



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

---

---

FACULTAD DE CIENCIAS

Una introducción a la estadística clínica

T E S I S

QUE PARA OBTENER EL TÍTULO DE:  
ACTUARIA

PRESENTA:  
ALMA ROSA BUSTAMANTE GARCÍA



DIRECTOR DE TESIS:  
MATEMÁTICA MARGARITA ELVIRA CHÁVEZ CANO



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

1. Datos del alumno.

Bustamante

García

Alma Rosa

51 26 75 97

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

303144287

2. Datos del tutor.

Mat.

Chávez

Cano

Margarita Elvira

3. Datos del sinodal 1.

Dra.

Fuentes

García

Ruth Selene

4. Datos del sinodal 2.

Act.

Vázquez

Alamilla

Jaime

5. Datos del sinodal 3.

M. en C.

Alonso

Reyes

María del Pilar

6. Datos del sinodal 4.

M. en C.

Flores

Díaz

José Antonio

7. Datos del trabajo escrito.

Una introducción a la estadística clínica

107 p.  
2011

8. Palabras clave.  
estadística clínica  
ensayos clínicos  
tratamiento  
prueba de no efecto

# Introducción

En la actualidad, nuestra vida cotidiana está completamente influenciada por los avances tecnológicos y científicos que surgen día con día, es difícil ignorar las ventajas que estos han aportado a nuestro desarrollo y a la mejora de las condiciones de vida en todos los aspectos. Pero esta condición tan favorable que ha alcanzado la ciencia para la humanidad no sería posible si científicos e investigadores no hubieran notado que es casi imposible que existan ciencias independientes que no necesiten nada de ninguna otra. Lo interdisciplinario de la Ciencia ha sido el punto clave para lograr el éxito en la investigación científica en muchos ámbitos, ejemplo de ello es la combinación de los conocimientos en Ciencias como Medicina y Matemáticas que ha significado grandes logros para acrecentar la calidad de vida de las personas.

La Medicina es una ciencia llena de probabilidades y de manejo de la incertidumbre, por eso ha sido indispensable implementar en la investigación clínico-científica Ciencias Matemáticas, como la Probabilidad y la Estadística.

La investigación debería entenderse como un proceso encaminado a responder una pregunta o resolver un problema, y si se habla de investigación clínica se debe entender como el proceso dedicado a resolver problemas relacionados completamente con la salud y el bienestar físico y psicológico de las personas. De la necesidad de adecuar los procedimientos de la Probabilidad y la Estadística a la investigación clínica surge la Estadística Clínica.

Es importante tomar en cuenta que al realizar estudios o experimentos clínicos no basta con disponer de tiempo, recursos y entusiasmo dirigidos a responder una pregunta, sino que además es necesario contar con conocimiento metodológico, estadístico y clínico.

La Estadística Clínica proporciona las herramientas e instrumentos metodológicos adecuados para la correcta planeación, realización y obtención de resultados de los estudios clínicos; y este trabajo tiene como objetivo principal introducir al lector en los principios básicos de diseño, análisis de datos e interpretación de información y resultados, para que con base en ellos sea capaz de realizar ensayos clínicos y obtener resultados veraces y útiles del objeto de estudio sin necesidad de adaptar a sus estudios clínicos los modelos de otros estudios ya realizados, además de desarrollar la capacidad de discernir entre el tipo de modelos que es conveniente usar y los que no lo son.

Se trata de abordar conjuntamente conceptos relacionados con la metodología, la estadística y clínicos ya que todos ellos son necesarios para interpretar y realizar un estudio. En consecuencia, en el primer capítulo de este trabajo se da una breve reseña de como ha evolucionado la investigación clínica y los ensayos clínicos a través del tiempo, además de proporcionar todos los conceptos clínicos básicos necesarios para la comprensión de los temas posteriores, sin olvidar tratar un poco, como se organizan y

realizan dichos ensayos.

En el segundo capítulo se analiza la forma de proceder de la estadística, basada en las herramientas y metodologías que proporcionan las bases necesarias para la inferencia, con el propósito de destacar la importancia de la aleatorización. Este capítulo es importante, ya que, después de haber introducido al lector en los conceptos clínicos y estadísticos necesarios, en él se ligan dichos conceptos para mostrar como la combinación hace que se pueda realizar investigación clínica veraz y confiable con la que se puede determinar la efectividad de un nuevo tratamiento o su fracaso. Adicionalmente, se revisan algunos tipos de pruebas que ayudan a saber si los tratamientos o terapias probados tienen al menos un efecto.

Por último, el tercer capítulo aborda la implementación del Análisis longitudinal en la investigación clínica describiendo algunos de los modelos relacionados más relevantes y mostrando la forma de utilizarlos por medio de aplicaciones sencillas pero claras. También se trata el tema de como debe procederse cuando el conjunto de datos está incompleto (datos faltantes), para hacer consciencia de que no se les puede dar el mismo tratamiento que ha conjuntos de datos completos.

Cabe mencionar que aunque el Análisis de Supervivencia es un tipo de análisis al que comúnmente se recurre ya que es de gran utilidad para muchos tipos de ensayos clínicos y además es de gran importancia, este trabajo no lo aborda debido a que hay mucha literatura disponible y de calidad al respecto. Es recomendable, para el lector que lo requiera, el estudio de este tipo de análisis.

# Índice general

<b>1. Introducción a los ensayos clínicos</b>	<b>1</b>
1.1. Desarrollo de los ensayos clínicos . . . . .	1
1.1.1. Las primeras investigaciones y ensayos clínicos . . . . .	2
1.1.2. Formalización de los Ensayos Clínicos . . . . .	3
1.2. Conceptos básicos . . . . .	6
1.3. Fases de los ensayos clínicos . . . . .	10
1.4. Participantes, Ética y Protocolo . . . . .	11
1.4.1. Participantes . . . . .	11
1.4.2. Ética . . . . .	12
1.4.3. Protocolo . . . . .	13
<b>2. Implementación de la Estadística Clínica</b>	<b>15</b>
2.1. El Método Científico y los Ensayos Clínicos Controlados . . . . .	15
2.2. Sir Ronald A. Fisher y la aleatorización . . . . .	17
2.2.1. La Dama Catadora de Té . . . . .	18
2.3. Diseño del Ensayo Clínico . . . . .	20
2.3.1. Diseño paralelo . . . . .	20
2.3.2. Diseño de tratamiento sucesivo . . . . .	21
2.3.3. Diseños alternativos . . . . .	23
2.4. Determinación del tamaño de la muestra y de la longitud del Ensayo . . . . .	24

2.4.1.	Monitoreo intermedio de los datos . . . . .	30
2.4.2.	Prueba Clásica Secuencial de Grupo . . . . .	31
2.5.	Unidades y Vector de Asignación de tratamientos . . . . .	35
2.6.	Métodos de asignación aleatoria de tratamientos . . . . .	38
2.7.	Prueba de Hipótesis de No Efecto del Tratamiento . . . . .	41
2.7.1.	La distribución de una Estadística de Prueba cuando el Tratamiento No tiene Efecto . . . . .	41
2.7.2.	Pruebas de Aleatorización comunes . . . . .	45
2.7.3.	Clases de Estadísticas de Prueba . . . . .	50
2.8.	Modelos para Efecto del Tratamiento . . . . .	55
2.8.1.	Respuestas cuando el tratamiento tiene un efecto . . . . .	55
2.8.2.	Modelo de No Interferencia entre unidades . . . . .	56
2.8.3.	Modelo de un Efecto Aditivo y modelos relacionados . . . . .	57
2.8.4.	Efectos positivos y efectos mayores . . . . .	57
2.9.	Intervalos de Confianza . . . . .	58
2.9.1.	Pruebas de Hipótesis Generales . . . . .	58
2.9.2.	Intervalos de Confianza . . . . .	60
2.10.	Estimación puntual . . . . .	61
2.10.1.	Ausencia de sesgos en estimaciones del efecto promedio . . . . .	61
2.10.2.	Estimador Hodges-Lehmann de un efecto aditivo . . . . .	62
<b>3.</b>	<b>Introducción al análisis de datos longitudinales</b>	<b>65</b>
3.1.	Definiciones básicas . . . . .	65
3.2.	Presentación gráfica . . . . .	66
3.3.	Algunos modelos para el análisis de datos longitudinales . . . . .	70
3.3.1.	Modelo con intercepto aleatorio . . . . .	71
3.3.2.	Modelo con intercepto y pendiente aleatorios . . . . .	79



3.3.3. Modelo con Estructura Covarianza Espacial de Potencia para el Error . . . . .	86
3.3.4. Modelo de regresión logística con intercepto aleatorio . . . . .	91
3.4. Manejo de Observaciones Faltantes . . . . .	96
3.4.1. Mecanismo para Observaciones Faltantes . . . . .	96
3.4.2. Distinción entre faltantes completamente al azar, al azar y no debido al azar . . . . .	97
3.4.3. Elección del modelo de efectos combinados adecuado para el conjunto de datos desbalanceados . . . . .	98
<b>1. Título Quinto de la Ley General de Salud: Materia de Investigación</b>	<b>110</b>
<b>2. Distribución Normal Estándar</b>	<b>113</b>
<b>3. Códigos SAS para los ejemplos del Capítulo 3</b>	<b>114</b>

# Capítulo 1

## Introducción a los ensayos clínicos

Abordar temas referentes a ensayos clínicos y a Estadística clínica, implica dar una definición breve pero útil de cada uno de estos conceptos. En atención a ello, se define a los *ensayos clínicos* como investigaciones científicas que se hacen para estudiar los riesgos y beneficios de nuevas terapias y tratamientos propuestos para su uso en humanos. Y la *Estadística clínica* es la Estadística matemática utilizada para que se lleve a cabo el diseño de estos ensayos, y principalmente, para ayudar a evaluar la efectividad de las terapias o tratamientos bajo estudio.

Para comprender la utilidad y la importancia de la Estadística Clínica en la implementación de Ensayos Clínicos es esencial familiarizarse con varios conceptos básicos que para muchos podrían ser nuevos. Por lo tanto, se hace una breve introducción a dichos conceptos para, posteriormente, situarse de manera sencilla en el objeto de estudio. También es importante, aún cuando sea de manera somera, conocer la forma en que se fueron desarrollando este tipo de experimentos o ensayos a lo largo de la historia, pues es así como puede crearse consciencia de la importancia de su implementación dentro de la investigación científica clínica. Considerando lo anterior, este primer capítulo hace alusión a cada uno de los aspectos ya mencionados.

Primero se hace una sucinta reseña del desarrollo de este tipo de investigación que es la clínica a través de los ensayos clínicos y su implementación a través de la historia. Seguido de esto, se proporciona una definición más estricta de lo que son dichos ensayos, cómo se desarrollan y conducen en la actualidad y cuál es su utilidad.

### 1.1. Desarrollo de los ensayos clínicos

Remontarse a los inicios de los ensayos clínicos conllevaría escribir un libro completo ya que, si se considera, que en los inicios del siglo XX ya se hacían varias pruebas de

tratamientos en humanos, se tendría que, al menos, mencionar cada una de estas pruebas, pues aunque no cumplieran con los estándares científicos actualmente establecidos, aportaban evidencia científica muy importante.<sup>1</sup>

Incluso, puede mencionarse el caso de un físico y filósofo árabe llamado Avicenna, quien vivió del año 980 al 1037. En su enciclopedia Canon de Medicina, Avicenna documentó una serie de reglas que debían seguirse para evaluar el efecto de los medicamentos sobre las enfermedades. Durante mucho tiempo sus propuestas fueron ignoradas o simplemente desconocidas por muchos, y fue hasta hace poco años que se reconoció que los experimentos planeados que él realizaba estaban tan bien diseñados que se podría afirmar que condujo ensayos clínicos, pues seguía los principios científicos elementales y, se puede decir que, se valía de las únicas bases fiables para evaluar la eficacia y seguridad de nuevos tratamientos.

### 1.1.1. Las primeras investigaciones y ensayos clínicos

Uno de los casos más famosos en la investigación clínica es el caso de James Lind, quien, en 1753, hizo un ensayo comparativo para encontrar el mejor tratamiento contra el escorbuto<sup>2</sup>. Este caso es famoso porque es un ejemplo de las investigaciones clínicas que se anticipan a las metodologías actuales y, sin embargo, proveyeron evidencia científica relevante. Lind designó a doce de sus pacientes a un experimento y con ellos formó 6 parejas, a cada pareja le asignó un medicamento o remedio distinto entre los siguientes: ingerir un cuarto de sidra al día, tomar veinticinco gotas de elixir de vitriolo diariamente, tomar dos cucharadas de vinagre al día, someterse a un tratamiento de agua de mar, comer nuez moscada o comer dos naranjas y un limón al día; y con dicho experimento concluyó que el mejor remedio consistía en comer naranjas y limones. Pero no todos los experimentadores han tenido la misma apreciación en cuanto a la metodología científica, ejemplo de ello es el doctor y farmacéutico Benjamín Rush que en 1794, afirmaba que la sangría era el mejor remedio para la fiebre amarilla que se manifestaba en Filadelfia en ese entonces. Rush sólo se basó en el resultado que este remedio había tenido en un sujeto, y considerando eso, puede comprenderse que pudieron existir muchos otros factores que quizá influyeron en la mejoría del sujeto y no precisamente pudo deberse a la sangría. Afortunadamente, tiempo después se descubrió que su afirmación era errónea.

A lo largo de la historia, hay una larga lista de experimentos exitosos y una más de fracasos experimentales pero que de alguna forma han servido y sirvieron para ir

---

<sup>1</sup>Aquí se mencionan varios casos, sin excluir la posibilidad de encontrar muchos casos más no menos importantes.

<sup>2</sup>Enfermedad ocasionada principalmente por la falta o deficiencia de vitamina C y por malas condiciones higiénicas; se caracteriza por un estado general de debilidad, dolores en las articulaciones, anemia, hinchazón de las encías y hemorragias múltiples.

logrando una forma de experimentación más segura y eficaz.

Es en 1834 que Pierre Louis da bases claras para el uso de métodos numéricos en la medicina, estableciendo que se debe contar con cuatro herramientas importantes para la experimentación, las cuales son:

1. Una observación minuciosa y exacta de los resultados en el paciente
2. Conocimiento del progreso natural de la enfermedad cuando no es tratada
3. Definición precisa de la enfermedad antes del tratamiento
4. Observación cuidadosa sobre desviaciones que se presenten al implementar el tratamiento

Después de asentar estas bases, comienzan a encontrarse una serie de experimentos, que al realizarse contribuían con nuevos métodos, procedimientos y metodologías<sup>3</sup> a los ensayos entonces vigentes.

Una herramienta adicional en el desarrollo de los ensayos clínicos fue la creación de anestias generales durante el siglo XIX, lo cual hizo posible que se realizaran pruebas de mortalidad durante las cirugías, en la cuales, la implementación de dichas anestias reflejó que había mayor supervivencia que en ausencia de ellas.

Al reunir todos estos medios, los experimentadores encontraron nuevas formas de probar sus tratamientos y/o estudios. En el Cuadro 1.1 se mencionan algunos de los experimentos más relevantes entre los años 1865 y 1950.

### **1.1.2. Formalización de los Ensayos Clínicos**

Después de la Segunda Guerra Mundial, comenzó a evolucionar la implementación y caracterización de los ensayos clínicos en la evaluación de nuevos medicamentos. Es así como se logra, que en 1950, se desarrollen los ensayos clínicos de manera más metodológica gracias a Sir Austin Bradford Hill quien influye en dicho desarrollo. Hill escribe varios artículos en los que describe cómo se deben conducir los ensayos clínicos y presenta conceptos fundamentales como son la definición de ensayo, control concurrente, asignación aleatoria, pacientes elegibles, tratamiento programado, objetivo de la evaluación y análisis estadístico.

---

<sup>3</sup>Método, es el significado general de modelo lógico que se sigue en la investigación científica. En cuanto al de Metodología, lo más adecuado es considerarla como el estudio y análisis, por medio de una serie de pasos a través de los cuales se articula el método con la realidad para resolver un problema, reservando los términos técnica y procedimiento para hacer alusión a los aspectos más específicos y concretos del método que se usan en cada investigación.

<b>Año</b>	<b>Autor o Sociedad de investigación</b>	<b>Experimento</b>
1865	Sutton	Usó agua de menta para mostrar la inmensa variación natural en el proceso de la enfermedad y la tendencia a la curación natural en algunos otros casos (podría considerarse el primer uso de un placebo)
1870	Lister	Comparación de la mortalidad en cirugías de amputación
1898	Fibiger	Ensayo comparativo de pacientes a quienes se les hace una asignación alternativa y usan el suero para la difteria y los que no se controlan
1915	Greenwood y Yule	Ensayo comparable de tratamientos para el cólera, al parecer con una asignación aleatoria
1927	Ferguson	En estudios de vacunas para el resfriado común, podría considerarse el primero en usar un ensayo ciego
1937	Colebrook y Purdie	Mostraron reducción de la mortalidad por fiebre puerperal al implementar sulfonamidas
1937	League of Nations Malaria Commission	Ensayo que describe prevención, tratamiento y conocimiento de la malaria
1938	Evans y Gaisford	Ensayo comparativo de sulfonamidas con una terapia no especificada contra la neumonía lobar
1948	Medical Research Council	El primer ensayo clínico con una apropiada aleatorización y un grupo control en el tratamiento de la tuberculosis pulmonar
1949	Harrison	Ensayo con asignación aleatoria logra demostrar la eficiencia de la penicilina (descubierta antes)
1950	Medical Research Council	Primer ensayo que usa un placebo control y se realiza como un ensayo doble ciego

Cuadro 1.1: Breve cuadro de la realización de ensayos importantes entre 1865 y 1950

Eventos como eran la escasez de registros, la falta de medios para la investigación, la falta de remedios activos y la automedicación eran situaciones que impedían el establecimiento de ensayos clínicos sobre bases científicas. Un factor importante para el desarrollo de los ensayos clínicos fue que cuando se implementaban los tratamientos, se evitaba la transmisión de las enfermedades que eran fácilmente contagiadas, cosa que era de gran interés para las sociedades y gobiernos, y por la cual, comenzaron a verse los ensayos como algo aceptable.

