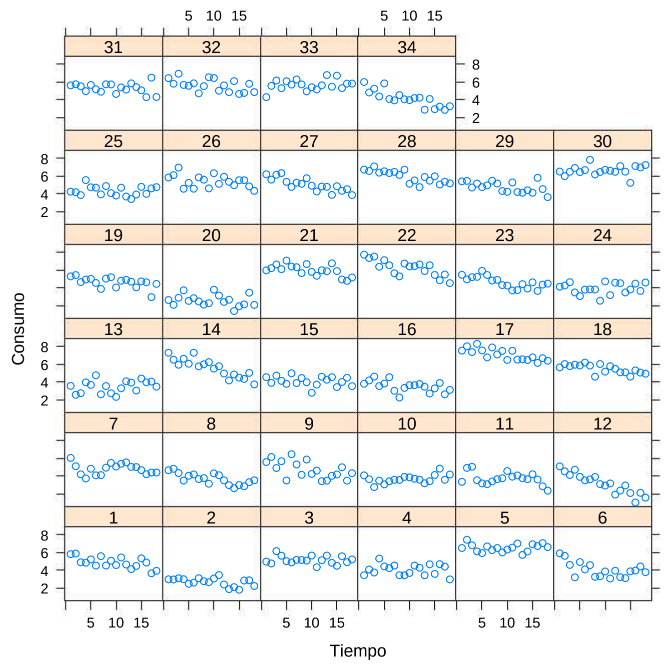

Figura 4-7: Gráfico del Nivel 3.

En la Tabla (4-7) podemos ver los valores numéricos ya representados en los gráficos de las figuras (4-5) a la (4-7). Así comprobamos las tendencias de cada uno de los grupos de datos.

Sustancia	*Individuo*	*Tiempo*	*Ace Pens*	*Ace Sens*	*Ace Emoc*	*Craving*	*Ansiedad*	*Depresión*	*Años Cons*	*Días S/Co*	*Tratam Prev*	*Unidades Cons*
4	8	1	16	18	12	22	34	67	7	10	1	2.31
4	8	2	14	10	13	18	42	68	7	10	1	2.19
4	8	3	17	13	15	20	36	67	7	10	1	1.59
4	8	4	21	12	15	21	33	67	7	10	1	2.22
4	8	5	24	18	16	18	27	70	7	10	1	2.27
4	8	6	18	16	16	22	39	71	7	10	1	2.21
4	8	7	22	18	16	18	24	69	7	10	1	2.21
4	8	8	20	17	17	19	30	64	7	10	1	1.77
4	8	9	23	18	18	18	26	67	7	10	1	1.57
4	8	10	28	25	20	20	16	73	7	10	1	2.22
4	8	11	33	20	18	21	17	68	7	10	1	1.96
4	8	12	30	22	20	20	15	70	7	10	1	1.85
4	8	13	33	21	28	21	19	68	7	10	1	1.97
4	8	14	34	23	26	18	10	72	7	10	1	1.72
4	8	15	39	22	32	19	6	73	7	10	1	1.96
4	8	16	43	28	34	18	1	74	7	10	1	1.85
4	8	17	39	29	33	18	1	77	7	10	1	1.29
4	8	18	42	27	34	19	1	71	7	10	1	1.09
...	...	...	...	...	...	...	...	...	...	...	...	...
5	15	1	11	13	15	15	76	62	13	20	1	51.6
5	15	2	11	14	17	15	75	64	13	20	1	71.08
5	15	3	12	19	15	13	74	62	13	20	1	74.35
5	15	4	15	22	17	14	73	64	13	20	1	64.73
5	15	5	21	21	21	13	69	60	13	20	1	32.49
5	15	6	18	20	26	13	68	63	13	20	1	62.17
5	15	7	24	25	30	11	70	57	13	20	1	84.66
5	15	8	26	33	32	11	67	64	13	20	1	66.9
5	15	9	23	32	34	10	72	55	13	20	1	51.22
5	15	10	24	31	36	11	73	55	13	20	1	58.44
5	15	11	26	38	31	7	73	50	13	20	1	61.29

Tabla 4-7: Tabla de datos simulados para el Caso II.

4.3. Estimación de parámetros

Uno de los objetivos del presente trabajo es determinar el tamaño de muestra y el número de observaciones repetidas por unidad muestral; es decir recomendar al experto qué valor deberían tomar las variables número de pacientes y número de observaciones por paciente de tal forma que el análisis de los datos (obtenidos al realizar el trabajo de campo) arroje un resultado estadísticamente significativo. Ante tal situación requerimos revisar qué tan precisas y confiables son las estimaciones de los parámetros para cada uno de los escenarios mostrados en la Tabla (4-1). Conocer diferentes resultados nos permitirá hacer comparaciones para finalmente poder decidir cuál escenario es el más recomendable.

La estructura general del modelo pide calcular estimadores puntuales e intervalos de confianza para cada uno de los parámetros involucrados en el correspondiente nivel, sin embargo, solamente presentaremos los de aquellos que pertenecen al Nivel 3. El motivo de esta consideración se fundamenta en que la variable dependiente, nivel de consumo, es nuestro principal centro de atención para poder responder al experto al respecto de la eficacia de su tratamiento. Es posible obtener el resto de los parámetros empleando el código del Apéndice A.

Para cada uno de los cuatro escenarios con sus respectivos subcasos se mostrará la tabla con los estimadores puntuales e intervalos de confianza del Nivel 3. Compararemos cada tabla con la información correspondiente mostrada en la Tabla (4-5) de parámetros sugeridos, para conocer sobre la precisión del modelo.