El progreso de los ensayos clínicos, a partir de 1950 hasta la fecha, ha sido de calidad científica consistente y prueba de ello son los siguientes avances:

- Ensayo clínico sobre el virus de la polio Salk
- Ensayos clínicos sobre las quimioterapias para el tratamiento de cáncer en Estados Unidos
- Ensayos clínicos para el tratamiento de los infartos agudos al miocardio
- Ensayos clínicos dentro de la industria farmacéutica

El cuadro 1.2 enlista algunos de los logros más importantes dentro de los ensayos clínicos, después de la formalización de los ensayos clínicos en 1950.

<b>Año</b>	<b>Científico o Soc. de Investigación</b>	<b>Experimento</b>
1962	Hill	Publicación de un libro de ensayos clínicos
1966	-	Publicación del Servicio de Salud Pública de Estados Unidos
1967	Chalmers	Estructura para separar los tratamientos monitoreados y los tratamientos de administración procesada
1977	Medical Research Council Working	Ensayo de hipertensión leve comparando bendrofluacida, propranolol y un placebo
1979	-	Establecimiento de la Sociedad para Ensayos Clínicos
1980	-	Primer caso de ensayos Clínicos Controlados, ensayo para probar el anturan para tratar el post-infarto
1982	-	Estudio de la meningitis bacterial con ensayo aleatorizado

Cuadro 1.2: Logros importantes en el campo de los ensayos clínicos después de 1950

## 1.2. Conceptos básicos

Existen varias definiciones de lo que son los ensayos clínicos, pero la mayoría de ellas coinciden en sugerir que los **ensayos clínicos** son experimentos planeados, diseñados para evaluar el tratamiento más apropiado para futuros pacientes con una condición médica en particular. Por ello, se menciona al inicio de este capítulo, que son investigaciones que se hacen para estudiar los riesgos y beneficios de nuevas *terapias y tratamientos*<sup>4</sup> propuestos para su uso en humanos.

En los ensayos clínicos, se *comparan* los resultados obtenidos en un grupo de pacientes tratado con el tratamiento de prueba (nuevo medicamento, terapia, etc.) con los resultados de otro grupo de pacientes que reciben un tratamiento control (tratamiento ya vigente, tratamiento estándar, placebo, etc.). Los pacientes de ambos grupos se inscriben al ensayo, son tratados y monitoreados durante el mismo periodo de tiempo hasta determinado momento.

Los ensayos clínicos son experimentos *controlados*, ya que, en éstos, la asignación de tratamientos a los sujetos es controlada por el experimentador. Dicha asignación de los pacientes a cada grupo es por aleatorización o por algún otro método y las mediciones que se hagan como resultado de los tratamientos deben ser el resultado de una prueba de laboratorio, de una evaluación de calidad de vida, de la calificación asignada a alguna característica específica o, en algunos casos extremos, de la muerte o supervivencia de un paciente.

Como consecuencia de esta definición un tanto limitada, los estudios comparativos donde participan animales, o estudios que se llevan a cabo *in vitro*, utilizando sustancias biológicas del hombre no califican como ensayos clínicos. La definición también excluye las investigaciones que involucran controles históricos.

Una característica esencial de los ensayos clínicos es que se usan resultados basados en una *muestra limitada* de pacientes para hacer inferencias acerca de cómo el tratamiento debería comportarse en la *población* de pacientes que requerirán el tratamiento en el futuro. Es aquí donde radica su importancia.

Pero, *¿cómo se conduce un ensayo clínico?* Primero, los médicos en los centros de atención primaria pueden recomendar a los pacientes inscribirse en ensayos clínicos adecuados. Antes de iniciar con un ensayo, los pacientes candidatos que han sido calificados pasan por una etapa conocida como Etapa de consentimiento informado, en el cual se les dan a conocer sus derechos y los riesgos y beneficios que tiene el ensayo. Después de esta etapa, si los candidatos siguen dispuestos a participar, firman un formato de consentimiento informado y están listos para participar en el ensayo. A los individuos

---

<sup>4</sup>Aquí se considera tratamiento a cualquier medicamento, cirugía, ejercicio, dieta, etc. o combinación de ellos que sea considerado una forma de tratar una enfermedad o deficiencia física.

que participan en un ensayo se les denomina *sujetos*.

Para comenzar el ensayo, todos los sujetos son sometidos a un tratamiento inicial (puede ser ingerir alguna píldora, someterse a una cirugía para algún implante, etc. pero esto lo determina el propósito del ensayo). Después de esto, los sujetos realizan varias visitas controladas para dar seguimiento al ensayo (la frecuencia de las visitas está predeterminada). Durante estas visitas, se revisa su salud y se toman y registran las mediciones necesarias. Según el ensayo y el propósito de éste, puede ser o no necesario que durante el periodo de seguimiento el sujeto siga realizando determinadas actividades como parte del tratamiento.

Al final de cada visita de seguimiento programada, cada sujeto tiene la opción de continuar en el estudio o desertar de él. Un sujeto que desea permanecer en el estudio llena una adición al formulario de consentimiento informado, un documento que confirma la voluntad del sujeto para continuar en el ensayo.

Dependiendo de lo larga que sea la participación de los sujetos dentro del ensayo, estos se pueden clasificar en cuatro categorías:

- Sujetos para quienes ocurren eventos adversos y les es necesario dejar el ensayo, donde un evento adverso es definido como contraer una enfermedad, desarrollar una cierta condición de salud o morir, dependiendo de la especialización del ensayo.
- Sujetos que abandonan o desertan del ensayo clínico antes de la última visita de seguimiento a causa de circunstancias ajenas a la salud, son denominados desertores. Por ejemplo, un sujeto que necesita mudarse de ciudad.
- Sujetos que voluntariamente abandonan el ensayo después de su última visita de seguimiento.
- Sujetos que permanecen inscritos en el ensayo hasta que éste finaliza.

Típicamente, un ensayo clínico es detenido cuando un número predeterminado de sujetos han sido *devengados*, es decir, han cumplido cierta condición. Por ejemplo, un estudio puede finalizar cuando se han seguido 250 sujetos durante todo un año.

Ocasionalmente, un ensayo es terminado antes. Por ejemplo, cuando los datos registrados ya representan un riguroso soporte de la eficacia de la terapia o tratamiento probado. Más adelante, se trata este punto con mayor detalle en lo referente a Monitoreos Intermedios.

Los ensayos clínicos tienen ciertas características esenciales:

- Son *prospectivos*, pues el registro de los datos está orientado al futuro, esto es, los participantes son seguidos desde puntos bien definidos en el tiempo llamados



tiempo cero o línea de base. Las líneas de base no siempre serán las mismas para todos los sujetos porque en la mayoría de ocasiones ellos pueden entrar al estudio en distintos tiempos.

- Comúnmente, los ensayos clínicos son *aleatorizados*, pero existen los *no aleatorizados*. Un ensayo aleatorizado es usado para comparar la eficacia de dos o más tratamientos a prueba, asignando uno distinto a cada grupo. Para poder validar un análisis estadístico cuando se comparan resultados, debería ser igualmente probable que un sujeto sea asignado a uno u otro grupo, es por eso que se usa la aleatorización.

En un ensayo no aleatorizado, todos los sujetos reciben un tratamiento experimental y, por tanto, no hay asignación al azar entre los distintos grupos. Para el análisis estadístico, el grupo a prueba es comparado con el grupo control histórico -que es, el de los sujetos tratados en el pasado con un tratamiento disponible.

Estadísticamente, se prefieren los ensayos clínicos aleatorizados debido a que la aleatorización es la que permite la *comparabilidad de los grupos*, lo que los convierte en el tipo de ensayo clínico por excelencia.

- Un ensayo clínico es un ensayo o experimento *controlado*, pues permite el control de las condiciones bajo estudio. Un ensayo clínico se inicia cuando surge una hipótesis a partir de estudios *no controlados*, que son observacionales, descriptivos o retrospectivos, o de estudios preclínicos. Frecuentemente se descubren en investigaciones preclínicas, con estudios observacionales, posibilidades terapéuticas que no tienen ningún beneficio al ser probadas bajo un ensayo clínico.

En la mayoría de los casos existe un grupo control (que recibe un tratamiento actual o un placebo) y un grupo de tratamiento (que recibe el tratamiento a probar). Un placebo es un tratamiento que es administrado en forma de medicamento (píldora, solución, inyección, etc.) pero no tiene un ingrediente medicinal activo. No hay tal cosa como un placebo para el grupo control en los ensayos que prueban nuevos dispositivos biológicos o prótesis. En su lugar, se implantan dispositivos que se utilizan ampliamente. Para este tipo de ensayo, un ensayo clínico no aleatorizado puede ser una mejor opción. Un ensayo clínico controlado permite el control de las condiciones bajo estudio.

La existencia de un grupo control permite la *comparabilidad de efectos*.

El elemento esencial del ensayo es la existencia de un grupo de comparación o **grupo control**, que permite probar si la nueva intervención (por ejemplo un nuevo fármaco) es mejor o no que las ya existentes o que no intervenir (placebo).

La existencia de un grupo control como tal, depende del diseño del ensayo (lo cual a su vez tiene dependencia del propósito del ensayo), pero cuando no existe un grupo control como el que se definió anteriormente (como podría suceder en el caso del ensayo de los nuevos dispositivos biológicos), existen grupos análogos

que hacen esta función de grupos comparables. Más adelante, cuando se tratan los tipos de diseño, se profundiza en cómo son estos grupos.

- El *cegamiento*, que es una condición impuesta sobre el ensayo para mantener en secreto el conocimiento del tratamiento asignado a cada sujeto, el curso del tratamiento u observaciones previas. Los procedimientos usualmente cegados son el tratamiento asignado o la evaluación del estatus de los sujetos de estudio. El cegamiento previene determinados sesgos en las diversas etapas del ensayo clínico y protege la secuencia después de la asignación al grupo de tratamiento.

De esta forma existen ensayos No ciego (ensayos en los que investigadores y participantes conocen el tratamiento asignado) y los ensayos Ciego (indican que el conocimiento del tratamiento asignado se oculta a los participantes, a los investigadores del ensayo y/o a los patrocinadores) y de estos últimos encontramos:

- Ciego simple. Una de las tres categorías de individuos, normalmente participantes, desconoce el tratamiento asignado a lo largo del desarrollo del estudio
- Ciego doble. Participantes, investigadores y patrocinadores, desconocen el tratamiento asignado
- Ciego triple. Se refiere a un ensayo doble ciego que mantiene blindado el análisis de los datos, hasta una etapa determinada del estudio

El cegamiento permite la *comparabilidad de la información*.

- Los ensayos clínicos pueden incluir sólo un centro clínico (*un ensayo de un solo centro*) o múltiples centros (*ensayos multicentro*). En un ensayo clínico multicentro, hay un centro coordinador y múltiples centros participantes. El centro de coordinación funciona como un centro de investigación donde se diseña el ensayo, se realiza el análisis estadístico de los datos coleccionados y se interpretan los resultados. Un centro de investigación participante es un centro médico donde los sujetos son reclutados, inicialmente tratados y admitidos para las visitas de seguimiento. Los ensayos multicentro son más difíciles de coordinar, pero estos reclutan sujetos más rápido y los resultados pueden ser generalizados a una población más grande porque los grupos de población son más amplios.

Después de haber descrito las características esenciales de los ensayos clínicos, se hace la siguiente aclaración: los estudios observacionales y los experimentos o ensayos controlados persiguen el mismo objetivo, inferir sobre los efectos de algún tratamiento, pero el control de las condiciones y la asignación aleatoria están presentes únicamente en los experimentos. A eso se debe que el tipo de ensayo clásico es el ensayo clínico aleatorizado controlado, por lo tanto, en adelante el desarrollo se hace con base en ellos.

### 1.3. Fases de los ensayos clínicos

Después de probar bajo condiciones de laboratorio y en animales, un nuevo producto (un nuevo medicamento o tratamiento), éste se prueba en humanos. Los ensayos que involucran humanos- los llamados ensayos clínicos- son divididos en cuatro fases:

- En la Fase I de los ensayos (o fase piloto), el nuevo producto es probado sobre un pequeño grupo de sujetos (normalmente de 20 a 80 personas). Usualmente las pruebas son hechas con grupos de personas que son voluntarios o, dependiendo de la especificación del ensayo, con voluntarios en estado avanzado de enfermedad. Esta fase es no ciego. Los voluntarios participantes están conscientes de la naturaleza y del propósito del producto que ellos están probando. Durante esta fase piloto, se evalúa la seguridad del producto, se determina una dosis óptima y se identifican los efectos secundarios provocados con el tratamiento.
- En la Fase II del ensayo, comienza la investigación clínica inicial. El producto es probado en un grupo de personas más grande (de 100 a 200 personas) para determinar si es efectivo y evaluar la tasa de eventos adversos. Los objetivos principales de esta Fase son:
  1. Identificar exactamente la población que puede ser beneficiada con el tratamiento.
  2. Verificar y estimar la efectividad de la dosis determinada en la Fase I.

Esta fase, como la Fase I, es abierta a los voluntarios y es no ciego.

- En la Fase III de los ensayos, da lugar la *investigación científica clínica* extensiva de los productos. La prueba se lleva a cabo sobre un grupo de personas mucho más grande para comparar el nuevo producto con un placebo o un tratamiento estándar, y para confirmar su eficacia y monitorear los efectos secundarios, o en su defecto, para confirmar que no es tan bueno o mejor que los ya existentes. Para algunas personas el término 'ensayo clínico' es sinónimo de la Fase III, pues es la fase más rigurosa y donde se lleva a cabo una amplia investigación clínica científica del nuevo tratamiento. Después de que se completa esta fase, los resultados son reportados a las instituciones competentes en cada nación y, sólo si éstos son benéficos, se solicita la aprobación para su comercialización.
- Después de que el producto es aprobado para su comercialización, los ensayos clínicos entran a la Fase IV- la última fase en la cual la post-comercialización toma lugar. En la Fase IV, se lleva a cabo una prueba en la población general después de que el producto es comercializado, esto es para coleccionar información adicional sobre la seguridad del producto y la eficacia sobre un extenso periodo de tiempo.

## 1.4. Participantes, Ética y Protocolo

### 1.4.1. Participantes

Para la correcta conducción de los ensayos clínicos es imprescindible la colaboración de los siguientes participantes:

- **Patrocinador.** El patrocinador de un ensayo clínico es un individuo, organización o compañía que inicia y financia el ensayo clínico, ya que tiene determinado interés en los resultados que se obtendrán de él. Por lo tanto, se encargará de hacer una buena selección de los investigadores (instituto o asociación) que lo ejecutarán, procurando que éstos cumplan con las regulaciones y hagan un análisis seguro y eficaz de los datos obtenidos con el producto probado.
- **Instituto o Asociación de Investigación Clínica.** Es el sitio de control del experimento y que trabaja a favor del patrocinador. Supervisa el ensayo y se asegura de que la investigación cumpla con todos los requerimientos necesarios y de que los investigadores y todo el personal estén bien entrenados y calificados, además de procurar que todos los materiales que se requieren para conducir el ensayo estén disponibles.
- **Sujetos.** Son los pacientes participantes dispuestos a someterse al experimento. Son los participantes más importantes dentro del ensayo ya que todo el experimento se basará en los resultados que serán observados en ellos y deberán ser suficientes en cantidad para poder generalizar los resultados que se obtengan a toda la población.

Cada uno de los participantes mencionados juega un papel muy importante en el ensayo clínico y es responsabilidad de cada uno de ellos que su participación sea lo más ética y apropiada posible para lograr resultados confiables y benéficos para la sociedad.

Cada participante tiene determinadas obligaciones dentro del experimento y son responsables de la forma en que se ejecuta. A continuación se especifica cuales son los compromisos de cada participante involucrado:

- **Instituto o Asociación de Investigación Clínica:** Este instituto o asociación es responsable de nombrar un investigador principal y asignarle un equipo de trabajo. El investigador principal es el encargado de vigilar el cumplimiento de los lineamientos de las Buenas Prácticas Clínicas (BPC, que se definen más adelante), del respeto a las leyes locales y del apego al protocolo. Debe organizar a su equipo de trabajo para que le ayude a cumplir con las expectativas en cuanto a número de pacientes reclutados, tiempos para hacerlo y calidad de los datos generados.