Los resultados mostrados, estimaciones de parámetros e intervalos de confianza, son obtenidos al emplear una sola vez el simulador, es decir, no se hizo un promedio de resultados recabados al realizar varias corridas del simulador. Se hicieron diferentes pruebas de estimaciones para cada uno de los subcasos de cada escenario y de forma empírica se llego a la conclusión de que eran similares; por tal motivo se presenta un único resultado del simulador.

Caso A

«¿Qué pasaría si el estudio se realizará con 10 pacientes y 7 mediciones de cada una de las variables por persona?»

Diez pacientes parece ser un número muy accesible para el experto; sin embargo, estadísticamente es considerablemente pequeño. Si bien es cierto que en total contamos con 70 datos, también recordemos que un modelo lineal mixto toma en cuenta que las observaciones de cada paciente están correlacionadas entre sí con lo cual tendremos más parámetros por estimar en comparación con un modelo de regresión lineal múltiple. Con esta información posiblemente no se obtengan resultados concluyentes.

Al comparar los resultados de las Tablas (4-8) y (4-5), a primera vista, aproximadamente la mitad de los estimadores puntuales están próximos a los reales (estos últimos son los valores propuestos, con los cuales se hicieron las simulaciones); por ejemplo, para el caso de φ, su estimador es 0.016 mientras que el valor poblacional es 0.017. Dos de los estimadores son negativos, muy distantes de los valores reales; por ejemplo s, su estimador es -0.051 mientras que el dato verdadero es 0.0076.

Notemos que todos los intervalos de confianza cubren a los valores verdaderos. La observación que hacemos al respecto de los intervalos es que todos (con excepción del correspondiente a σ_ξ) cubren al cero dando paso a dos posibilidades: la primera es que las variables asociadas a los parámetros podrían no ser necesarias para la explicación de la variable independiente, esto cuando el estimador sea cero o muy cercano a cero; la segunda es que no tendríamos seguridad del signo que adquiere el estimador del parámetro pues podría ser tanto positivo como negativo, lo cual a su vez no permitiría concluir al respecto de un aumento o disminución en la variable nivel de consumo. En otras palabras, el parámetro resultaría no significativo si se realizara la prueba de hipótesis asociada.

Caso I

«¿Qué pasaría si el estudio se realizará con 12 ó bien 30 pacientes y 12 mediciones de cada una de las variables por persona?»

Al comparar los resultados de la Tabla (4-9) con sus respectivos de la Tabla (4-5), tenemos que para ambos escenarios (con 12 ó 30 pacientes), a pesar de que dos de los estimadores son negativos cuando deberían ser positivos, aproximadamente un poco más de la mitad del total de los estimadores de los parámetros son valores próximos a los reales. La diferencia entre ambos escenarios es que para el caso de 30 pacientes los estimadores puntuales mejoran sobre los obtenidos en el caso con 12 pacientes.

Parámetro	*Estimador puntual*	*Intervalo de confianza* $\alpha = 0.05$
κ	2.207	(-0.998, 5.213)
λ	0.0001	(-0.060, 0.059)
φ	0.016	(-0.003, 0.035)
ψ	0.013	(-0.008, 0.035)
p	0.012	(-0.176, 0.202)
q	-0.004	(-0.018, 0.010)
s	-0.051	(-0.714, 0.605)
σ_ξ	0.492	(0.389, 0.578)
σ_{g}	0.0001	(0.000, 0.907)
σ_{h}	0.000	(0.000, 0.063)
σ_{m}	0.018	(0.000, 0.028)
σ_{n}	0.008	(0.000, 0.020)

Tabla 4-8: Estimaciones del Nivel 3 para el Caso A.

Los intervalos de confianza mejoran conforme aumenta el número de pacientes. Con 12 pacientes todos los intervalos (con excepción de los correspondientes a ψ y σ_ξ) cubren el cero. Cuando el número de personas incrementa a 30 sucede que únicamente 7 intervalos de un total de 12, es decir, un 58 % cubren el cero. En general observamos mejores resultados con 30 pacientes que con sólo 12.

Al no ser significativos los parámetros de los efectos fijos, puede ocurrir que en realidad no estén influyendo sobre la respuesta (como p, q y s) o puede ser que haya efectos aleatorios que tengamos que quitar del modelo. Resulta un poco sorprendente que al final quizá el único efecto aleatorio que valdría la pena tomar es σ_{m} más el residuo σ_ξ. El hecho de no resultar significativos, en este caso, lo atribuimos a que el número de mediciones por paciente es pequeño. Como notaremos más adelante esta última situación se presenta para los subcasos de los dos escenarios restantes, lo cual nos lleva a pensar que realmente el número de parámetros por estimar es muy grande y la información muy escasa o bien podría influir la incertidumbre acumulada (no tomada en cuenta) al pasar de un nivel a otro.