- **Patrocinador:** El patrocinador es responsable de implementar y mantener sistemas para asegurar la calidad y control de calidad con procedimientos estándar de operación escritos para asegurar que los estudios sean conducidos y los datos sean generados, documentados (registrados) y reportados en cumplimiento con el protocolo, las BPC y los requerimientos regulatorios que apliquen. Todos los acuerdos establecidos por el patrocinador con el investigador/institución y con cualquier otra parte involucrada en el estudio clínico deberán ser por escrito. El programa de control de calidad implementado por el patrocinador debe incluir actividades tanto en el ámbito interno como externo, asegurando que se encuentren correctamente documentados todos los procedimientos que tienen que ver con el estudio. Dicho programa debe incluir visitas regulares de monitoreo a los centros de investigación, detección oportuna de errores en la conducción del estudio y corrección inmediata de los mismos, con estrategias que aseguren su prevención en el futuro, incluyendo entrenamiento y reentrenamiento en caso necesario.
- **Sujetos:** Su responsabilidad es cumplir con el mayor apego posible las instrucciones de los investigadores, tratar de asistir a todas las visitas de seguimiento y procurar, en la medida de lo posible, no abandonar el experimento. Sus aseveraciones y declaraciones acerca de cómo se sienten y de su estado de salud deben ser objetivas, francas y exactas.

### 1.4.2. Ética

La protección de los derechos y el bienestar de los pacientes que participan en investigaciones científicas constituyen el propósito actual de la ética de la investigación. Lamentablemente, se han señalado investigaciones que no han considerado a los participantes como personas, sino como objetos de estudio. Por esta razón, la Declaración de Helsinki, en 1964, estableció uno de los primeros antecedentes en principios éticos de investigaciones en seres humanos.

Universalmente, se acepta que no debe hacerse investigación médica en personas sin seguir los postulados y lineamientos nacional e internacionalmente aprobados.

En México, la Secretaría de Educación, la Secretaría de Salud y El Consejo Nacional de Ciencia y Tecnología trabajan conjuntamente para orientar el desarrollo de la investigación científica y tecnológica destinada a la salud, valiéndose de las leyes y reglamentos emitidos para este propósito. La Ley General de Salud, busca el seguimiento de los postulados internacionalmente válidos y cuenta, en su Título Quinto, con una resolución en Materia de Investigación<sup>5</sup>, la cual está realizada con apego a la Declaración de Helsinki. Se establece que toda institución de salud en donde se realice investigación

---

<sup>5</sup>Anexo I. Título quinto de La Ley General de Salud: Materia de Investigación (México)

para la salud deberá tener un Comité de Ética en el caso de que realice investigación en seres humanos.

Las Buenas Prácticas Clínicas (BPC), en inglés, *Good Clinical Practice*, es un estándar internacional para el diseño, conducción, desarrollo, monitoreo, auditoría, registro, análisis y reporte de estudios de investigación clínica en los que participan seres humanos como sujetos de estudio. El cumplimiento de las BPC asegura dos aspectos fundamentales en los estudios en los que se pone a prueba un medicamento o un equipo de diagnóstico previo a su comercialización:

- a) la protección de los derechos, la seguridad, y el bienestar de los sujetos de estudio, y
- b) la credibilidad de los datos generados, necesaria para el registro de una innovación terapéutica en cualquier parte del mundo.

Los lineamientos de las BPC son el estándar que se sigue durante el desarrollo de proyectos de investigación clínica. A este respecto, se encuentran claramente especificadas las responsabilidades y funciones del investigador y del patrocinador, los contenidos básicos del protocolo de investigación del ensayo clínico, el manual del investigador y los documentos esenciales para la conducción de un estudio clínico. Aquí se cubren todos los aspectos de preparación, monitoreo, reporte y archivo de estudios clínicos e incluye los lineamientos de las BPC de la Unión Europea, Japón, EUA, Australia, Canadá, los Países Nórdicos, y la OMS (Organización Mundial de la Salud), que deben ponerse en práctica cuando se generan datos clínicos que pretendan someterse a autoridades regulatorias o en cualquier proyecto de investigación clínica que pueda tener un impacto en la seguridad y bienestar de los sujetos participantes.

### 1.4.3. Protocolo

El protocolo de investigación es un documento científico (y legal) que formaliza la propuesta de investigación y sus condiciones. Está formado por una secuencia sistematizada de apartados, abarca toda una serie de información relevante y describe el proceso dinámico a cumplir por los investigadores, o por terceros. Por lo tanto, el protocolo corresponde a la versión escrita del plan propuesto para estudiar e investigar el tema de interés, es la directriz fundamental de trabajo que, de manera ordenada, indica una vía idónea para solución del problema planteado. Dado que la conceptualización del problema reduce a términos concretos y explícitos el asunto de interés, el talento y la experiencia del investigador es de gran importancia al intentar la transformación de una idea en un verdadero proyecto o plan de acción, y para asegurar que el problema es relevante como para merecer su resolución y, con ello, proveer beneficios a la sociedad.

El esquema para presentar un protocolo debe de cumplir con los siguientes elementos:

1. Resumen
2. Índice
3. Información general
4. Introducción y justificación
5. Objetivos
6. Tipo de estudio, diseño y muestra (y lugar)
7. Selección de pacientes - sujetos
8. Intervención (tratamientos)
9. Variables de evaluación
10. Plan de operación
11. Procesamiento y análisis de datos
12. Aspectos éticos
13. Consideraciones prácticas
14. Referencias bibliográficas
15. Anexos y apéndices

# Capítulo 2

## Implementación de la Estadística Clínica

La Estadística Clínica se puede definir sencillamente como la Estadística implementada para el desarrollo de los Ensayos Clínicos, es decir, es todo el conjunto de métodos necesarios para recoger, clasificar, representar y resumir datos (obtenidos en los ensayos clínicos), así como para hacer inferencias (extraer conclusiones) científicas y clínicas a través de ellos<sup>1</sup>.

En este capítulo se describe de forma sencilla como se hace la implementación de estos métodos estadísticos en los ensayos clínicos, y como se aplican en cada etapa del ensayo.

### 2.1. El Método Científico y los Ensayos Clínicos Controlados

Cabe señalar que a partir de ahora sólo es necesario concentrarse en la Fase III de los ensayos clínicos (que es en la cual se lleva a cabo la investigación científica clínica) y en estudiar cual es el razonamiento científico para conducirlos.

Debe tomarse en cuenta que el objetivo principal de la investigación clínica, es el de encontrar tratamientos novedosos o mejoras a los ya existentes en busca de beneficios para los pacientes, y esto sólo se puede lograr si los investigadores con visión e imaginación son capaces de proponer tratamientos y terapias innovadores que parezcan tener oportunidades reales de éxito.

---

<sup>1</sup>Definición ligeramente modificada de la presentada en el libro Bioestadística para Ciencias de la Salud, Martín Andrés, Juan de Dios Luna del Castillo.



Es natural, que si un investigador o un grupo de investigación propone un nuevo tratamiento o terapia, éstos se mostrarán muy entusiastas con los resultados y los estudios pre-clínicos de las Fases I y II pueden ser considerados muy prometedores. Lamentablemente, no todos los nuevos proyectos resultan tan efectivos como era esperado, al pasar por la Fase III, la mayoría de los tratamientos y terapias son truncados pues, esta fase, es una prueba estadística muy rigurosa.

La regla y característica fundamental de esta fase, es que es *comparativa*, pues se hace una comparación entre el tratamiento estándar vigente y el innovador sobre grupos distintos de pacientes; cuando hay ausencia de un tratamiento estándar se hace la comparación con un grupo de pacientes no tratados. Además para tener una evaluación imparcial es necesario que a cada uno de los pacientes se les asigne aleatoriamente el tratamiento estándar o el innovador. De esta forma, los ensayos controlados aleatorizados son los métodos de investigación clínica más generalmente aceptados. Se puede hacer mención de ensayos clínicos sencillos pero trascendentales y que han seguido esta metodología, por ejemplo, el primer ensayo para la meningitis bacterial<sup>2</sup> representa la situación más sencilla donde se ha comparado un nuevo tratamiento (cefuroxina) con un tratamiento estándar (ampicilina y cloranfenicol) para probar si era más efectivo para matar la bacteria.

Otra situación común se ve reflejada en el ensayo clínico realizado para comparar los efectos del anturán como nuevo tratamiento para reducir la mortalidad de los pacientes después de que éstos han sufrido un infarto contra los de un placebo consistente en una tableta que los pacientes no podían distinguir del tratamiento.

Los ensayos clínicos, al ser experimentos, deben cumplir con la forma de proceder del método científico y cumplir con cada uno de los pasos que éste sugiere. Puede observarse en la Figura 1.1 una sugerencia de la forma de proceder al ejecutar un ensayo.

Primero, basándose en una idea inicial acerca del mejoramiento de una terapia o tratamiento, es necesario formular una definición precisa de lo que se le exigirá al ensayo para así poder definir específicamente cada una de las hipótesis a contrastar buscando la eficacia y seguridad del tratamiento. En este paso se especifica exactamente el tipo de paciente, los tratamientos a comparar y el método que será empleado para evaluar la respuesta de cada paciente.

El siguiente paso es diseñar detalladamente un ensayo aleatorio y documentarlo de manera que quede asentado todo en el protocolo elaborado. Este diseño debe cubrir todos los requerimientos científicos, éticos y organizacionales de manera que el ensayo pueda proceder eficazmente de acuerdo al plan. En este paso se deben resolver dos aspectos:

---

<sup>2</sup>Infección de las meninges producida por bacterias; es una enfermedad grave que produce fiebre, cefalea, vómitos, rigidez de nuca, etc.

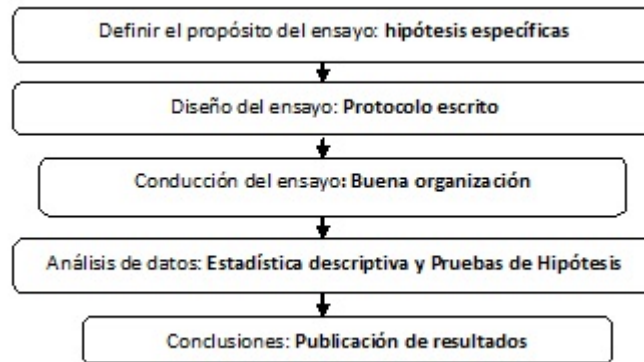


Figura 2.1: Diagrama de la secuencia de los pasos a seguir en un ensayo clínico

- **Tamaño de la muestra:** El ensayo debe contar con el número suficiente de pacientes que le permita dar una estimación precisa y razonable de la respuesta de los pacientes al tratamiento correspondiente.
- **Evasión del sesgo:** La selección, la evaluación y los cuidados auxiliares a los pacientes no deben diferir entre tratamientos, de manera que el ensayo no se vea afectado por factores ajenos a los de los tratamientos.

Al conducir el ensayo, este debe contar con una buena organización que permita obtener los datos necesarios de manera ordenada y correcta. El método estadístico debe ser aplicado a los resultados obtenidos de los pacientes para hacer las pruebas correspondientes a las hipótesis antes especificadas. En particular, se deben usar pruebas de significancia que permitan evaluar que tan fuerte es la evidencia existente para afirmar que hay una genuina diferencia entre tratamientos. Finalmente, es necesario dar las conclusiones observando a fondo los resultados de cada tratamiento, de tal manera que al publicarse puedan ser utilizadas por otros investigadores clínicos.

El objetivo de un ensayo clínico es obtener conclusiones veraces para una cuestión médica relevante. Esto requiere que las conclusiones estén basadas en una evaluación imparcial y objetiva.

## 2.2. Sir Ronald A. Fisher y la aleatorización

Usualmente en la literatura sobre inferencia estadística se menciona a Sir Ronald A. Fisher como el inventor de la aleatorización. Dicha atribución se le hace ya que logró establecer en los experimentos, lo que él mismo llamó, 'las bases que motivan la inferencia'.

Para dar una explicación de tales bases, se valió de, lo que ahora es uno de los ejemplos más famosos de la literatura estadística, el ejemplo de la Dama Catadora de Té, el cual plasmó en su libro *The Design of Experiments* (Diseño de experimentos), ejemplo mismo que a continuación será introducido para dar una idea de aquello que Fisher llamó las bases que motivan la inferencia (propiedades de la aleatorización), esto es, la relación que existe entre la aleatorización y los métodos estadísticos.

La selección de Fisher sobre este ejemplo fue muy acertada, ya que ilustra de manera muy simple puntos clave de la aleatorización.

### 2.2.1. La Dama Catadora de Té

En el país de origen de Fisher, Inglaterra, una bebida tradicional para el desayuno es el té con leche. Una dama declaraba que ella era capaz de discernir si en una taza de té con leche, se había vertido primero el té o primero la leche. Movidio por estas declaraciones, Fisher diseña el siguiente experimento:

Se preparan ocho tazas de té con leche, en cuatro de ellas se vierte primero el té y en las otras cuatro se vierte primero la leche; entonces las tazas son presentadas ante la dama en orden aleatorio para que ésta dé su juicio sobre el orden de la preparación.

La dama es informada previamente de la forma en que es llevado el experimento y se le explica que el orden aleatorio no es determinado arbitrariamente por un humano, sino que se utilizan aparatos físicos usados en juegos de azar o se usa una colección de números aleatorios para tal ordenación.

La tarea de la dama es hacer la cata y dividir las ocho tazas en dos grupos de cuatro, agregando, si es posible, qué tratamiento recibió cada grupo.

Entonces Fisher se plantea la siguiente pregunta: ¿Qué se espera si la dama no tiene ninguna facultad de discriminación, esto es, si ella no es capaz de discernir el tratamiento que recibe cada taza de té con leche?

En respuesta a esto, Fisher propone: Supóngase, por un momento, que la dama no es capaz de discriminar cómo se preparó una taza de té con leche, es decir, ella da la misma respuesta sin importar cuáles son las cuatro tazas que reciben primero la leche, por ejemplo. Entonces, se tiene que hay  $\binom{8}{4} = 70$  posibles combinaciones de cómo las tazas pueden ser divididas en dos grupos de cuatro, y la aleatorización asegura que estas combinaciones sean igualmente probables, así dicha probabilidad es de  $1/70$  para cada combinación.

En otras palabras, la probabilidad de que el orden aleatorio de las tazas coincida perfectamente con el juicio dado por la dama es de  $1/70$ . Si la dama clasifica correcta-

mente las tazas, la probabilidad,  $0.014 = 1/70$ , es el nivel de significancia para probar la hipótesis nula de que ella no tiene ninguna facultad de discriminación, ya que si

$H_0$ : La dama no tiene ninguna facultad de discriminación, entonces

Probabilidad(*elegir la combinación correcta* | *no tiene facultad de discriminación*) =

Probabilidad(*elegir la combinación correcta* |  $H_0$  es cierta) =

Probabilidad(*Rechazar  $H_0$*  |  $H_0$  es cierta) =  $0.014$

Fisher entonces prosigue describiendo la aleatorización como *la base que motiva la inferencia y la base física de la validez de la prueba*, es decir, la considera el gran elemento de juicio. En su libro, Fisher hace una larga discusión de esta base y decide demostrar con verdaderos fundamentos como se logra que funcione la inferencia. Así formó su llamada 'base razonada' o 'elemento de juicio', una base que no dependiera del mismo patrón de respuestas, o de los mismos datos, exceptuando la aleatorización.

El ejemplo de la Dama, formulado por Fisher, sirvió perfectamente para su propósito. La Dama no es una muestra de una población de Damas, y aún si se imagina que ella lo fuera, el experimento y sus hipótesis sólo la conciernen a ella. Sus ocho juicios no son observaciones independientes, ya que el diseño del experimento requiere una división de cuatro y cuatro, y la decisión de la Dama en las tazas posteriores a la cuarta está influenciada por la decisión en las cuatro primeras. Por ejemplo, para la quinta taza, la Dama sabe que ya probó, al menos, una preparación en la que se vertió primero la leche y, al menos, una donde se vertió primero el té. No hay forma de interpretar o, mejor dicho, de malinterpretar los datos de este experimento con los de otro que involucre la muestra de una población o con uno que sea una serie de eventos idénticos e independientes. Además la inferencia de Fisher es justificada, ya que, la única distribución de probabilidad que se utiliza es la creada por el experimentador.

Lo interesante es conocer cuáles son las herramientas en el argumento de Fisher que permiten generalizar sus bases a otro tipo de experimentos y son las siguientes:

1. Los experimentos no requieren, de hecho no es razonable exigir, que las unidades experimentales sean homogéneas, es decir, que no tengan variabilidad en sus respuestas. Las unidades experimentales homogéneas no hacen una descripción real de las operaciones de una fábrica, de los pacientes de un hospital, de campos de cultivo, etc.
2. Los experimentos no requieren, de hecho no es razonable exigir, que las unidades experimentales sean una muestra aleatoria cualquiera de una población de unidades.

Las muestras aleatorias cualesquiera de unidades experimentales no son la realidad de la industria de laboratorio, de un ensayo clínico o de un experimento de agricultura.

3. Para inferencias válidas acerca del efecto del tratamiento sobre las unidades del experimento es suficiente requerir que los tratamientos sean asignados de forma aleatoria a las unidades experimentales - estas unidades deben ser, ambas, heterogéneas y no una muestra aleatoria cualquiera de una población.
4. Las probabilidades entran al experimento únicamente a través de la asignación aleatoria de los tratamientos, un proceso controlado por el experimentador. Una cantidad que no es afectada por la asignación aleatoria de los tratamientos es una cantidad fija al describir las unidades en el experimento.

## 2.3. Diseño del Ensayo Clínico

La forma en que es diseñado el ensayo clínico influye en cómo debe ser ejecutado y en el tipo de análisis que se le hará a los resultados. Los diseños más comunes actualmente son los siguientes:

- Diseño paralelo
- Diseño de tratamiento sucesivo, que puede ser de dos tipos:
  - Diseño de tratamiento de reemplazo
  - Diseño cruzado (Crossover)
- Diseño de ensayos alternativos, que puede ser de distintos tipos, entre ellos:
  - Diseño factorial
  - Ensayos de equivalencia
  - Aleatorización por conglomerados (Cluster)

### 2.3.1. Diseño paralelo

Este tipo de diseño es el más usado por su simplicidad. El término paralelo hace referencia a la coincidencia en el tiempo de aplicación de los tratamientos a cada uno de los grupos participantes en el ensayo.

Se forman los dos grupos, tratamiento y control, y a cada uno se le aplica, de forma aleatoria, el tratamiento correspondiente, entonces se sigue la evolución de los mismos

a lo largo de la duración del ensayo. La sencillez de este diseño radica en que se elige un sólo tratamiento para cada grupo, con un tamaño de muestra previamente fijado.

Si se tiene una muestra de tamaño  $N$  y de esta muestra un número de unidades  $m$  pertenece al grupo tratamiento (por lo tanto,  $N-m$  pertenecen al grupo control), y los tratamientos a probar y control son A y B respectivamente, entonces una representación gráfica del ensayo paralelo es la siguiente:

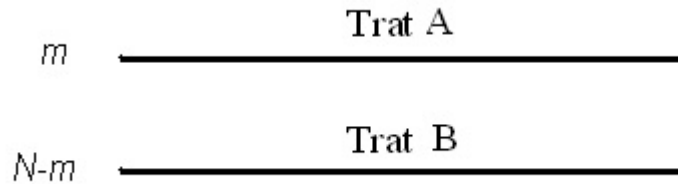


Figura 2.2: Representación de un ensayo con diseño paralelo

### 2.3.2. Diseño de tratamiento sucesivo

En este tipo de diseño, cada sujeto es asignado aleatoriamente a un grupo. Dentro de cada uno de los grupos se establece una secuencia de tratamientos, de forma que cada sujeto recibe más de un tratamiento.

La forma más frecuente es el diseño sucesivo en dos periodos, donde se aplican dos tratamientos de manera sucesiva, dejando entre ellos un periodo con uso de placebos, favoreciendo así la disipación de los efectos residuales del primer tratamiento, por ello, este periodo también se conoce como periodo intermedio de lavado (PIL).

Básicamente se usan dos tipos de diseño de tratamiento sucesivo: el diseño de tratamiento de reemplazo y el diseño cruzado,

- El diseño de tratamiento de reemplazo es usado para recolectar datos sobre efectos que tiene el cambiar de un tratamiento A a uno de dos tratamientos alternativos, por ejemplo, tratamiento B o tratamiento C. Los sujetos de estudio se dividen en dos grupos iguales. Ambos grupos reciben el tratamiento A durante un primer periodo. Después del periodo intermedio de lavado, a cada grupo se le aplica un tratamiento alternativo distinto, B o C. Las observaciones hechas entre los pacientes tratados con A y B se comparan con los resultados observados entre los pacientes tratados con A y C; como se esquematiza en la Figura 2.3.

Se observa que aunque no existe un grupo control perfectamente definido, existen los grupos comparables necesarios para el análisis estadístico.

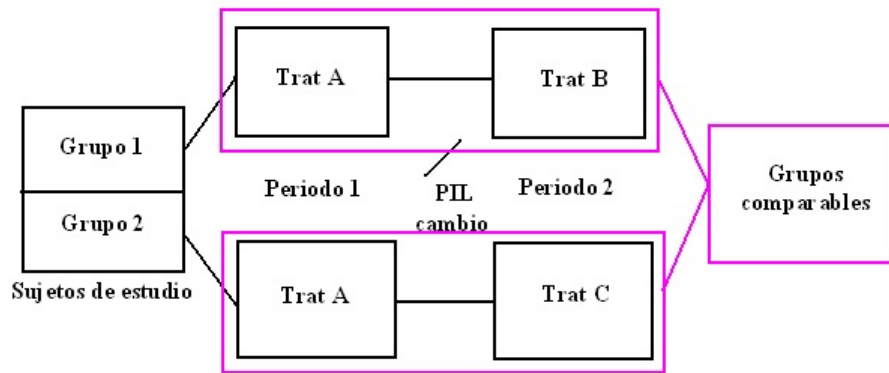


Figura 2.3: Representación de un ensayo con diseño de tratamiento de reemplazo

- El diseño cruzado es un diseño en el cual, los sujetos de estudio se dividen en dos grupos (grupo 1 y grupo 2). El grupo 1 recibe el tratamiento A durante un primer periodo y el tratamiento B en el segundo. El grupo 2 recibe los tratamientos durante los mismos periodos pero en orden inverso al grupo 1 (primero B, luego A). Los diseños cruzados permiten ajustar las variaciones de persona a persona haciendo que cada sujeto sirva como su propio control. En este diseño se exige con frecuencia un menor número de sujetos en relación con otros diseños, y su esencia es que cada sujeto sirve como su propio control. Los diseños cruzados pueden ser de dos vías (con dos tratamientos), como el que se explicó anteriormente o de varias vías (con varios tratamientos)<sup>3</sup>.

A continuación se presenta un esquema de diseño cruzado de dos vías:

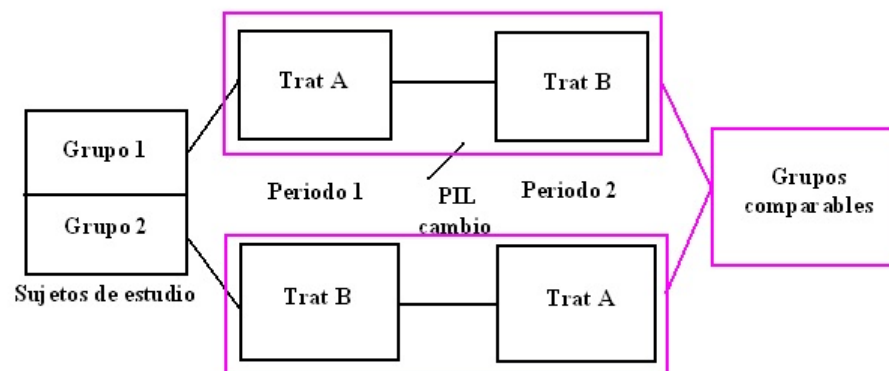


Figura 2.4: Representación de un ensayo con diseño cruzado

<sup>3</sup>Esto hará que el número de grupos pueda variar según los propósitos que se persigan (distintas combinaciones y ordenaciones en la administración de los tratamientos).

### 2.3.3. Diseños alternativos

Son diseños a los que se recurre porque cubren requerimientos que los diseños estándar no cubren y proveen combinaciones necesarias de los diseños anteriores.

- Diseño factorial. La evaluación de dos o más tratamientos en un mismo ensayo puede ser implementada usando un diseño de tipo paralelo, con un aumento en el número de grupos. Sin embargo, se requiere aumentar el tamaño de muestra y puede ser ineficiente. Además es inadecuado si también hay interés en considerar combinaciones de los tratamientos y de uno o varios factores adicionales asignados aleatoriamente con el fin de analizar las interacciones entre ellos. El diseño alternativo adecuado para esta situación es de tipo factorial, donde se pueden asignar de manera aleatoria dos o más tratamientos en forma independiente, siempre y cuando no exista una interacción; de tal manera que los sujetos pueden no recibir ningún tratamiento, uno de ellos o eventualmente todos; haciendo, además (sólo si son necesarias), combinaciones de otros factores, por ejemplo, distintas dosis.

El siguiente cuadro<sup>4</sup> presenta un ejemplo de cómo se puede hacer un ensayo con diseño factorial:

*Diseño factorial. A cada letra se le asigna aleatoriamente un bloque de 14 pacientes*

Tratamiento
A: Diltiazem – dosis 1
B: Diltiazem – dosis 2
C: Diltiazem – dosis 3
D: Diltiazem – dosis 4
E: Hidrolorotiazida – dosis 1
F: Hidrolorotiazida – dosis 2
G: Hidrolorotiazida – dosis 3
H: Diltiazem – dosis 1 + hidroclorotiazida – dosis 1
I: Diltiazem – dosis 1 + hidroclorotiazida – dosis 2
J: Diltiazem – dosis 1 + hidroclorotiazida – dosis 3
K: Diltiazem – dosis 2 + hidroclorotiazida – dosis 1
L: Diltiazem – dosis 2 + hidroclorotiazida – dosis 2
M: Diltiazem – dosis 2 + hidroclorotiazida – dosis 3
N: Diltiazem – dosis 3 + hidroclorotiazida – dosis 1
O: Diltiazem – dosis 3 + hidroclorotiazida – dosis 2
P: Diltiazem – dosis 3 + hidroclorotiazida – dosis 3
Q: Diltiazem – dosis 4 + hidroclorotiazida – dosis 1
R: Diltiazem – dosis 4 + hidroclorotiazida – dosis 2
S: Diltiazem – dosis 4 + hidroclorotiazida – dosis 1
T: Placebo

Cuadro 2.1: Ejemplo de diseño factorial

<sup>4</sup>Tomado del libro Bases Metodológicas del Ensayo Clínico por Fernando Tato



- Diseño de equivalencia. Se usan cuando se quiere demostrar que dos tratamientos son efectivamente similares respecto a la respuesta del paciente. Estos diseños son no sesgados y evalúan diferencias en tratamiento cercanas a cero y con un estrecho intervalo de confianza. Se practican porque existen tratamientos que pueden diferir en seguridad, efectos secundarios, formas de aplicación, costos, entre otras características; y el hecho de mostrar 'equivalencia' tiene importancia para el uso subsiguiente de uno o ambos tratamientos.
- Diseño de Aleatorización por conglomerados (*Clusters*<sup>5</sup>). En este tipo de diseños, una sola respuesta ante determinado tratamiento es dada por un grupo de sujetos, es decir, un conglomerado de sujetos es considerado una unidad.

## 2.4. Determinación del tamaño de la muestra y de la longitud del Ensayo

El número total de sujetos que participarán en un ensayo clínico a realizar es lo que conocemos como tamaño de muestra del ensayo, el cual deber ser lo suficientemente grande como para obtener información estadísticamente significativa acerca del producto probado sin llegar a convertirse en un obstáculo para la realización misma. El tamaño de muestra del ensayo determina la longitud del ensayo.

Es común, que en los ensayos se tome como tamaño de muestra el mínimo requerido para detectar la eficacia del nuevo producto. Hay ensayos en los que en lugar de pre-determinar el tamaño de muestra se predetermina el número de años-paciente, que es el tiempo acumulado por todos los sujetos del ensayo. Entonces, el número actual de sujetos que deberían participar en el ensayo depende de la frecuencia de la inscripción de los pacientes al ensayo, lo cual puede ser algo difícil de calcular.

El tamaño de muestra o el número de años-paciente debe ser determinado antes de iniciar el ensayo y debe ser documentado en el protocolo.

Al momento de calcular el tamaño de muestra se deben tener en cuenta las siguientes consideraciones:

- Definir el objetivo del ensayo clínico, que puede ser de tres tipos:
  - Que el paciente alcance un porcentaje de cambio en algunas mediciones médicas
  - Que el paciente alcance determinado estado o nivel de cambio

---

<sup>5</sup>Traduciendo literalmente del inglés, 'clusters' es 'agrupaciones'.

- Que se alcance determinado nivel en una tasa específica de eventos adversos también conocida como tasa de complicación (calculada como el número de eventos adversos que han ocurrido durante todo el periodo entre el tiempo total en el ensayo por todos los sujetos).
- Elegir la familia de distribuciones de probabilidad para los datos y estimar los parámetros necesarios de acuerdo a datos históricos.
- Identificar la hipótesis nula,  $H_0$ , y la alternativa,  $H_1$ , según el tipo de objetivo que se persigue con el ensayo (basándose en qué se quiere probar: el nuevo producto es mejor que el estándar, el nuevo producto tiene ventajas distintas al estándar, el nuevo producto es más eficaz que el estándar pero también más riesgoso).
- Determinar,  $\alpha$ , el nivel de significancia que es el tamaño del error tipo I (máxima probabilidad de rechazar la hipótesis nula dado que es cierta), generalmente se determina de 0.01 ó 0.05.

Representa la probabilidad,  $p$ , de hacer válida la diferenciación entre tratamientos, cuando dicha diferencia no existe en realidad. Al aplicarse pruebas estadísticas adecuadas sobre los parámetros o variables de interés, se debe rechazar la hipótesis nula,  $H_0$ , si  $p \leq \alpha$ .

- Determinar,  $\beta$ , el tamaño del error tipo II (probabilidad de no rechazar la hipótesis nula,  $H_0$ , dado que es falsa), usualmente se elige entre 0.05 y 0.25.

Representa la probabilidad de aceptar que no hay diferencia significativa en los tratamientos cuando en realidad si la hay.

El control de esta probabilidad o tamaño del error tipo II, sólo se puede asumir en la etapa del diseño y sólo si se contrastan hipótesis simples, por lo tanto resulta mucho más importante su obtención y derivación conceptual inmediata a la realización de la prueba.

- Considerar la potencia de la prueba, que en realidad ofrece un sustento para el diseño del ensayo, ya que estima la probabilidad de detectar la diferencia entre tratamientos cuando realmente ésta existe.

La potencia de la prueba,  $\pi$ , se calcula como:

$$\pi = 1 - (\text{tamaño del error tipo II}) = 1 - \beta$$

entonces, la potencia de la prueba representa la probabilidad de rechazar la hipótesis nula dado que una hipótesis alternativa **específica** es válida.

Haciendo referencia a los dos últimos puntos, es necesario e importante considerar que el valor  $\beta$  sólo se puede obtener cuando se hace un contraste de hipótesis (nula y

alternativa) simples, pues cuando se tiene un contraste de una hipótesis nula cualquiera contra una hipótesis alternativa compuesta, el comportamiento de esta probabilidad está determinado por una función llamada la Función Potencia, la cual proporciona una probabilidad específica para cada uno de los posibles valores que pueda tomar la variable o parámetro sobre el cual se están postulando hipótesis o supuestos, por ello, cuando se tiene una hipótesis alternativa específica o simple (lo que se recalca en los puntos anteriores) sí se puede obtener el valor de la función potencia en el punto que señala dicha hipótesis alternativa y este valor será obtenido a través de  $\beta$ . La Función Potencia da la probabilidad de rechazar que hay diferencia significativa entre los tratamientos para cada uno de los rangos de diferencia que existan.

Como consecuencia de lo anterior, si se tuviera una hipótesis alternativa compuesta (donde no se establece una hipótesis sobre un valor específico sino sobre todo un conjunto de valores), tampoco se podría obtener el valor de la potencia de la prueba, pues se carece de un valor específico  $\beta$ .

Una forma de obtener un valor específico  $\beta$ , es contrastar hipótesis simples valiéndose del valor *mínimo detectable* (o tamaño del efecto) en las respuestas para el grupo tratamiento y el grupo control. Este mínimo detectable, se establece por los investigadores según lo prometedor que, ellos consideran, es el nuevo tratamiento.

La longitud del ensayo depende del objetivo del ensayo; si se persigue un porcentaje de cambio o un cambio de estado/nivel, se debe calcular un tamaño de muestra; pero si se persigue una tasa de complicación debe calcularse un número de años-paciente.

Para ilustrar un poco como se hace este cálculo del tamaño de muestra, a continuación se presentan dos ejemplos<sup>6</sup> muy sencillos pero ilustrativos de la forma correcta de proceder.

**Ejemplo 1:** Se realiza un ensayo clínico para probar la eficacia de un nuevo tratamiento en la disminución de la presión sanguínea en pacientes que sufren de hipertensión. Los sujetos control reciben un tratamiento que está vigente en el mercado. Los investigadores centran el objetivo del experimento en un porcentaje de reducción en la presión sanguínea al momento de la diástole<sup>7</sup>.

Se denotarán por  $\mu_{tr}$  y  $\mu_c$  los verdaderos valores del porcentaje **medio** de reducción en la presión de la sangre para el grupo tratamiento y para el grupo control, respectivamente. Los investigadores acuerdan que los siguientes serán los datos utilizados para el cálculo del tamaño de muestra:

- Las hipótesis de interés son  $H_0: \mu_{tr} = \mu_c$  vs  $H_1: \mu_{tr} > \mu_c$ . Se hace un contraste

---

<sup>6</sup>Ejemplos tomados del libro *Clinical Statistics* de Olga Korosteleva

<sup>7</sup>Fase de relajación de las paredes ventriculares que produce el retroceso de la sangre de las arterias hacia el corazón: la sístole y la diástole son los dos movimientos del corazón.

de una sola cola (prueba unilateral) porque los investigadores confían en que el nuevo tratamiento no puede ser peor que el vigente.