A pesar de que se aprecia una mejoría en los estimadores puntuales y en los intervalos de confianza al incrementar el número de pacientes en el Caso I, pareciera que se pudieran mejorar aún más los resultados ya que los obtenidos en la presente situación, al igual que en el Caso A, no permitirán concluir al respecto de los cambios en la variable consumo. Observemos que

Parámetro	*Estimador puntual (12 Pacientes)*	*Intervalo de confianza* $\alpha = 0.05$		*Estimador puntual (30 Pacientes)*	*Intervalo de confianza* $\alpha = 0.05$
κ	1.359	(-0.533, 3.292)		2.133	(1.264, 2.973)
λ	0.038	(-0.019, 0.093)		0.025	(-0.003, 0.051)
φ	0.014	(-0.002, 0.031)		0.019	(0.010, 0.027)
ψ	0.023	(0.002, 0.044)		0.014	(0.006, 0.023)
p	-0.044	(-0.194, 0.115)		-0.005	(-0.056, 0.049)
q	-0.004	(-0.062, 0.054)		0.007	(-0.007, 0.022)
s	0.506	(-0.640, 1.604)		-0.054	(-0.318, 0.217)
σ_ξ	0.574	(0.501, 0.641)		0.551	(0.505, 0.594)
σ_{g}	0.000	(0.000, 0.930)		0.000	(0.000, 0.513)
σ_{h}	0.036	(0.000, 0.086)		0.033	(0.000, 0.057)
σ_{m}	0.020	(0.000, 0.033)		0.016	(0.008, 0.021)
σ_{n}	0.014	(0.000, 0.029)		0.009	(0.000, 0.016)

Tabla 4-9: Estimaciones del Nivel 3 para el Caso I.

los intervalos del Caso A son menos confiables que los del Caso I los cuales a su vez podrían mejorar.

Caso II

«¿Qué pasaría si el estudio se realizará con 12, 25, 34 ó bien 42 pacientes y 18 mediciones de cada una de las variables por persona?»

Este escenario considera un poco más del doble de mediciones de cada una de las variables en comparación con el Caso A, lo cual, en primera instancia, podría sugerir mejores estimaciones a pesar de que la cantidad de pacientes no incremente en el mismo grado. El incremento del número de observaciones por paciente se ve muy limitado por el número de sesiones de terapia. Sin embargo, ahora en el presente caso consideraremos alternativas diferentes a las dictadas por las características del tratamiento.

En la realización del estudio 18 observaciones por paciente parece ser una cantidad razonable, a pesar de que el experto no haya considerado anteriormente esta posibilidad. Por tal motivo este Caso II merece atención especial; parece ser que será una de las mejores opciones. Revisemos qué sucede si se consideran 12, 25, 34 y 42 pacientes.

Al comprar las Tablas (4-10) y (4-11) contra la Tabla (4-5) las estimaciones puntuales se

Parámetro	***Estimador puntual (12 Pacientes)***	***Intervalo de confianza*** $\alpha = 0.05$	***Estimador puntual (25 Pacientes)***	***Intervalo de confianza*** $\alpha = 0.05$
κ	1.088	(-1.381, 3.516)	2.268	(0.509, 4.056)
λ	0.003	(-0.034, 0.040)	0.046	(0.023, 0.069)
φ	0.021	(-0.005, 0.046)	0.012	(-0.001, 0.026)
ψ	0.020	(0.005, 0.036)	0.016	(0.004, 0.028)
p	0.033	(-0.113, 0.183)	-0.032	(-0.129, 0.066)
q	0.018	(-0.028, 0.066)	0.013	(0.006, 0.021)
s	-0.020	(-0.624, 0.601)	-0.113	(-0.654, 0.416)
σ_{ξ}	0.503	(0.452, 0.551)	0.577	(0.536, 0.614)
σ_{g}	0.199	(0.000, 1.053)	0.097	(0.000, 0.748)
σ_{h}	0.035	(0.000, 0.069)	0.000	(0.000, 0.044)
σ_{m}	0.019	(0.000, 0.029)	0.018	(0.000, 0.027)
σ_{n}	0.000	(0.000, 0.020)	0.017	(0.000, 0.025)

Tabla 4-10: Estimaciones del Nivel 3 para el Caso II.

aproximan cada vez más a los valores reales conforme el número de pacientes incrementa. Este proceso ocurre de forma progresiva, si comparamos el caso de 12 pacientes contra el de 42 pacientes vemos que en el segundo los estimadores puntuales son más próximos a los reales. Otra situación que se presenta, aunque no requiere especial atención, es que para el parámetro s el estimador obtenido siempre es negativo cuando debería ser positivo.

Para el caso de los intervalos de confianza observamos un fenómeno similar, a mayor número de pacientes los intervalos adquieren mayor precisión. Para el escenario con 12 pacientes 10 intervalos cubren al cero; con 25 pacientes solamente 7; y para el caso de 34 y 42 pacientes solamente 4 cubren al cero.

Observemos que al comprar los dos últimos escenarios, en el último, los estimadores puntuales se observan más cercanos a los valores reales; sin embargo, la cantidad de intervalos que cubren al cero es la misma. En este sentido resulta prudente considerar al caso con 34 pacientes como la opción óptima para ser posible de realizarse.

Caso III

«¿Qué pasaría si el estudio se realizará con 12 ó bien 42 pacientes y 30 mediciones de cada una de las variables por persona?»