- La probabilidad del error tipo I,  $\alpha = \text{Probabilidad}(\text{rechazar } H_0 \mid H_0 \text{ es cierta})$  es establecida en 0.05.
- Se considera que la diferencia mínima detectable,  $\delta = \mu_{tr} - \mu_c$ , sea del 5%, es decir,  $\delta = 5$ .
- La probabilidad del error tipo II,  $\beta = \text{Probabilidad}(\text{no rechazar } H_0 \mid H_1: \mu_{tr} - \mu_c = \delta \text{ se cumple})$  se fija en 0.25.
- Los datos obtenidos en la fase II del ensayo sugieren que la distribución de probabilidad que representa su comportamiento es aproximadamente normal con una desviación estándar igual a 15,  $\sigma = 15$ .
- La prueba Z para medias de dos muestras se aplica sobre dos grupos del mismo número de sujetos, n, en cada grupo.

El objetivo de este ejemplo es encontrar el valor de n, el tamaño de cada grupo requerido en el ensayo clínico. Se denota por  $\overline{X}_{tr}$  y  $\overline{X}_c$  a los valores medios desconocidos de los porcentajes de reducción que serán observados en la Fase III del ensayo clínico en el grupo tratamiento y en el grupo control, respectivamente. Se supone que los dos grupos son independientes.

Bajo  $H_0$ , la estadística de prueba

$$Z = \frac{\overline{X}_{tr} - \overline{X}_c}{\sigma\sqrt{2/n}} \sim N(0, 1)$$

La región de no rechazo - la región en la que  $H_0$  no es rechazada - es de la forma

$$\{Z < k\} = \left\{ \frac{\overline{X}_{tr} - \overline{X}_c}{\sigma\sqrt{2/n}} < k \right\} = \left\{ \overline{X}_{tr} - \overline{X}_c < k\sigma\sqrt{2/n} \right\}$$

para alguna constante k real positiva. Si se cumple la hipótesis alternativa  $H_1: \mu_{tr} - \mu_c = \delta$ , entonces

$$\overline{X}_{tr} - \overline{X}_c \sim N(\delta, 2\sigma^2/n)$$

Tomando en cuenta lo anterior, las probabilidades de los errores tipo I y tipo II definen dos ecuaciones para n y k:

$$1 - \alpha = \phi(k) \dots\dots\dots \textcircled{1},$$

$$\beta = \phi\left(k - \frac{\delta}{\sigma\sqrt{2/n}}\right) \dots\dots\dots \textcircled{2}$$

ya que

$$1 - \alpha = \text{Probabilidad}(Z < k \mid Z \sim N(0, 1)) = \phi(k) \text{ para ecuación } \textcircled{1} \text{ y}$$

$$\beta = \text{Probabilidad}(\overline{X}_{tr} - \overline{X}_c < k\sigma\sqrt{2/n} \mid \overline{X}_{tr} - \overline{X}_c \sim N(\delta, 2\sigma^2/n)) =$$

$$\text{Probabilidad}\left(\frac{\overline{X}_{tr} - \overline{X}_c}{\sigma\sqrt{2/n}} < k \mid \frac{\overline{X}_{tr} - \overline{X}_c - \delta}{\sigma\sqrt{2/n}} \sim N(0, 1)\right) =$$

$$\phi\left(k - \frac{\delta}{\sigma\sqrt{2/n}}\right) \text{ para ecuación } \textcircled{2}$$

Donde  $\phi$  denota la función de distribución acumulada de una variable aleatoria que se distribuye  $N(0,1)$ .

Despejando  $n$  de la ecuación  $\textcircled{2}$  y sustituyendo  $k$  por  $\phi^{-1}(1 - \alpha)$  obtenido de un despeje en la ecuación  $\textcircled{1}$  puede deducirse que

$$n = 2(\sigma/\delta)^2 \left(\phi^{-1}(1 - \alpha) - \phi^{-1}(\beta)\right)^2 \dots\dots\dots \textcircled{3}$$

En realidad, el valor de  $n$ , por ser un entero, se elige como el valor entero más cercano que exceda al valor obtenido con  $\textcircled{3}$ , lo cual provoca que la probabilidad del error tipo II sea ligeramente menor que el valor que se especificó inicialmente (en la ecuación  $\textcircled{2}$  se puede ver como  $\beta$  depende de  $n$  y que conforme ésta última crece,  $\beta$  decrecerá). Para este ejemplo, al sustituir los valores correspondientes a  $\alpha = 0.05$ ,  $\beta = 0.25$ ,  $\sigma = 15$  y  $\delta = 5$ , y usando una tabla de la función de distribución de la Normal Estándar como la del Anexo 2, resulta que  $n = 96.83$ , lo cual implica que  $n = 97$  para cada uno de los dos grupos. El tamaño real del error tipo II es de 0.249.

**Ejemplo 2:** Se lleva a cabo un ensayo clínico no aleatorizado para probar como funciona el implante de una nueva válvula del corazón. A los investigadores les gustaría estudiar tasas de varias complicaciones relacionadas con el implante de la válvula. Sin embargo, sólo debe elegirse una de estas tasas como la tasa principal u objetivo

para poder hacer la estimación de la longitud del ensayo. Entonces a continuación se explicará cómo se elige dicha tasa.

En el análisis estadístico:

- Se compara una tasa de complicación bajo el nuevo implante,  $R$ , con un valor histórico  $R_h$ .
- Se contrasta la hipótesis nula  $H_0 : R \geq 2R_h$  con la hipótesis alternativa  $H_1 : R < 2R_h$ . Nótese que la hipótesis nula indica que la nueva válvula funciona mucho peor que una válvula 'histórica' ya existente, si no se rechaza la hipótesis nula, la válvula no debería ser patentada.
- El tamaño del error tipo I es  $\alpha = 0.05$
- El tamaño del error tipo II, suponiendo que  $R = R_h$ , es 0.2.
- El número de complicaciones  $X$  es modelado como una variable aleatoria con función de distribución Poisson con media  $\lambda = RT$  sobre un periodo fijo de tiempo,  $T$ .
- La media histórica del número de complicaciones es  $\lambda_h = R_h T$ . Las hipótesis nula y alternativa pueden ser escritas como  $H_0 : \lambda \geq 2\lambda_h$  y  $H_1 : \lambda < 2\lambda_h$  respectivamente. El valor específico de la hipótesis alternativa para la cual se calcula  $\beta$  es  $\lambda = \lambda_h$ . El valor de  $\lambda_h$  puede ser calculado de las ecuaciones para  $\alpha$  y  $\beta$ .
- La tasa histórica  $R_h = 0.012$  o 1.2% para endocarditis (inflamación del revestimiento del corazón y las válvulas) es la más pequeña entre todas las tasas de complicación históricas consideradas. Por lo tanto, el número de años-paciente correspondiente,  $T = \lambda_h / R_h$ , es el mayor. Entonces, la tasa de endocarditis debe ser elegida como la principal para el ensayo.

Con estos datos ya se puede calcular el número de años-paciente. Se denotará con  $x$  el número observado de complicaciones por endocarditis. Por la forma de las hipótesis, la región de aceptación es de la forma  $\{x > x_0\}$ , donde el valor crítico  $x_0$  es un entero positivo.

Para un  $x_0$  fijo,  $\alpha = \max_{\lambda \geq 2\lambda_h} P(X \leq x_0)$  corresponde al caso  $\lambda = 2\lambda_h$ .

Por lo tanto las ecuaciones para  $\alpha$  y  $\beta$  son

$$1 - \alpha = P(X > x_0 | \lambda = 2\lambda_h) \text{ y } \beta = P(X > x_0 | \lambda = \lambda_h)$$

donde  $X \sim \text{Poisson}(\lambda)$ .

Estas ecuaciones definen un sistema de dos ecuaciones no lineales con dos incógnitas,  $x_0$  y  $\lambda_h$ :

$$1 - \alpha = \sum_{k=x_0+1}^{\infty} \left( \frac{(2\lambda_h)^k}{k!} \right) e^{-2\lambda_h} \text{ y}$$

$$\beta = \sum_{k=x_0+1}^{\infty} \left( \frac{\lambda_h^k}{k!} \right) e^{-\lambda_h} \dots\dots\dots \textcircled{4}$$

El sistema no puede ser resuelto exactamente bajo la restricción de que  $x_0$  es un entero. Sin embargo, puede utilizarse una versión integrable de la distribución de Poisson. Para cualquiera  $Y \sim \text{Poisson}(\lambda_0)$ , y para cualquier real positivo  $y$ , se cumple que

$$P(Y > y) = \int_0^{\lambda_0} \left( \frac{u^y}{\Gamma(y+1)} \right) e^{-u} du$$

donde  $\Gamma(y+1) = \int_0^{\infty} v^y e^{-v} dv$  es la función gamma. De aquí, las ecuaciones  $\textcircled{4}$  pueden escribirse de la siguiente forma:

$$1 - \alpha = \int_0^{2\lambda_h} \left( \frac{u^{x_0}}{\Gamma(x_0+1)} \right) e^{-u} du \text{ y}$$

$$\beta = \int_0^{\lambda_h} \left( \frac{u^{x_0}}{\Gamma(x_0+1)} \right) e^{-u} du \dots\dots\dots \textcircled{5}$$

La solución numérica de estas ecuaciones es  $x_0 = 11.296$  y  $\lambda_h = 9.287$ . Para  $R_h = 0.012$ , se requiere una longitud de ensayo de  $T = \frac{\lambda_h}{R_h} = \frac{9.287}{0.012} = 774$  años-paciente. La solución aproximada para las ecuaciones  $\textcircled{4}$  es  $x_0 = 12$  y  $\lambda_h = 9.72$ , con lo cual se obtiene una longitud de  $T = \frac{\lambda_h}{R_h} = \frac{9.72}{0.012} = 810$  años-paciente.

### 2.4.1. Monitoreo intermedio de los datos

Dentro de los ensayos clínicos, a lo largo de su implementación, pueden realizarse monitoreos de los datos a manera de análisis intermedios, conocidos como monitoreos intermedios de los datos. Un monitoreo intermedio de datos es un análisis de los datos llevado a cabo mientras el ensayo sigue en proceso y sirve para determinar si el ensayo puede ser detenido, es decir, si puede ser terminado tempranamente debido a que los datos analizados hasta el momento, ya son suficientes para determinar que el nuevo tratamiento es superior al estándar y que es eficaz o, en su defecto, para determinar que el producto probado es peligroso y/o menos eficaz.

Los investigadores clínicos deben decidir, a priori, si el ensayo será conducido hasta la totalidad del estudio o si se harán análisis intermedios durante el ensayo. Si los investigadores tienen suficiente confianza en el producto probado, lo más razonable es elegir el camino de los monitoreos intermedios. Una vez que se opta por este camino, se debe pre-especificar el número de monitoreos intermedios, así como estimar estadísticamente el tamaño de las muestras que se emplearán en los análisis intermedios y todo esto debe ser documentado en el protocolo. De esta manera surge otro problema, que es el de determinar el tamaño de las muestras que se utilizarán en el análisis intermedio, si bien, parece un problema sencillo, debe hacerse notar que es un problema distinto al planteado en los ejemplos anteriores y por ello necesita un tratamiento particular.

Un método estadístico común, utilizado para calcular el tamaño de las muestras intermedias es la Prueba Clásica Secuencial de Grupo, que en la siguiente sección se trata con detalle.

Los métodos secuenciales son muy recurridos en la actualidad, principalmente en epidemiología genética, pero también en los ensayos clínicos en los que es habitual el uso de análisis intermedios. Esto, con el objetivo de determinar si los resultados que se han obtenido hasta el momento hacen que sea inaceptable, desde el punto de vista ético, la continuidad del estudio, al encontrarse diferencias de efectos inesperadamente grandes o tasas de efectos secundarios demasiado altas. Además de que puede significar una reducción considerable de costos humanos y económicos.

## 2.4.2. Prueba Clásica Secuencial de Grupo

La prueba clásica secuencial de grupo es un método estadístico utilizado particularmente para la determinación de tamaños de muestra en los grupos necesarios para la realización del monitoreo intermedio de los datos dentro de un ensayo clínico, y la forma de aplicarla es la siguiente: Al llegar el momento en que se determinó que se llevaría a cabo el monitoreo se tienen disponibles los datos de  $n$  sujetos en cada grupo (tratado y control), se realizará entonces un análisis intermedio sobre  $2n$  datos. Los grupos se comparan estadísticamente y, si no se rechaza la hipótesis alternativa, el ensayo clínico se detendrá. Por otro lado, si la hipótesis alternativa se rechaza, deberá continuar el ensayo hasta que haya disponibles otros  $2n$  datos,  $n$  nuevos de cada grupo y se analizará de nuevo. En este caso, la prueba estadística se conducirá sobre  $4n$  sujetos ( $4n$  datos) y, nuevamente, si la hipótesis alternativa no es rechazada, el ensayo se detendrá; por el contrario, si es rechazada, se continúa con evaluaciones periódicas cada vez que se tienen  $2n$  datos nuevos disponibles, y así se puede continuar hasta tener  $X$  conjuntos disponibles de  $2n$  datos cada uno. Al llegar a este punto, se realiza la última prueba estadística, que es decisiva y se termina el ensayo.

Se puede emplear el mismo procedimiento para ensayos no aleatorizados. Se con-



ducen pruebas intermedias con grupos de tamaño  $2n, 4n, \dots, X * 2n$ , o en su caso, de intervalos de tiempo iguales a  $2t, 4t, \dots, X * 2t$  años-paciente para casos como el del último ejemplo.

La probabilidad del error tipo I para las  $X$  pruebas estadísticas intermedias es una constante  $\alpha'$ . Para un número fijo  $X$ , los valores de  $\alpha'$  y  $n$  se pueden encontrar si  $\alpha$  y  $\beta$  -los tamaños de los errores tipo I y II, respectivamente- han sido especificados.

El tamaño del error tipo I se define como la probabilidad de que al menos haya una diferencia significativa dado que la hipótesis nula es cierta. El tamaño del error tipo II es la probabilidad de que todas las diferencias sean insignificantes dado que se cumple una hipótesis alternativa específica.

Con el siguiente ejemplo se ilustra como debe aplicarse la prueba secuencial de grupo para un monitoreo intermedio de datos en el ensayo clínico del ejemplo 1 tratado en la sección anterior<sup>8</sup>.

**Ejemplo 3.** En el ejemplo 1, las hipótesis de interés fueron  $H_0: \mu_{tr} = \mu_c$  y  $H_1: \mu_{tr} > \mu_c$ ,  $\alpha = 0,05$ ,  $\beta = 0,25$ ,  $\delta = 5$ ,  $\sigma = 15$ . Para llevar a cabo esta prueba, se necesita una muestra de tamaño 97 para cada grupo, esto es, en caso de no usar una prueba secuencial (o la prueba secuencial es con  $X = 1$ ), que es lo que se obtuvo en el ejemplo 1.

Se considerará ahora el caso, en que  $X = 2$ . Sean  $\overline{x_{tr}^{(i)}}$  y  $\overline{x_c^{(i)}}$  las medias muestrales respectivas al  $i$ -ésimo conjunto de datos de  $2n$  sujetos,  $i = 1$  ó  $2$ . Se denotará  $\overline{x_{tr}} = \frac{\overline{x_{tr}^{(1)}} + \overline{x_{tr}^{(2)}}}{2}$  y  $\overline{x_c} = \frac{\overline{x_c^{(1)}} + \overline{x_c^{(2)}}}{2}$  las medias muestrales respectivas a cada grupo cuando se trabaje sobre el conjunto de  $4n$  sujetos.

La primera prueba estadística de  $H_0: \mu_{tr} = \mu_c$  contra  $H_1: \mu_{tr} > \mu_c$  al nivel de significancia  $\alpha'$  es efectuada sobre el conjunto inicial de  $2n$  sujetos. Bajo  $H_0$ ,  $\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}} \sim N(0, 2\sigma^2/n)$ . La región en la que no se rechaza la hipótesis nula (región de 'no rechazo') es

$$\left\{ \frac{\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}}}{\sigma\sqrt{2/n}} < k \right\} = \left\{ \overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}} < k\sigma\sqrt{2/n} \right\}$$

La relación entre el nivel de significancia  $\alpha'$  y el valor crítico,  $k$ , de la región de aceptación está dada por la fórmula  $\phi^{-1}(1 - \alpha')$ , o equivalentemente,  $\alpha' = 1 - \phi(k)$ . Bajo una hipótesis alternativa específica  $H_1: \mu_{tr} - \mu_c = \delta$ , entonces  $\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}} \sim N(\delta, 2\sigma^2/n)$ .

Si en esta primera prueba no se rechaza la hipótesis nula, se lleva a cabo la segunda

---

<sup>8</sup>Ejemplo tomado del libro Clinical Statistics de Olga Korosteleva

prueba, y en dado caso la segunda prueba de  $H_0: \mu_{tr} = \mu_c$  contra  $H_1: \mu_{tr} > \mu_c$  al nivel de significancia  $\alpha'$  es efectuada sobre el conjunto de  $4n$  sujetos. La diferencia

$$\bar{x}_{tr} - \bar{x}_c = \frac{\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}}}{2} + \frac{\overline{x_{tr}^{(2)}} - \overline{x_c^{(2)}}}{2}$$

es la suma de  $\frac{\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}}}{2}$  y  $\frac{\overline{x_{tr}^{(2)}} - \overline{x_c^{(2)}}}{2}$  que son dos variables independientes que bajo  $H_0$  tienen una distribución  $N(0, \sigma^2/2n)$  y bajo una hipótesis alternativa específica  $H_1: \mu_{tr} - \mu_c = \delta$  tienen distribución  $N(\delta/2, \sigma^2/2n)$ .

De esta manera, bajo  $H_0$ , la distribución de  $\bar{x}_{tr} - \bar{x}_c$  es  $N(0, \sigma^2/n)$ . Por lo tanto, la región de 'no rechazo' para la segunda prueba es

$$\left\{ \frac{\bar{x}_{tr} - \bar{x}_c}{\sigma\sqrt{1/n}} < k \right\} = \left\{ \left( \overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}} \right) + \left( \overline{x_{tr}^{(2)}} - \overline{x_c^{(2)}} \right) < 2k\sigma\sqrt{1/n} \right\}$$

Bajo una hipótesis alternativa específica  $H_1: \mu_{tr} - \mu_c = \delta$  la distribución de  $\bar{x}_{tr} - \bar{x}_c$  es  $N(\delta, \sigma^2/n)$ . Las definiciones de  $\alpha$  y  $\beta$  proporcionan dos ecuaciones para  $k$  y  $n$ . La primera ecuación es

$$1 - \alpha = P \left\{ \frac{\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}}}{\sigma\sqrt{2/n}} < k, \frac{\bar{x}_{tr} - \bar{x}_c}{\sigma\sqrt{1/n}} < k \right\} = P \left\{ Z_1 < k, Z_1 + Z_2 < k\sqrt{2} \right\}$$

donde  $\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}} \sim N(0, 2\sigma^2/n)$  y  $\bar{x}_{tr} - \bar{x}_c \sim N(0, \sigma^2/n)$ , además  $Z_1 = \frac{\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}}}{\sigma\sqrt{2/n}}$  y  $Z_2 = \frac{\overline{x_{tr}^{(2)}} - \overline{x_c^{(2)}}}{\sigma\sqrt{2/n}}$  son variables aleatorias independientes que se distribuyen  $N(0, 1)$ . La segunda ecuación es

$$\beta = P \left\{ \frac{\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}}}{\sigma\sqrt{2/n}} < k, \frac{\bar{x}_{tr} - \bar{x}_c}{\sigma\sqrt{1/n}} < k \right\} = P \left\{ Z_3 + \frac{\delta}{\sigma\sqrt{2/n}} < k, Z_3 + Z_4 + 2\frac{\delta}{\sigma\sqrt{2/n}} < k\sqrt{2} \right\}$$

donde  $\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}} \sim N(\delta, 2\sigma^2/n)$  y  $\bar{x}_{tr} - \bar{x}_c \sim N(\delta, \sigma^2/n)$ , además  $Z_3 = \frac{\overline{x_{tr}^{(1)}} - \overline{x_c^{(1)}} - \delta}{\sigma\sqrt{2/n}}$  y  $Z_4 = \frac{\overline{x_{tr}^{(2)}} - \overline{x_c^{(2)}} - \delta}{\sigma\sqrt{2/n}}$  son variables aleatorias independientes que se distribuyen  $N(0, 1)$ . Para

simplificar la notación, sea  $n^* = \left(\frac{1}{2}\right) \left(\frac{\delta}{\sigma}\right)^2 n$ . En términos de  $k$  y  $n^*$ , las dos ecuaciones anteriores quedan de la siguiente forma

$$1 - \alpha = P\left(Z_1 < k, Z_1 + Z_2 < k\sqrt{2}\right)$$

$$\beta = P\left(Z_1 + n^* < k, Z_3 + Z_4 + 2\sqrt{n^*} < k\sqrt{2}\right).$$

La solución numérica de las últimas ecuaciones para  $\alpha = 0.05$ ,  $\beta = 0.25$  es  $k = 1.875$  ( $\alpha' = 0.030$ ) y  $n^* = 3.029$ . Por lo que, el tamaño de los grupos para los monitoreos intermedios es el menor entero mayor que  $2\left(\frac{\sigma}{\delta}\right)^2 n^* = 54.522$ , esto es  $n = 55$ . La probabilidad del error tipo II que corresponde a este tamaño de grupo es 0.246.

Entonces, en lugar de acumular 97 sujetos en cada grupo, y probar las hipótesis una vez al nivel de significancia 5 %, el método secuencial de grupo con  $X = 2$  sugiere que los investigadores prueben al nivel de significancia del 3 % con 55 sujetos en cada grupo y, si la hipótesis nula no es rechazada, probar una segunda vez al nivel de significancia del 3 % con dos grupos de 110 sujetos cada uno. Cuando los investigadores se encuentran muy seguros de la eficacia del producto probado, pueden optar por los métodos secuenciales, ya que es muy probable que los análisis intermedios den la evidencia suficiente para demostrar que es mejor que los productos estándar y, de esta forma, se pueda detener el ensayo clínico antes de lo esperado. Sin embargo, si los resultados obtenidos no son tan buenos como se esperaba, el ensayo debería continuar hasta que se hayan acumulado 110 sujetos para cada grupo, lo cual representa un ensayo más largo que uno sin monitoreos intermedios (pues en ese caso, bastaría con acumular 97 sujetos por grupo).

En el caso general, para cualquier  $X$ , las cantidades  $k$  y  $n^*$  pueden ser expresadas como se muestra a continuación:

$$1 - \alpha = P\left(\bigcap_{m=1}^N \{Z_1 + Z_2 + \dots + Z_m < k\sqrt{m}\}\right)$$

$$\beta = P\left(\bigcap_{m=1}^N \{Z_1 + Z_2 + \dots + Z_m + m\sqrt{n^*} < k\sqrt{m}\}\right)$$

donde  $Z_1, Z_2, \dots, Z_N$ , son variables aleatorias independientes con distribución  $N(0, 1)$ . De forma esquemática, la región de 'no rechazo' de la  $m$ -ésima prueba para  $m = 1, 2, \dots, N$ , puede representarse de la siguiente forma:

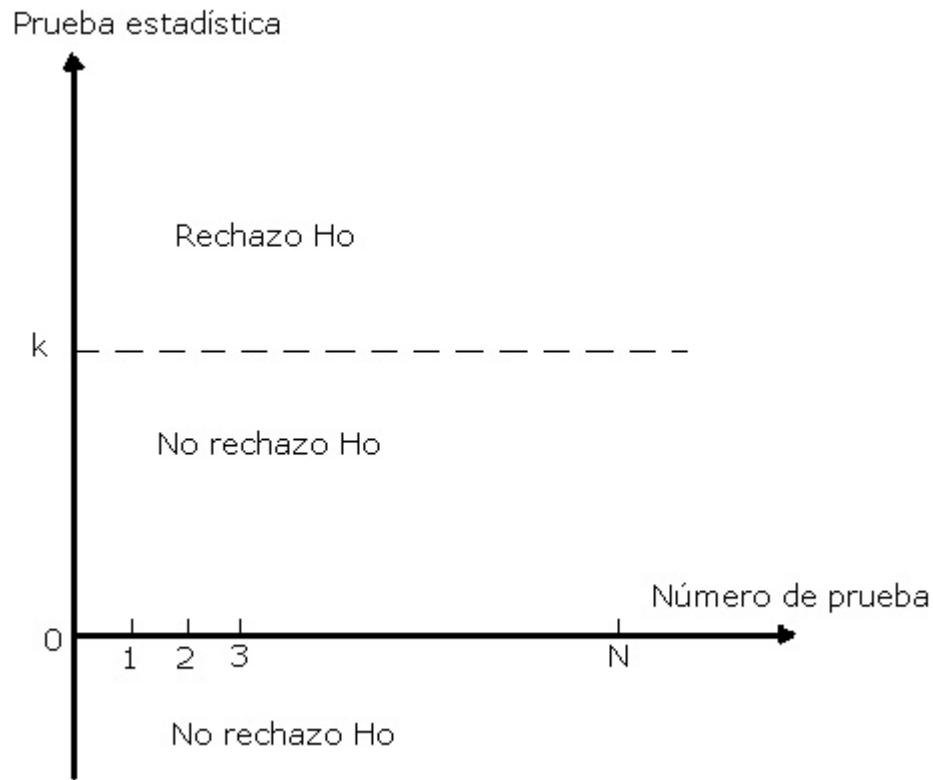


Figura 2.5: Representación de la región de 'no rechazo' para la  $m$ -ésima prueba

Nótese que la frontera de esta región es una línea horizontal. Esto corresponde a una constante  $\alpha'$ , la probabilidad del error tipo I para las pruebas intermedias. Existen otros métodos secuenciales en los que la frontera de la región de 'no rechazo' no es una línea horizontal y  $\alpha'$  no es la misma para todas las pruebas intermedias, pero dichos métodos no son tratados aquí.

## 2.5. Unidades y Vector de Asignación de tratamientos

Se mencionó con anterioridad que las personas participantes en un ensayo clínico se conocen como sujetos, pero se debe tomar en consideración que, aunque frecuentemente a cada sujeto se le asigna un tratamiento, para determinados experimentos muy raros, cada sujeto aporta más de una respuesta o dato, ya que se puede estar haciendo un experimento en el que se aplica más de un tratamiento a un mismo sujeto. De la misma forma, existen experimentos en los que un grupo de sujetos es necesario para aportar un sólo dato, ya que es al grupo de sujetos al que se aplica un determinado tratamiento

(se conocen como ensayos con 'clusters' o intervenciones comunitarias). Estos dos casos mencionados son muy raros en los ensayos clínicos pero existen en general en los experimentos aleatorizados, por ello, para hacer una generalización de todos los tipos de ensayos que se pueden realizar, se considerará que cada dato importante para el análisis estadístico es una unidad, sin importar si provienen varios datos de un mismo individuo o uno de varios de ellos.

Después de hacer esta aclaración, se puede decir que, en cada ensayo se tienen  $N$  unidades disponibles para la experimentación, este valor de  $N$  fue determinado como el tamaño de muestra necesario. Cada unidad es una oportunidad de aplicar o rechazar el tratamiento. Las  $N$  unidades son divididas en  $S$  estratos o subclases según las co-variables, es decir, con base en las características de los pacientes que fueron medidas y/o registradas antes de la asignación de los tratamientos. Al formar los estratos antes de aplicar el tratamiento, este último no afecta el estrato al que pertenece determinada unidad. Después de hacer esta división en estratos, se considera que el estrato  $s$  tiene  $n_s$  unidades para  $s = 1, 2, \dots, S$ , por lo tanto

$$N = \sum_{s=1}^S n_s$$

Para la  $i$ -ésima unidad en el estrato  $s$ , se asigna a  $Z_{si}$  el valor 1 si dicha unidad recibe el tratamiento a prueba y se asigna a  $Z_{si}$  el valor 0 si la unidad recibe el tratamiento control. Entonces, se denota  $m_s$  como el número de unidades que, dentro del estrato  $s$ , tienen asignado el tratamiento a prueba, por lo tanto,

$$m_s = \sum_{i=1}^{n_s} Z_{si}$$

Introducida esta notación, también se denota, como  $m$ , al número de unidades que tienen asignado el tratamiento a prueba del total de unidades  $N$ , de donde se deduce que

$$m = \sum_{i=1}^S m_i$$

Lo anterior, puede esquematizarse de la siguiente forma:

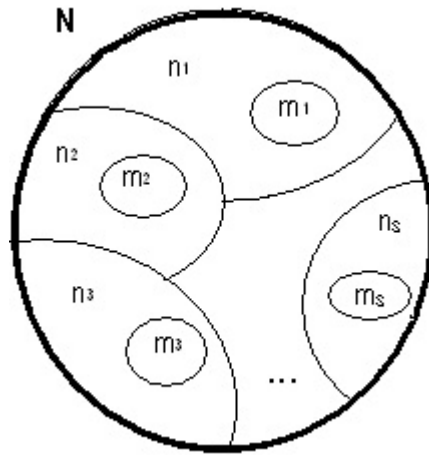


Figura 2.6: Representación de la división en estratos de la muestra de unidades

Finalmente, se denotará como  $Z$  al vector columna que contiene todos los valores asignados a  $Z_{si}$  para cada una de las unidades, esto es,

$$Z = \begin{bmatrix} Z_{1,1} \\ Z_{1,2} \\ \vdots \\ Z_{1,n_1} \\ Z_{2,1} \\ \vdots \\ Z_{s,n_s} \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_S \end{bmatrix}$$

donde  $Z_s = \begin{bmatrix} Z_{s,1} \\ Z_{s,2} \\ \vdots \\ Z_{s,n_s} \end{bmatrix}$  para  $s = 1, 2, \dots, S$ . Usando esta notación se puede hacer la representación de los tipos de ensayos más comunes. Por ejemplo:

- Si la muestra no es dividida en distintos estratos, por no considerar covariables, esto quiere decir que todas las unidades están contenidas en un único estrato, entonces  $S = 1$ , y se conoce como un **ensayo completamente aleatorizado**.
- Si  $S \geq 2$ , se tienen dos o más estratos y se dice que se trabaja con un **experimento aleatorizado por bloques**.

- Si  $n_s = 2$  y  $m_s = 1$  para  $s = 1, 2, \dots, S$ , entonces hay  $S$  pares de unidades con base en las covariables, donde cada par contiene una unidad de tratamiento y una unidad del grupo control, y se dice que se tiene un **experimento aleatorizado pareado**.
- Si  $n_s \geq 2$  y  $m_s = 1$  para  $s = 1, 2, \dots, S$ , entonces hay  $S$  estratos donde existe una unidad de tratamiento y una o más unidades del grupo control, esto significa que se tiene un **experimento controlado múltiple**.

## 2.6. Métodos de asignación aleatoria de tratamientos

Al ejecutar un experimento aleatorizado, la asignación del tratamiento correspondiente a cada sujeto (o unidad) es determinada por el experimentador, esto quiere decir, que es él quién determina si una unidad recibe el tratamiento control o el nuevo tratamiento, por lo tanto, también los valores del vector  $Z$  quedan determinados por el experimentador, quien se vale de algún mecanismo aleatorio conocido, tal como es una tabla de números aleatorios para dicha asignación.

Mencionar que el mecanismo aleatorio utilizado es conocido, sugiere que la distribución de la variable aleatoria  $Z$  es conocida ya que fue creada por el experimentador. Al mecanismo aleatorio elegido para la asignación se le exige que cumpla con un requerimiento esencial, este es que, antes de que los tratamientos sean asignados, cada unidad tenga una probabilidad distinta de cero de recibir tanto el tratamiento nuevo como el tratamiento control, es decir,

$$0 < \Pr(Z_{si} = 1) < 1$$

para  $i = 1, 2, \dots, N$  y  $s = 1, 2, \dots, S$ .

Se denotará,  $\Omega_0$ , al conjunto de todos los posibles valores de  $Z$ , es decir, todos los posibles valores de  $Z$  a los cuales el mecanismo otorga una probabilidad distinta de cero de aparecer.

En la práctica, se utilizan distintos mecanismos para determinar  $Z$ , pero el mecanismo más simple asigna los tratamientos de forma independiente a las diferentes unidades, tomando  $\Pr(Z_{si} = 1) = 1/2$  para toda  $s, i$ .

En el proceso de aleatorización simple, cada uno de los sujetos tiene la misma probabilidad de ser asignado a cualquiera de los grupos del ensayo. Este tipo de aleatorización se lleva a cabo usando tablas de números aleatorios (en la práctica es común el uso de

software generador de números aleatorios como es Excel o Minitab). Por ejemplo, si existe un grupo tratamiento y un grupo control, se puede establecer que cuando aparezcan los dígitos 0, 2, 4, 6 y 8, entonces el sujeto o la unidad irá al grupo tratamiento, en cambio, si aparecen los dígitos 1, 3, 5, 7 y 9 el sujeto formará parte del grupo control.

Considerando que para el ensayo clínico se cuenta con  $N$  unidades, entonces, el conjunto  $\Omega_0$  contiene  $2^N$  posibles valores de  $Z$ . Cada  $N$ -ada de ceros y unos (cada asignación en  $\Omega_0$ ) tiene la misma probabilidad de aparecer, por lo tanto,  $\Pr(Z = z) = \frac{1}{2^N}$  para toda  $z \in \Omega_0$ . Se denotará a,  $|\Omega_0|$ , como el número de elementos de  $\Omega_0$ , entonces  $|\Omega_0| = 2^N$ . Una peculiaridad de este mecanismo es que existe una probabilidad distinta de cero de que todas las unidades sean asignadas al mismo tratamiento (es posible que todas las unidades queden dentro del grupo control o todas queden dentro del grupo del nuevo tratamiento) aunque esta probabilidad es pequeña conforme aumenta  $N$ , el tamaño de la muestra. En la práctica, con este mecanismo surge un problema importante y nace cuando  $S$  es bastante grande en comparación con  $N$  (aunque  $S < N$ ). Si  $S$  es muy grande (con un valor cercano a  $N$ ), ocurre que existen muchos estratos y cada uno de ellos puede contener una cantidad pequeña de elementos lo que aumenta las probabilidades de que dentro de un estrato todas las unidades sean asignadas al mismo tratamiento (control o nuevo). Afortunadamente hay mecanismos que evitan esta posibilidad.