Parámetro	***Estimador puntual (34 Pacientes)***	***Intervalo de confianza*** $\alpha = 0.05$	***Estimador puntual (42 Pacientes)***	***Intervalo de confianza*** $\alpha = 0.05$
κ	1.577	(0.645, 2.470)	1.732	(0.842, 2.608)
λ	0.039	(0.019, 0.057)	0.022	(0.006, 0.039)
φ	0.013	(0.006, 0.021)	0.018	(0.009, 0.028)
ψ	0.019	(0.011, 0.028)	0.020	(0.011, 0.029)
p	0.059	(0.012, 0.106)	0.009	(-0.052, 0.069)
q	0.008	(0.003, 0.012)	0.011	(0.004, 0.017)
s	-0.177	(-0.503, 0.154)	-0.113	(-0.414, 0.193)
σ_{ξ}	0.546	(0.513, 0.579)	0.0002	(0.522, 0.581)
σ_{g}	0.151	(0.000, 0.563)	0.000	(0.000, 0.589)
σ_{h}	0.030	(0.000, 0.046)	0.016	(0.000, 0.033)
σ_{m}	0.003	(0.000, 0.011)	0.018	(0.005, 0.021)
σ_{n}	0.014	(0.003, 0.018)	0.552	(0.010, 0.024)

Tabla 4-11: Estimaciones del Nivel 3 para el Caso II.

Observemos que los estimadores puntuales mejoran, es decir, se aproximan a los valores reales cuando incrementamos el número de pacientes a un poco más del triple. Los intervalos de confianza adquieren mayor precisión debido a la misma situación; menor cantidad de ellos cubren al cero.

Podemos decir que el último escenario revisado: 42 pacientes y 30 observaciones por persona es el mejor en cuanto a estimaciones puntuales. Para el caso de los intervalos de confianza 5 de un total de 12, es decir, 41.5 % cubren al cero. Esta última situación es similar al escenario con 34 pacientes y 18 mediciones por sujeto. Estadísticamente, el segundo escenario del Caso III es una buena opción, sin embargo, para el experto sería muy conflictivo alcanzar a recolectar tal cantidad de información en el tiempo que tiene disponible.

4.3.1. Intervalos de confianza por tipo de sustancia

En la ecuación (3-7) del Capítulo 3 mostramos el modelo lineal mixto que considera a la variable nivel de consumo (H) como independiente. Recordemos que esta variable está reescalada, es decir, para obtener el nivel de consumo dependiendo el tipo de sustancia psicoactiva tenemos que multiplicar el resultado que nos da el modelo indicado por la correspondiente desviación estándar mostrada en la Tabla (3-2). Para el modelo (3-7) hemos obtenido estimadores

Parámetro	***Estimador puntual (12 Pacientes)***	***Intervalo de confianza*** $\alpha = 0.05$		***Estimador puntual (42 Pacientes)***	***Intervalo de confianza*** $\alpha = 0.05$
κ	1.730	(0.722, 2.706)		1.686	(0.973, 2.382)
λ	0.016	(-0.008, 0.043)		0.036	(0.022, 0.051)
φ	0.035	(0.022, 0.046)		0.023	(0.016, 0.029)
ψ	0.027	(0.016, 0.039)		0.019	(0.013, 0.025)
p	0.037	(-0.027, 0.106)		0.007	(-0.029, 0.042)
q	0.004	(-0.002, 0.010)		0.009	(0.006, 0.012)
s	-0.234	(-0.549, 0.096)		0.088	(-0.124, 0.291)
σ_ξ	0.503	(0.465, 0.541)		0.518	(0.497, 0.539)
σ_{g}	0.183	(0.000, 0.465)		0.000	(0.000, 0.393)
σ_{h}	0.000	(0.000, 0.033)		0.022	(0.000, 0.037)
σ_{m}	0.000	(0.000, 0.012)		0.014	(0.007, 0.019)
σ_{n}	0.012	(0.000, 0.017)		0.010	(0.000, 0.015)

Tabla 4-12: Estimaciones del Nivel 3 para el Caso III.

puntuales e intervalos de confianza para cada uno de los parámetros, ahora veremos qué indica cada uno de ellos en términos del estudio psicológico.

Mostremos un ejemplo; para los demás estimadores la interpretación es similar. Tomemos $\hat{\lambda}$ el estimador del parámetro λ. Este estimador representa el incremento que experimenta el nivel de consumo cuando la variable *craving* (D) aumenta su valor en una unidad mientras el resto de las variables se mantienen constantes.

Si obtenemos un intervalo de confianza para $\hat{\lambda}$ y multiplicamos cada extremo del mismo por la desviación estándar de cada una de las sustancias obtendremos los diferentes intervalos de confianza por tipo de sustancia. Abajo presentamos estos resultados para el Nivel 3 del modelo para los Casos A y II (para el escenario con 34 pacientes) en las Tablas (4-13) y (4-14) respectivamente.

Al revisar la Tabla (4-14) vemos que, por ejemplo, tomando al intervalo de confianza correspondiente al estimador del parámetro λ para la sustancia alcohol, lo podemos interpretar como la incertidumbre del incremento que experimenta el nivel de consumo de alcohol cuando la variable *craving* (D) aumenta su valor en una unidad mientras las demás variables se mantienen constantes.

En general observamos que los intervalos de confianza mostrados en la Tabla (4-14) no

Parámetro	***Alcohol***	***Cocaína***	***Mariguana***	***Piedra***	***Pvc***
κ	(-36.58, 191.16)	(-0.50, 2.61)	(-0.64, 3.37)	(-0.50, 2.61)	(-20.78, 108.61)
λ	(-2.22, 2.18)	(-0.03, 0.03)	(-0.04, 0.04)	(-0.03, 0.03)	(-1.26, 1.24)
φ	(-0.09, 1.27)	(-0.001, 0.02)	(-0.002, 0.02)	(-0.001, 0.02)	(-0.05, 0.72)
ψ	(-0.29, 1.28)	(-0.004, 0.02)	(-0.01, 0.02)	(-0.004, 0.02)	(-0.16, 0.73)
p	(-6.46, 7.42)	(-0.09, 0.10)	(-0.11, 0.13)	(-0.09, 0.10)	(-3.67, 4.22)
q	(-0.65, 0.37)	(-0.01, 0.01)	(-0.01, 0.01)	(-0.01, 0.01)	(-0.37, 0.21)
s	(-26.16, 22.18)	(-0.36, 0.30)	(-0.46, 0.39)	(-0.36, 0.30)	(-14.87, 12.60)

Tabla 4-13: Intervalos de confianza del Nivel 3 por sustancia psicoactiva para el Caso A.