La forma más común de evitar este problema, es fijar el número,  $m_s$ , de unidades a las que se aplica el nuevo tratamiento dentro del estrato  $s$ . Entonces, las únicas asignaciones  $z$  para  $Z$  que tienen una probabilidad distinta de cero son aquellas con  $m_s$  sujetos tratados<sup>9</sup> en el estrato  $s$  para  $s = 1, 2, \dots, S$ . De esta forma se entiende claramente que se evitará el problema. Si de forma sensata se fija que  $m_s$  sea igual a  $n_s/2$  para cada  $s$ , al final de la asignación habrá un balance, pues la mitad de las unidades en cada estrato recibirán el tratamiento nuevo y la otra mitad el tratamiento control.

Si  $m_s$  se fija de esta forma, entonces  $\Omega_0$  contiene  $K$  posibles asignaciones de tratamiento diferentes, donde

$$K = \prod_{s=1}^S \binom{n_s}{m_s} \text{ y } K = |\Omega_0|$$

Y el vector  $Z$  queda determinado como:

$$Z_s = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_s \end{bmatrix}$$

---

<sup>9</sup>Se mencionan como 'sujetos tratados' porque son quienes reciben el tratamiento nuevo y no el tratamiento control, y a partir de este momento se hace referencia a ellos como 'sujetos tratados'.



donde cada  $Z_s$  es una  $n_s$ -ada con  $m_s$  unos y con  $n_s - m_s$  ceros para  $s = 1, 2, \dots, S$ . En el mecanismo de asignación más común, cada uno de los  $K$  elementos de  $\Omega_0$  tiene la misma probabilidad de aparecer,  $\Pr(Z = z) = 1/K$  para toda  $z$  en  $\Omega_0$ . Este tipo de experimentos aleatorizados con probabilidades iguales y  $m_s$  fijo, son llamados experimentos aleatorizados uniformes. Cuando además de estas condiciones,  $S = 1$ , entonces son comúnmente conocidos como experimentos completamente aleatorizados.

No se debe olvidar que según se fijen los valores de  $S$ ,  $m_s$  y  $n_s$  tendremos los distintos casos: ensayos completamente aleatorizados, ensayos aleatorizados por bloques, ensayos aleatorizados pareados y ensayos controlados múltiples.

La siguiente proposición es frecuentemente útil. Establece que en un experimento aleatorizado uniforme, las asignaciones en cada estrato son independientes de la asignación en los demás.

**Proposición 1.** *En un experimento aleatorizado uniforme, las  $Z_1, Z_2, \dots, Z_S$  son mutuamente independientes, y  $\Pr(Z_s = z_s) = \frac{1}{\binom{n_s}{m_s}}$  para cada  $n_s$ -ada  $z_s$  que contiene  $m_s$  unos y  $n_s - m_s$  ceros.*

Los diseños aleatorizados uniformes son los experimentos aleatorizados más comunes que involucran dos tratamientos, pero son aplicados otros más, especialmente en los ensayos clínicos.

Es de gran utilidad mencionar uno de estos métodos de aleatorización para subrayar que los experimentos aleatorizados no necesitan dar a cada asignación de tratamientos  $z \in \Omega_0$  la misma probabilidad. Un rasgo distintivo de muchos ensayos clínicos es que las unidades son pacientes que abordan el tratamiento en distintos momentos dentro de un periodo de meses o años. Como consecuencia, es imposible conocer al inicio del ensayo, el número de sujetos que tendrá cada estrato, lo que hace imposible la aplicación de un diseño aleatorizado por bloques.

Efron (1971) propuso un método para la asignación de tratamientos que es superior en muchos aspectos a los procedimientos tradicionales. Primero se fija una probabilidad  $p$  donde  $1/2 < p < 1$ . Cada vez que un paciente llega al estrato  $s$  se recalcula una medida llamada de *desequilibrio*. Cuando el  $i$ -ésimo paciente llega al estrato  $s$ , se calcula una nueva medida de *desequilibrio* en el estrato  $s$ ,  $DES_{si}$ , definido por el número de pacientes hasta entonces asignados al nuevo tratamiento dentro de ese estrato menos el número de pacientes hasta entonces asignados al tratamiento control dentro del mismo estrato. Si el  $DES_{si} = 0$ , se debe asignar al nuevo paciente uno de los dos tratamientos con probabilidad  $1/2$  para cada uno. Pero, si  $DES_{si} < 0$ , significa que el estrato  $s$  tiene menos pacientes asignados al nuevo tratamiento que pacientes asignados al tratamiento control, entonces se procede asignando al nuevo paciente el nuevo tratamiento con probabilidad  $p$  y el tratamiento control con probabilidad  $1 - p$ . De lo contrario, si  $DES_{si} > 0$ , el estrato  $s$  tiene más pacientes tratados que pacientes control entonces la

asignación se hace dando una probabilidad de  $1 - p$  para el nuevo tratamiento y de  $p$  para el tratamiento control.

Efron examinó muchos aspectos de este método, en particular, mostró que era mucho mejor para producir un equilibrio (balance) en los grupos tratado y control que la asignación independiente.

## 2.7. Prueba de Hipótesis de No Efecto del Tratamiento

### 2.7.1. La distribución de una Estadística de Prueba cuando el Tratamiento No tiene Efecto

Al diseñar un experimento, debe considerarse la posibilidad que sugiere que el tratamiento no tiene absolutamente ningún efecto. Por esto, es recomendable aplicar una prueba de No Efecto del Tratamiento. Dicha consideración puede tomarse en cuenta ya que en un experimento aleatorizado, esta prueba puede ser ejecutada sin suponer hipótesis de ningún tipo, confiando solo en la asignación aleatoria de los tratamientos. Esta afirmación se puede hacer como consecuencia de que Fisher discutió el problema de la Dama catadora de té con sumo cuidado para demostrarlo. Hay actividades, tales como estimar los efectos del tratamiento o construir intervalos de confianza que requieren contar con algunos supuestos, frecuentemente supuestos inocuos, pero supuestos al fin. La contribución de la aleatorización a la inferencia formal es más clara cuando es expresada en términos de la Prueba de No Efecto.

Si no existen dificultades tales como incumplimiento o pérdida de seguimiento, los supuestos juegan un rol menos importante en experimentos aleatorizados y no tienen ningún efecto sobre la aleatorización en las pruebas de hipótesis de No Efecto. En contraste, la inferencia en un experimento no aleatorizado requiere supuestos que no son del todo inocuos. Por lo tanto, está permitido seguir los pasos de Fisher y desarrollar este punto con cuidado.

Cada unidad del ensayo exhibe una respuesta, la cual es observada tiempo después de aplicar el tratamiento. Decir que el tratamiento no tiene efectos sobre esta respuesta es decir que cada unidad debería exhibir el mismo valor de la respuesta independientemente de si es asignado al grupo tratamiento o al grupo control. Esta es la definición de **No Efecto**.

Si cambiar el tratamiento asignado a una unidad cambió la respuesta de la unidad entonces ciertamente el tratamiento tiene al menos un efecto.

En el proceso de la aleatorización, el azar y las probabilidades entran en juego a través de la asignación aleatoria de los tratamientos, es decir, a través del mecanismo de aleatorización conocido con el que se seleccionó  $Z$  de  $\Omega$ . Por lo tanto, las únicas cantidades aleatorias que se encontrarán serán  $Z$  y las que dependan de  $Z$ .

Si un tratamiento no tiene ningún efecto, la respuesta de una unidad del ensayo quedará fija, en el sentido de que aunque fuera seleccionada de  $\Omega$ , una asignación de tratamientos distinta, la respuesta no debería cambiar en esa unidad debido al No Efecto.

Para probar la hipótesis nula de No Efecto del tratamiento, se denotará como  $r_{si}$  a la respuesta que tuvo ante el tratamiento la  $i$ -ésima unidad perteneciente al estrato  $s$  y a la  $N$ -ada de respuestas de las  $N$  unidades se le denotará por un vector  $\mathbf{r}$ . Al trabajar con subíndices para  $r_{si}$ , se enfatiza que al suponer la hipótesis nula,  $r_{si}$  es una cantidad construida y no una variable aleatoria.

Una estadística de prueba  $t(Z, \mathbf{r})$ , es una cantidad calculada a partir de la asignación<sup>10</sup>  $Z$  seleccionada y de la respuesta  $\mathbf{r}$ .

Por ejemplo, se puede definir como estadística de prueba, a la diferencia de medias muestrales de pacientes tratados menos pacientes control, es decir,

$$t(Z, r) = \frac{Z^T r}{Z^T \mathbf{1}} - \frac{(1 - Z)^T r}{(1 - Z)^T \mathbf{1}}$$

donde  $\mathbf{1}$  es una  $N$ -ada de uno's.

Dada cualquier estadística de prueba  $t(Z, \mathbf{r})$  la tarea es calcular un nivel de significancia para una prueba que refute la hipótesis nula de No Efecto del Tratamiento cuando  $t(Z, \mathbf{r})$  es grande. De manera más precisa:

1. La hipótesis nula de No Efecto es tentativamente tomada como verdadera, por lo tanto  $\mathbf{r}$  es fija
2. Una asignación  $Z$  ha sido seleccionada de  $\Omega$  usando un mecanismo aleatorio conocido
3. El valor observado,  $T$ , de la estadística de prueba  $t(Z, \mathbf{r})$  ha sido calculado
4. Se busca la probabilidad de que haya un valor de la estadística de prueba tan o más grande que el observado bajo la hipótesis nula.

---

<sup>10</sup>A partir de ahora se mencionará como 'asignación' a la 'asignación de tratamientos' con el fin de abreviar.

El nivel de significancia será simplemente la suma de las probabilidades de aleatorización de la asignación  $z \in \Omega$  que representan los valores de  $t(Z, \mathbf{r})$  iguales o más grandes que el valor observado,  $T$ , esto es,

$$\Pr(t(Z, \mathbf{r}) \geq T) = \sum_{z \in \Omega} [t(Z, \mathbf{r}) \geq T] * \Pr(Z = z)$$

donde  $r_{si} = \begin{cases} 1 & \text{si se presenta la respuesta esperada} \\ 0 & \text{en cualquier otro caso} \end{cases}$

y la probabilidad  $\Pr(Z = z)$  es determinada por el mecanismo aleatorio que se utilizó para la asignación de los tratamientos. Este es un cálculo directo, aunque no siempre sencillo ya que algunas veces  $\Omega$  puede ser extremadamente grande.

En el caso de un experimento aleatorizado uniforme, existe una expresión más simple para el nivel de significancia considerando que  $\Pr(Z = z) = 1/K = 1/|\Omega|$ . Esto es, la proporción de las asignaciones de tratamiento  $z \in \Omega$  que dan valores de la estadística de prueba  $t(Z, \mathbf{r})$  más grandes o iguales a  $T$ , es decir,

$$\Pr(t(Z, \mathbf{r}) \geq T) = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{|\Omega|} = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{K}$$

Como una forma de ilustrar esto de manera más clara, se retoma el caso de la Dama Catadora de Té utilizado por Fisher. La mujer catava  $N = 8$  tazas de té, todos en un solo estrato, por lo que  $S = 1$ . Cada asignación posible es una (8-ada) serie de 8 números que contiene cuatro 1's y cuatro 0's. Por ejemplo, la asignación elegida podría ser  $z = (1, 0, 0, 1, 1, 0, 0, 1)^T$  y debería significar que a las tazas 1, 4, 5 y 8 se les agregó primero la leche y a las otras cuatro tazas se les agregó primero el té. El conjunto de asignaciones de tratamiento  $\Omega$  contiene todas las posibles 8-adas que contienen cuatro 1's y cuatro 0's, entonces  $\Omega$  contiene  $|\Omega| = K = \binom{8}{4} = 70$  asignaciones distintas.

La actual asignación fue seleccionada aleatoriamente en el sentido de que  $\Pr(Z = z) = \frac{1}{K} = \frac{1}{70}$  para todas las  $z \in \Omega$ <sup>11</sup>.

La respuesta de la mujer para la taza  $i$  es  $r_i$ , donde

$$r_i = \begin{cases} 1 & \text{si responde que se agrega primero leche} \\ 0 & \text{si responde que se agrega primero te} \end{cases}$$

Entonces  $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8)^T$ . Se debe recordar que la dama debía clasificar exactamente cuatro tazas a las que se les colocó primero la leche, por lo tanto,

---

<sup>11</sup>Nótese que  $z^T \mathbf{1} = 4$  para toda  $z \in \Omega$

$\mathbf{1}^T * \mathbf{r} = 4$ . La estadística de prueba es el número de tazas correctamente identificadas y esto se escribe formalmente como

$$t(Z, \mathbf{r}) = Z^T \mathbf{r} + (\mathbf{1} - Z)^T (\mathbf{1} - \mathbf{r}) = 2Z^T \mathbf{r}$$

donde la segunda igualdad se obtiene del hecho de que  $\mathbf{1}^T * \mathbf{1} = 8$ ,  $Z^T * \mathbf{1} = 4$  y  $\mathbf{1}^T * \mathbf{r} = 4$ . Para ilustrar de una forma más concreta, se adopta el supuesto de que  $r = (1, 1, 0, 0, 0, 1, 1, 0)$ , lo que quiere decir que la dama clasificó las tazas primera, segunda, sexta y séptima como tazas en las que se vertió primero la leche. Decir que el tratamiento no tiene efecto es decir que la dama daría esta clasificación sin importar en que orden fueron agregados el té y la leche en las tazas. Si al preparar las tazas para el experimento, se cambian las tazas a las que se les agrega primero la leche y esto hace que la dama cambie sus respuestas, entonces significa que ella está discerniendo algo, y el tratamiento tiene algún efecto, sin embargo no sabemos si es ligero o errático.

Hay únicamente una asignación de tratamientos  $z \in \Omega$  que coincide con las respuestas de la dama, es decir,  $z = (1, 1, 0, 0, 0, 1, 1, 0)$ , por lo tanto, si  $t(Z, \mathbf{r}) = 8$  el nivel de significancia es

$$\Pr(t(Z, \mathbf{r}) \geq 8) = \frac{1}{70}$$

Esto quiere decir que la probabilidad de que las respuestas coincidan perfectamente por accidente es de  $1/70 = 0.014$ , lo que es una probabilidad muy pequeña. En otras palabras, si el tratamiento no tiene ningún efecto, la oportunidad de que la asignación aleatoria de los tratamientos coincida perfectamente con las ocho respuestas de la dama es de  $1/70$ .

Considerando que hay ocho respuestas, se puede dar el caso de que la asignación sólo concuerde en algunas respuestas y no precisamente en las ocho que dio la dama, pero por la forma del experimento, es imposible que existan siete respuestas concordantes, ya que si se encuentra que una de las ocho respuestas no coincide con la de la asignación esto provocará un error en otra taza, entonces existirán, necesariamente, a lo más seis respuestas concordantes. La cuestión ahora es saber cuántas asignaciones  $z \in \Omega$  pueden coincidir exactamente con seis respuestas correctas de la dama. Sabiendo que  $r = (1, 1, 0, 0, 0, 1, 1, 0)$ , una asignación con seis respuestas concordantes podría ser  $z = (1, 0, 1, 0, 0, 1, 1, 0)$ , donde la asignación no coincide con las respuestas para la segunda y tercera taza. Se observa que cualquiera de los cuatro 1's puede ser cambiado por un 0 y cada uno de los cuatro 0's puede ser cambiado por un 1, por lo tanto existen  $4*4 = 16$  asignaciones con exactamente seis respuestas concordantes. A partir de esto, se puede afirmar que son 17 las asignaciones que permiten que haya seis o más respuestas concordantes en la asignación, por lo tanto, seis respuestas concordantes brindan un

nivel de significancia igual a la  $\Pr(t(Z, \mathbf{r}) \geq 6) = (1 + 16)/70 = 17/70 = 0,24$ , lo que ya no es una probabilidad pequeña. No debería ser sorprendente ver seis o más respuestas que coincidan con la asignación si el tratamiento no tuviera ningún efecto - pasa por azar tan frecuentemente como ver dos caras cuando se lanzan dos monedas.

El punto clave merece repetición. La probabilidad entra al cálculo únicamente a través de la asignación aleatoria de los tratamientos. La distribución de probabilidad necesaria es conocida, no supuesta. El nivel de significancia resultante no depende de supuestos de ningún tipo. Si el mismo cálculo fuera hecho en un experimento no aleatorizado, se tendría que suponer que la distribución de la asignación de los tratamientos,  $\Pr(Z = z)$ , es alguna distribución en particular, tal vez se tendría que suponer la hipótesis de que las asignaciones son igualmente probables,  $\Pr(Z = z) = 1/K$ .

En un estudio no aleatorizado puede haber muy pocas bases para defender esta hipótesis y hacerlo significará abrir el caso a discusión y debate. Dicho de otra forma, el punto importante es recalcar formalmente que experimentos aleatorizados no están expuestos a ciertos problemas o dificultades a los que los estudios no aleatorizados sí están expuestos de manera legítima.