Parámetro	***Alcohol***	***Cocaína***	***Mariguana***	***Piedra***	***Pvc***
κ	(23.66, 90.56)	(0.32, 1.23)	(0.42, 1.60)	(0.32, 1.23)	(13.45, 51.46)
λ	(0.71, 2.10)	(0.01, 0.03)	(0.01, 0.04)	(0.01, 0.03)	(0.41, 1.19)
φ	(0.21, 0.75)	(0.003, 0.01)	(0.004, 0.01)	(0.003, 0.01)	(0.12, 0.43)
ψ	(0.41, 1.01)	(0.01, 0.01)	(0.01, 0.02)	(0.01, 0.01)	(0.23, 0.57)
p	(0.43, 3.89)	(0.01, 0.05)	(0.01, 0.07)	(0.01, 0.05)	(0.25, 2.21)
q	(0.11, 0.46)	(0.002, 0.01)	(0.002, 0.01)	(0.002, 0.01)	(0.06, 0.26)
s	(-18.44, 5.63)	(-0.25, 0.08)	(-0.32, 0.10)	(-0.25, 0.08)	(-10.48, 3.20)

Tabla 4-14: Intervalos de confianza del Nivel 3 por sustancia psicoactiva para el Caso II con 34 pacientes.

presentan rangos que sean importantes para el experto. Por ejemplo, para el psicólogo no dice nada el hecho de que un paciente tome media copa de alcohol ó que tome 2 copas a la semana, esto no implica una disminución relevante en la cantidad de consumo.

Capítulo 5

Aportaciones y Conclusiones

En términos generales diremos que hemos construido un modelo estadístico *ad hoc* para la representación de un estudio en psicología. Uno de los propósitos de este trabajo fue establecer un modelo que expresara la relación entre las variables del estudio tal cual fue discernida mediante interacción con el experto. Se buscó conocer si la variación existente en la variable nivel de consumo de sustancias psicoactivas (variación que se espera sea decreciente) se puede explicar como función un conjunto de variables independientes. El análisis de datos experimentales a través de este modelo nos permitirá conocer si el tratamiento psicológico que manipula a todas estas variables es o no eficaz, es decir, podríamos conocer si se logra que haya una disminución significativa en el consumo de sustancias psicoactivas por parte de las personas que toman el tratamiento.

El modelo quedó establecido como un modelo de 3 etapas, cada una de las cuales conformada por uno o tres modelos lineales con efectos mixtos o también llamados modelos lineales mixtos. Este planteamiento requirió tanto de la revisión de las bases teóricas de esta clase tan particular de modelos estadísticos, como de la interacción continua psicólogo-estadístico (el trabajo multidisciplinario se ve reflejado en cada paso de la formulación del modelo ya señalado) obteniendo así un modelo pertinente y útil.

Con ayuda de un simulador se hicieron diferentes juegos de datos considerados cada uno como una muestra aleatoria real. Con la información construida de esta forma se puso en marcha la metodología estadística que será empleada una vez que se tenga la información real. La metodología estadística arrojará como resultados estimadores puntuales e intervalos de confianza para cada uno de los parámetros. Si los resultados son estadísticamente significativos se podrán generalizar hacia el resto de la población mexicana consumidora de sustancias psicoactivas. El análisis de diferentes escenarios ha permitido que lleguemos a proponer como la opción más viable a aquella que toma en cuenta 34 pacientes y 18 observaciones por paciente para el estudio. Un experimento con estas características permitirá al experto obtener resultados estadísticamente significativos (en caso de sí existir efectos fijos y aleatorios como los propuestos).

5.1. Aportaciones y conclusiones específicas al respecto del fenómeno psicológico

El Capítulo I nos presentó un problema a resolver en el área de la salud mental. Después de todos los análisis hechos al respecto es momento de indicar las conclusiones a las cuales llegamos. Esperemos que los comentarios y las conclusiones aquí mostrados ayuden al experto a re-direccionar, es decir, a reajustar las características del trabajo experimental que actualmente está por iniciar.

Pues bien, concluimos con lo siguiente:

a) El modelo que relaciona al conjunto de variables en la forma expuesta por el experto es un modelo por etapas, lineal con efectos mixtos.

b) El software estadístico R cuenta con una librería, $lme4$, que permite el análisis de datos cuyo comportamiento es representado a través del dicho modelo lineal mixto.

c) A través del análisis de diferentes escenarios simulados se logra concluir que un número aproximado de pacientes requeridos para el estudio es 34 y un número aproximado de mediciones por paciente serían 18. Se sugiere al experto que, de ser posible, extienda el número de sesiones del tratamiento, o de otra forma, prevea un seguimiento de los pacientes durante un tiempo pertinente de tal forma que se puedan conseguir las 18 observaciones por persona, aproximadamente.

d) Al respecto de realizar el estudio con la cantidad de información propuesta por el psicólogo antes del inicio del presente trabajo, hemos de decir que si bien las estimaciones puntuales son un tanto certeras, los intervalos de confianza no permitirán obtener resultados concluyentes pues hemos visto que todos cubren tanto valores positivos como negativos incluido el cero.

e) Al analizar los resultados se observa que la variable número de tratamientos previos resulta en todos los escenario tener un estimador negativo, además, el respectivo intervalo de confianza cubre valores tanto positivos como negativos, a partir de lo cual se concluye que esta variable no aporta información al modelo.

f) El emplear información simulada de un modelo, lo que llamamos Nivel 1, para estimar los parámetros correspondientes de los modelos lineales mixtos del Nivel 2 y similar para el Nivel 3, implica variabilidad adicional no considerada. Esto es, quizá los intervalos de confianza e inferencia en el Nivel 3 tienen una variabilidad distinta a la que arroja el modelo ajustado en esa etapa. No se prestó especial atención en este hecho debido a que cuando se tengan los datos reales este problema no estará presente; aunque reconocemos que esta situación pudiese ser la causa de que algunos de los efectos aleatorios resultaran ser no significativos.

g) Finalmente es importante reconocer que se han dado por hecho ciertas afirmaciones al respecto del modelo que posiblemente podrían no resultar del todo ciertas cuando se analicen los datos reales, como lo es la hipótesis de linealidad. Se ha propuesto un modelo lineal; sin embargo, los datos recolectados en el trabajo de campo podrían sugerir un modelo no lineal. Por otro lado, posiblemente no se han considerado todas las variables necesarias en la formulación del modelo, lo cual se podría mejorar. Así mismo se especificaron gran cantidad de efectos aleatorios, una vez que se cuente con información real se podría considerar anular algunos de ellos; esto con la finalidad de mejorar los estimadores. Luego, se hicieron suposiciones al respecto de las distribuciones de los efectos aleatorios y de la distribución de las perturbaciones involucradas en los modelos presentados en cada nivel; sin embargo, cuando se cuente con los datos reales se tendrá que revisar que estas afirmaciones se cumplan o estén lo más cerca de cumplirse para poder afirmar que el modelo construido con la estructura ya dicha explica el fenómeno en cuestión. Quizás se pudieran considerar modelos lineales mixtos generalizados con alguna distribución de los efectos aleatorios mas acorde, digamos poisson, binomial negativa, *etc.* Reconocer que puede suceder lo anteriormente dicho, es parte de mencionar las debilidades que presenta el trabajo, para que una vez que se cuente con información real se preste especial atención en la validación y, de ser posible, se trabaje sobre ello para mejorar el modelo.

5.2. Conclusiones y comentarios generales

El proceso de planteamiento de un modelo conlleva una serie de pasos que son indispensables para una adecuada formulación del mismo. En primer lugar, se trata de prestar atención a algún fenómeno natural o bien plantearse alguna pregunta al respeto de algún fenómeno construido por la mano del hombre. Una vez con las interrogantes en mente el trabajo en conjunto del estadístico y el experto en el área permitirá que se identifiquen las variables y la relación que puede existir entre ellas, es decir, la manera en la cual interactúan.

Cuando se ha detectado alguna teoría estadística que permita explicar el comportamiento de las variables, hay que empezar primero por revisar el modelo más básico de dicha metodología. Hacerlo de esta forma permite conocer las bases bajo las cuales se sustenta el modelo para posteriormente modificar las hipótesis de tal forma que se pueda llegar a un modelo más sofisticado el cual posiblemente sea más *ad hoc*. Algunas ocasiones bastará con la primera aproximación del modelo; todo depende de la pregunta a resolver.

Después de revisado el paso anterior procedemos a los cálculos de estimadores correspondientes con la metodología estadística propuesta. Los resultados obtenidos nos permitirán concluir al respecto de los cuestionamientos planteados por el experto.

La interacción estadístico-usuario de la cual hemos venido hablando a lo largo del presente estudio, es una forma de trabajo gratificante para ambos profesionistas. Permite construir modelos con base en hipótesis lo más realistas posibles y además deja nuevo conocimiento en un área diferente para cada una de las partes. Actualmente, debido a la profundidad con la cual cada profesión trabaja, es que se ha recurrido al trabajo interdisciplinario, pues para una sola persona es complicado poder manipular libremente diferentes áreas de investigación.

La importancia de la formulación de un pre-estudio a través del empleo de un simulador permite a los investigadores conocer algunas de las consecuencias que puede tener el trabajar con experimentos que tengan determinadas características. Simular el experimento propuesto por un investigador podría resultar ser de gran utilidad pues se podrían hacer mejorías de la propuesta del diseño experimental antes de iniciar el trabajo de campo. Estas mejorías se podrían ver reflejadas en una nueva manipulación de variables ó en una mejor distribución de: tiempo, recursos económicos, recursos humanos, *etc.*

Apéndice A

Simulador (Código en R)

```
library(lme4)
library(lattice)
rm(list = ls())

n <- 34     # Número de pacientes
obs <- 18    # Número de observaciones por persona

sdut1_1 <- 1.47
sdut1_2 <- 0.72
sdut2_1 <- 1.48
sdut2_2 <- 0.74
sdut3_1 <- 1.5
sdut3_2 <- 0.71
sduu <- 2.8

sdwc_1 <- 1.56
sdwc_2 <- 0.11
sdwc_3 <- 0.11
sdwc_4 <- 0.12
sdwa_1 <- 6.24
sdwa_2 <- 0.15
sdwa_3 <- 0.4
sdwa_4 <- 0.44
sdwd_1 <- 6.12
sdwd_2 <- 0.15
sdwd_3 <- 0.39
sdwd_4 <- 0.43
sdww <- 1.58

sdz_1 <- 0.01
sdz_2 <- 0.017
sdz_3 <- 0.015
sdz_4 <- 0.013
sdzz <- .55

d_0 <- 11.01
d_1 <- 1.08
b_0 <- 20.52
b_1 <- -0.12
b_2 <- -0.14
b_3 <- -0.13
bd_0 <- 72.5
bd_1 <- -0.32
bd_2 <- -0.31
bd_3 <- -0.32
k_0 <- 1.72
k_1 <- 0.025
k_2 <- 0.0192
k_3 <- 0.0181
k_4 <- 0.009
k_5 <- 0.0081
k_6 <- 0.0072

sd_coc <- 0.5
sd_pie <- 0.5
sd_mar <- 0.65
sd_alc <- 36.67
```

```
sd_pvc <- 20.83

ind <- rep(factor(seq(from=1, to=n)), each=obs)

b_poblacion <-read.csv("C:/Users/ANA PAULINA PEREZ R/Desktop/poblacion.csv")

años_cons <-c()
dias_sin_cons <-c()
tipo_sust <-c()
sd_sust <-c()

for (i in 1:n)
{
aleatorio <- sample(1:29, 1, replace=T)
años <- rep(b_poblacion[aleatorio,1], each=obs)
dias <- rep(b_poblacion[aleatorio,2], each=obs)
sustancia <- rep(b_poblacion[aleatorio,3], each=obs)
sd_sustancia <-rep(b_poblacion[aleatorio,4], each=obs)
años_cons <-c(años_cons,años)
dias_sin_cons <-c(dias_sin_cons,dias)
tipo_sust <-c(tipo_sust,sustancia)
sd_sust <-c(sd_sust,sd_sustancia)
}

tratam_previos <- rep(sample(1:3, n, replace=T), each=obs)

#---############## Nivel 1 ##############---#
#############################################
tiem <- rep(c(1:obs),n)
ut1_1 <- rep(rnorm(n,0,sdut1_1), each=obs)
ut1_2 <- rep(rnorm(n,0,sdut1_2), each=obs)
ut2_1 <- rep(rnorm(n,0,sdut2_1), each=obs)
ut2_2 <- rep(rnorm(n,0,sdut2_2), each=obs)
ut3_1 <- rep(rnorm(n,0,sdut3_1), each=obs)
ut3_2 <- rep(rnorm(n,0,sdut3_2), each=obs)

t_1 <- (d_0 + ut1_1) + (d_1*tiem + ut1_2*tiem) + rnorm(n*obs,0,sduu)
m_t1 <- lmer(t_1 ~ tiem+(1|ind)+(0+tiem|ind))
summary(m_t1)
coef(m_t1)
confint(m_t1, method="boot")
t_1 <- round(t_1,0)
t_1[t_1<9] <- 9
t_1[t_1>45] <- 45
xyplot(t_1~tiem|ind, xlab="Tiempo", ylab="Aceptación de pensamientos")

t_2 <- (d_0 + ut2_1) + (d_1*tiem + ut2_2*tiem) + rnorm(n*obs,0,sduu)
m_t2 <- lmer(t_2 ~ tiem+(1|ind)+(0+tiem|ind))
summary(m_t2)
coef(m_t2)
confint(m_t2, method="boot")
t_2 <- round(t_2,0)
t_2[t_2<9] <- 9
t_2[t_2>45] <- 45
xyplot(t_2~tiem|ind, xlab="Tiempo", ylab="Aceptación de sensaciones")

t_3 <- (d_0 + ut3_1) + (d_1*tiem + ut3_2*tiem) + rnorm(n*obs,0,sduu)
m_t3 <- lmer(t_3 ~ tiem+(1|ind)+(0+tiem|ind))
summary(m_t3)
coef(m_t3)
confint(m_t3, method="boot")
t_3 <- round(t_3,0)
t_3[t_3<9] <- 9
t_3[t_3>45] <- 45
xyplot(t_3~tiem|ind, xlab="Tiempo", ylab="Aceptación de emociones")

#---############## Nivel 2 ##############---#
#############################################
wc_1 <- rep(rnorm(n,0,sdwc_1), each=obs)
wc_2 <- rep(rnorm(n,0,sdwc_2), each=obs)
wc_3 <- rep(rnorm(n,0,sdwc_3), each=obs)
wc_4 <- rep(rnorm(n,0,sdwc_4), each=obs)
wa_1 <- rep(rnorm(n,0,sdwa_1), each=obs)
```

```
wa_2 <- rep(rnorm(n,0,sdwa_2), each=obs)
wa_3 <- rep(rnorm(n,0,sdwa_3), each=obs)
wa_4 <- rep(rnorm(n,0,sdwa_4), each=obs)
wd_1 <- rep(rnorm(n,0,sdwd_1), each=obs)
wd_2 <- rep(rnorm(n,0,sdwd_2), each=obs)
wd_3 <- rep(rnorm(n,0,sdwd_3), each=obs)
wd_4 <- rep(rnorm(n,0,sdwd_4), each=obs)

Cr <- (b_0 + wc_1) + (b_1*t_1 + wc_2*t_1) + (b_2*t_2 + wc_3*t_2) + (b_3*t_3 + wc_4*t_3) +
rnorm(n*obs,0,sdww)
mc <- lmer(Cr ~ t_1+t_2+t_3+(1|ind)+(0+t_1|ind)+(0+t_2|ind)+(0+t_3|ind))
summary(mc)
coef(mc)
confint(mc, method="boot")
Cr <- round(Cr,0)
Cr[Cr<0] <- 0
Cr[Cr>24] <- 24
xyplot(Cr~t_1|ind, xlab="Aceptación de pensamientos", ylab="Craving")
xyplot(Cr~tiem|ind, xlab="Tiempo", ylab="Craving")

An <- (bd_0 + wa_1) + (bd_1*t_1 + wa_2*t_1) + (bd_2*t_2 + wa_3*t_2) + (bd_3*t_3 + wa_4*t_3) +
rnorm(n*obs,0,sdww)
ma <- lmer(An ~ t_1+t_2+t_3+(1|ind)+(0+t_1|ind)+(0+t_2|ind)+(0+t_3|ind))
summary(ma)
coef(ma)
confint(ma, method="boot")
An <- round(An,0)
An[An<1] <- 1
An[An>84] <- 84
xyplot(An~t_1|ind, xlab="Aceptación de pensamientos", ylab="Ansiedad")
xyplot(An~tiem|ind, xlab="Tiempo", ylab="Ansiedad")

De <- (bd_0 + wd_1) + (bd_1*t_1 + wd_2*t_1) + (bd_2*t_2 + wd_3*t_2) + (bd_3*t_3 + wd_4*t_3) +
rnorm(n*obs,0,sdww)
md <- lmer(De ~ t_1+t_2+t_3+(1|ind)+(0+t_1|ind)+(0+t_2|ind)+(0+t_3|ind))
summary(md)
coef(md)
confint(md, method="boot")
De <- round(De,0)
De[De<1] <- 1
De[De>84] <- 84
xyplot(De~t_1|ind, xlab="Aceptación de pensamientos", ylab="Depresión")
xyplot(De~tiem|ind, xlab="Tiempo", ylab="Depresión")

#---############## Nivel 3 ##############---#
#############################################
z_1 <- rep(rnorm(n,0,sdz_1), each=obs)
z_2 <- rep(rnorm(n,0,sdz_2), each=obs)
z_3 <- rep(rnorm(n,0,sdz_3), each=obs)
z_4 <- rep(rnorm(n,0,sdz_4), each=obs)

consumo <- (k_0 + z_1) + (k_1*Cr + z_2*Cr) + (k_2*An + z_3*An) + (k_3*De + z_4*De) +
k_4*años_cons + k_5*dias_sin_cons + k_6*tratam_previos + rnorm(n*obs,0,sdzz)

m3 <- lmer(consumo ~
Cr+An+De+años_cons+dias_sin_cons+tratam_previos+(1|ind)+(0+Cr|ind)+(0+An|ind)+(0+De|ind))
summary(m3)
coef(m3)
confint(m3,level = 0.95, method="boot", nsim=2500)
xyplot(consumo~tiem|ind, xlab="Tiempo", ylab="Consumo")

#---############# Ajuste de la variable nivel de consumo #############---#
consumo2 <- round(consumo*sd_sust,2)
consumo2[consumo2<0] <- 0
plot(consumo2)
########################################################################

base <- data.frame(sustancia=tipo_sust, individuos=ind, tiempo=tiem, acept_pensa=t_1,
acept_sens=t_2, acept_emoc=t_3, craving=Cr, ansiedad=An, depresion=De, años_cons=años_cons,
días_sin_cons=dias_sin_cons, tratam_previos=tratam_previos,unidades_cons=consumo2)

write.csv(base,"basefin.csv")
```

Bibliografía

[1] Bates, Douglas *Computational Methods for Mixed Models.* Department of Statistics. University of Wisconsin-Madison, EUA. 2011.

[2] Casella, George & Berger, Roger *Statistical Inference.* Brooks/cole, Cengage Learning. California, EUA. 2002.

[3] Cheng, Chi-Lun & Van Ness, John *Statistical regression with measurement error.* John Wiley Sons. Nueva York, EUA. 1999.

[4] Demidenko, Eugene. *Mixed Models. Theory and Applications.* John Wiley & Sons. Nueva Jersey, EUA. 2004.

[5] Faraway, Julian. *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models.* CRC press. Florida, EUA. 2005.

[6] Fitzmaurice, Garrett, *et al*. *Longitudinal Data Analysis.* Chapman & Hall/CRC. Florida, EUA. 2009.

[7] Galecki, Andrzej & Burzykowski, Tomasz. *Linear Mixed-Effects Models Using R. A Step-by-step Approach.* Springer Science & Business Media. Nueva York, EUA. 2013.

[8] Hodges, Joseph & Lehmann, Erich. *Basic concepts of probability and statistics.* SIAM. California, EUA. 1970.

[9] Jiang, Jiming. *Linear and Generalized Linear Mixed Models and Their Applications.* Springer Science & Business Media. Nueva York, EUA. 2007.

[10] Pérez-Romero, Luis Angel. *Prevención Estructurada de Recaídas Basada en Destrezas de Enfrentamiento y de Atención Plena (Mindfulness) y Aceptación para el Mantenimiento del Cambio en Problemas de Adicción a Sustancias Psicoactivas.* (Tesis de Maestría). Universidad Nacional Autónoma de México. Residencia en Psicología de las Adicciones. 2014.

[11] R Core Team (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* URL http://www.R-project.org/.

[12] Torres Menárguez, Ana. *No hace falta ser matemático para ser experto en 'big data'.* El País. Madrid, España. 16 de Febrero de 2015.