## 2.7.2. Pruebas de Aleatorización comunes

En la práctica, las pruebas de hipótesis más usadas son pruebas de aleatorización en las que el nivel de significancia puede determinarse como se mencionó anteriormente, por medio de la siguiente expresión

$$\Pr(t(Z, \mathbf{r}) \geq T) = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{|\Omega|} = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{K}$$

Pero existen otro tipo de pruebas que se derivan de una manera distinta. Por ello, a continuación, de forma sencilla, se hace una pequeña revisión de algunas de las más comunes. En todos los casos, el experimento es un experimento uniforme aleatorizado con  $\Pr(Z = z) = 1/K$  para toda  $z \in \Omega$ .

Primero se revisará la prueba exacta de Fisher. Esta prueba es para una tabla de contingencia de  $2 \times 2$  y se usa solamente como una prueba para el ejemplo de la Dama Catadora de Té. En este caso, hay un solo estrato, por lo tanto,  $S = 1$ , el resultado  $r_i$  es binario,  $r_i = 1$  ó  $r_i = 0$ , y la estadística de prueba es el número de respuestas iguales a 1 en el grupo tratado, es decir,  $t(Z, \mathbf{r}) = Z^T \mathbf{r}$ .

La tabla de contingencia de  $2 \times 2$  contiene los valores de  $z_i$  y  $r_i$  para el ejemplo de Fisher, como sigue:

		RESPUESTA $r_i$	
		1	0
ASIGNACIÓN AL GRUPO CONTROL O AL GRUPO TRATAMIENTO $Z_i$	1	$Z^T \mathbf{r}$	$\mathbf{1}^T \mathbf{Z} - Z^T \mathbf{r}$
	0	$\mathbf{1}^T \mathbf{r} - Z^T \mathbf{r}$	$Z^T \mathbf{r}$

Cuadro 2.2: Tabla de contingencia

Debido a la forma en que se diseñó el experimento, se tiene que  $\mathbf{1}^T \mathbf{r} = N/2$  y  $\mathbf{1}^T \mathbf{Z} = N/2$  por lo que la tabla queda de la siguiente forma:

		RESPUESTA $r_i$		Total
		1	0	
ASIGNACIÓN AL GRUPO CONTROL O AL GRUPO TRATAMIENTO $Z_i$	1	$Z^T \mathbf{r}$	$N/2 - Z^T \mathbf{r}$	$N/2$
	0	$N/2 - Z^T \mathbf{r}$	$Z^T \mathbf{r}$	$N/2$
Total		$N/2$	$N/2$	$N$

Cuadro 2.3: Ejemplo particular del problema de Fisher

Bajo la hipótesis de No Efecto, la distribución de aleatorización de la estadística de prueba  $Z^T \mathbf{r}$  es la distribución hipergeométrica.

El análogo de la prueba exacta de Fisher cuando hay dos o más estratos,  $S \geq 2$ , pero las respuestas individuales  $r_{si}$  siguen siendo binarias es la prueba de Mantel-Haenszel, pero en este caso, los datos deben estar registrados en una tabla de contingencia de  $2 \times 2 \times S$ , de acuerdo al vector de asignación de tratamientos,  $Z$ , al vector respuesta,  $\mathbf{r}$ , y al estrato  $s$  correspondiente.

Por la forma de proceder análoga a la prueba exacta de Fisher, la estadística de prueba es, nuevamente, el número de respuestas 1 existentes en las unidades del grupo tratado, por lo que la estadística queda  $t(Z, \mathbf{r}) = Z^T \mathbf{r} = \sum_{s=1}^S \sum_{i=1}^{n_s} z_{si} r_{si}$ . Bajo la hipótesis nula y la contribución del estrato  $s$ ,  $\sum_{i=1}^{n_s} z_{si} r_{si}$  tiene nuevamente una distribución hipergeométrica y, en este caso,  $t(Z, \mathbf{r})$  tiene una distribución igual a la distribución de la suma de  $S$  variables hipergeométricas independientes.

Un atractivo técnico de esta estadística de prueba es que la aproximación para muestras grandes funciona bien para una tabla de  $2 \times 2 \times S$  con  $N$  grande aún cuando  $S$  también sea grande y esto provoque que haya pocos sujetos en cada uno de los  $S$  estratos.

La estadística de Mantel-Haenszel es usada extensivamente en epidemiología además de otros campos, en particular, es ampliamente usada en experimentos controlados múltiples en los cuales  $m_s = 1$  para cada  $s$ .

Un caso especial de la prueba de Mantel-Haenszel se aplica cuando se trata de un experimento aleatorizado pareado, ya que algunos cálculos se simplifican, en este caso especial la prueba es también conocida como prueba de McNemar.

Desarrollando estos métodos para tablas de  $2 \times 2 \times S$  en una forma distinta, Birch y Cox muestran que estas tres pruebas con respuestas binarias poseen una propiedad de optimalidad, por lo tanto, tiene sentido afirmar que la Prueba Exacta de Fisher, la Prueba de Mantel-Haenszel y la Prueba de McNemar son las mejores pruebas para los problemas que atiende cada una de éstas. Específicamente ellos muestran que la estadística de prueba  $t(Z, \mathbf{r}) = Z^T \mathbf{r}$  junto con el nivel de significancia son la prueba uniformemente más potente contra alternativas definidas en términos de cocientes de probabilidad.

Una extensión de la prueba de Mantel-Haenszel conocida simplemente como la extensión de Mantel, es una prueba en la que las respuestas  $r_{si}$  son reducidas, en cantidad, a un número menor de valores. Estos nuevos valores son puntuaciones numéricas que involucran el valor de todas las  $r_{si}$  iniciales, pues cada valor representa una de varias categorías en las que se han introducido las  $r_{si}$ .

A continuación se presenta un ejemplo de este tipo de experimentos. La New York Heart Association clasifica a cada paciente con problemas coronarios en una de cuatro categorías basándose en el grado de limitación en la actividad física que le provoca la sintomatología de la enfermedad coronaria como el dolor en el pecho, las cuatro categorías son:

1. Sin limitaciones para la actividad física
2. Ligera limitación, confortable en el descanso pero al momento de la actividad física se presenta dolor o malestar
3. Limitación marcada, con la menor actividad física se presentan los síntomas y molestias; y
4. Completa limitación, no hay habilidad para la actividad física sin dolor ni malestar, éstos incluso se presentan en el descanso.

Para este experimento, el vector de resultados,  $\mathbf{r}$ , está formado por las respuestas individuales,  $r_{si}$ , es decir, por los valores 1, 2, 3 y/o 4, ya que son estos los valores que puede tomar cada  $r_{si}$ , según la categoría en la que caiga el individuo  $i$  en el estrato  $s$ . Además, los datos deberían registrarse en una tabla de contingencia de  $2 \times 4 \times s$



correspondiente a la tabla  $Z \times r \times S$ . La estadística de prueba de Mantel es la suma de los valores asignados a las  $r_{si}$  de las unidades tratadas, es decir, la suma de las puntuaciones respuesta de los pacientes en el grupo tratado, por lo tanto,  $t(Z, \mathbf{r}) = Z^T \mathbf{r}$ .

En el caso de un solo estrato,  $S = 1$ , la prueba de Suma de rangos de Wilcoxon se usa comúnmente para analizar resultados ante la existencia de muchas categorías o rangos.

En esta prueba, las respuestas son ordenadas, según la puntuación, de menor a mayor. Si las  $N$  respuestas fueran números diferentes, los rangos deberían ser los números  $1, 2, \dots, N$ . Si algunas respuestas fueran iguales, entonces se usa el promedio de sus rangos preliminares para determinar su rango definitivo. Denotando por  $q_i$  el rango de  $r_i$ , entonces el vector  $q = (q_1, \dots, q_N)^T$  es el vector de rangos de las respuestas. Por ejemplo, si  $N = 4$ , y se tiene  $\mathbf{r} = (r_1, r_2, r_3, r_4)^T = (2,3 \ 1,1 \ 2,3 \ 7,9)^T$  entonces al ordenar las respuestas se obtiene:

$r_i$ ordenadas	Valores $r_i$ ordenados	Rango preliminar	Rango definitivo
$r_2$	1.1	1	$q_2 = 1$
$r_1$	2.3	2	$q_1 = 2.5$
$r_3$	2.3	3	$q_3 = 2.5$
$r_4$	7.9	4	$q_4 = 4$

Cuadro 2.4: Ejemplo de la determinación de los rangos en un caso particular

Por lo tanto,  $q = (2,5 \ 1 \ 2,5 \ 4)^T$ , todo esto partiendo de que  $r_2$  es la respuesta con la menor puntuación,  $r_4$  la respuesta con la mayor puntuación, y  $r_1$  y  $r_3$ , al ser iguales, comparten los rangos 2 y 3, por lo cual se les asigna como rango definitivo el  $(2 + 3)/2 = 2,5$ .

Se puede notar que los rangos  $q$  están en función de las respuestas  $\mathbf{r}$ , las cuales son valores fijos, por lo tanto,  $q$  también es un vector fijo.

La estadística de la suma de rangos es la suma de los rangos de las unidades tratadas,  $t(Z, \mathbf{r}) = Z^T \mathbf{q}$ , y su nivel de significancia es determinado por

$$\Pr(t(Z, \mathbf{r}) \geq T) = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{|\Omega|} = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{K}$$

En el caso de  $S$  pares con  $n_s = 2$  y  $m_s = 1$  para  $s = 1, 2, \dots, S$ , es comúnmente usada la prueba de rangos con signos de Wilcoxon. Así se tiene,  $(Z_{s1}, Z_{s2}) = (1, 0)$  si

la primera unidad del par  $s$  recibe el tratamiento o  $(Z_{s_1}, Z_{s_2}) = (0, 1)$  si es la segunda unidad la que recibe el tratamiento. En esta prueba también se asignan rangos pero no a cada una de las respuestas, sino a cada una de las diferencias absolutas del par de respuestas. Entonces a las diferencias  $|r_{s_1} - r_{s_2}|$  se les asignan los rangos del 1 al  $S$ , con los rangos promedio usados para los valores que sean iguales. Se denota con  $d_s$ , el rango que  $|r_{s_1} - r_{s_2}|$  obtuvo. La estadística de rangos con signos es la suma de los rangos de los pares en los que la unidad tratada tiene una respuesta más alta que la unidad control. Formalmente se tiene,  $C_{s_1} = 1$  si  $r_{s_1} > r_{s_2}$  y  $C_{s_1} = 0$  en cualquier otro caso, y de manera similar,  $C_{s_2} = 1$  si  $r_{s_2} > r_{s_1}$  y  $C_{s_2} = 0$  en cualquier otro caso, por lo tanto,  $C_{s_1} = C_{s_2} = 0$  si  $r_{s_1} = r_{s_2}$ .

Entonces,  $Z_{s_1}C_{s_1} + Z_{s_2}C_{s_2}$  será igual a 1 si la unidad tratada en el par  $s$  tiene una respuesta más alta que la unidad control y será 0 en cualquier otro caso. De esto se sigue que la estadística de rangos con signo es  $\sum_{s=1}^S (d_s \sum_{i=1}^2 C_{s_i} Z_{s_i})$ . Se puede notar que  $d_s$  y  $C_{s_i}$  son funciones de  $\mathbf{r}$  y, por lo tanto, son valores fijos bajo la hipótesis nula de No efecto del tratamiento. Además si  $r_{s_1} = r_{s_2}$ , entonces el par  $s$  contribuye cero al valor de la estadística sin importar como son asignados los tratamientos.

Para respuestas estratificadas, un método, que es usado algunas veces, involucra el cálculo de estadísticas de suma de rangos de forma separada en cada uno de los  $S$  estratos, y toma la suma de estas  $S$  estadísticas de suma de rangos como la estadística de prueba. Esto es, la estadística de suma de rangos estratificados es de la forma  $t(Z, \mathbf{r}) = Z^T \mathbf{q}$  de manera similar a la estadística de suma de rangos; sin embargo, los rangos en  $q$  ya no son una permutación de los números  $1, 2, \dots, N$ , sino más bien de los números  $1, \dots, n_1, 1, \dots, n_2, \dots, 1, \dots, n_s$ , con los ajustes necesarios en cada relación. En adición,  $\Omega$  cambió.

Hodges y Lehmann encontraron que la estadística de suma de rangos estratificados es ineficiente cuando  $S$  es muy grande al compararlo con  $N$ . En particular, para datos pareados donde  $S = N/2$ , la prueba de rango estratificado es equivalente a la prueba de los Signos, la cual es, sustancialmente, menos eficiente que la prueba de rangos con signo para datos de distribuciones de cola corta tales como la Normal. Ellos sugieren como un método alternativo, el método de rangos alineados: se obtiene en cada estrato, la media de las respuestas, creando respuestas alineadas que son asignadas a los rangos del 1 al  $N$ , ignorando por un momento el estrato. Escribiendo  $q$  para estos rangos alineados, la estadística de rangos alineados es la suma de los rangos alineados en el grupo tratado,  $t(Z, \mathbf{r}) = Z^T \mathbf{q}$ .

Otra estadística importante es la empleada en la Prueba de la Mediana. Asignando  $C_{s_i} = 1$  si  $r_{s_i}$  es más grande que la mediana de las respuestas del estrato  $s$  y  $C_{s_i} = 0$  en cualquier otro caso, se obtiene el vector  $C$ , que es la  $N$ -ada que contiene todos los  $c_{s_i}$ . Entonces  $t(Z, \mathbf{r}) = Z^T C$  es el número de respuestas en unidades tratadas que exceden a la mediana de su estrato. Con un solo estrato,  $S = 1$ , la prueba de la mediana es bastante buena en muestras grandes, si las respuestas tienen una distribución exponencial doble,

una distribución con una cola más pesada que la normal. Algunas veces se hacen pruebas análogas en las que la mediana es reemplazada por algún otro cuantil o por otras medidas de localización.

Al comenzar a trabajar con una estadística  $t(Z, \mathbf{r})$ , su distribución de aleatorización puede ser determinada por

$$\Pr(t(Z, \mathbf{r}) \geq T) = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{|\Omega|} = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{K}$$

Esto es cierto incluso con estadísticas que son comúnmente referidas a distribuciones teóricas en lugar de, por ejemplo, la Prueba de Dos Muestras o la Prueba  $t$ -pareada, entre otras. Welch y Wilk estudiaron la relación entre la distribución de aleatorización y la distribución teórica de estadísticas que fueron inicialmente derivados de supuestos de normalidad e independencia en la distribución de las respuestas. Ellos sugieren que la distribución teórica puede ser vista como una aproximación computacionalmente conveniente a la deseada, pero una distribución de aleatorización computacionalmente difícil. Es decir, sugieren que las pruebas  $t$ , así como las pruebas de rangos o las pruebas de Mantel-Haenszel, pueden ser justificadas solamente con base en el uso de la aleatorización en el diseño de un experimento, sin referencia a la normal. Estos resultados dependen del buen comportamiento de los momentos de las sumas de los cuadrados de las respuestas sobre la distribución de aleatorización; por lo tanto, éstas dependen de la ausencia de respuestas extremas. Estos resultados son importantes como una relación conceptual entre la teoría de la Normal y la inferencia de la aleatorización.

### 2.7.3. Clases de Estadísticas de Prueba

La semejanza existente entre las estadísticas de prueba comúnmente usadas (como las tratadas con anterioridad) es sorprendente pero no accidental. Grandes clases de estadísticas comparten las propiedades importantes de los métodos, por lo que es más simple y menos repetitivo discutir las clases de estadísticas y no método por método.

Aunque las estadísticas que ya se trataron fueron desarrolladas por distintas personas en distintos tiempos y con distintos propósitos, existe semejanza entre ellas por una buena razón. Conforme el tamaño de muestra,  $N$ , aumenta, el número  $K$  de asignaciones de tratamientos contenidas en  $\Omega$  crece muy rápidamente, y el cálculo directo de  $\Pr(t(Z, \mathbf{r}) \geq T)$  se hace muy complicado aún con la computadora más rápida. Para verificar tómese un caso simple con  $S = 1$  y con una división equitativa de los  $N$  sujetos entre los grupos tratado y control, es decir,  $m = N/2$  sujetos tratados y  $N - m = N/2$  sujetos control. Esto significa que hay  $K = \binom{N}{N/2}$  asignaciones de tratamiento distintas en  $\Omega$ . Si se agregara una unidad más al experimento, se incrementaría el tamaño de

muestra a  $N + 1$  y, por lo tanto, se puede mostrar fácilmente que  $K$  se incrementaría multiplicándose por un factor de  $(N + 1)/(N/2 + 1)$ , que significa, casi duplicar el valor de  $K$ , aún más, cuando  $N$  es más grande. Rigurosamente hablando, suponiendo que la computadora más rápida puede calcular  $\Pr(t(Z, \mathbf{r}) \geq T)$  directamente, para a lo más, una muestra de tamaño  $N$ , es decir, puede calcular los  $2^N$  elementos contenidos en  $\Omega$ , si se agregaran diez elementos más a la muestra, el número de elementos que tendría que calcular sería  $2^{N+10} = 2^N 2^{10}$ , entonces esto haría  $2^{10} = 1024$  veces más grande el número de elementos a calcular, lo cual resulta poco práctico para  $N$ 's grandes.

La solución usual a este problema es aproximar  $\Pr(t(Z, \mathbf{r}) \geq T)$  usando una muestra grande o una aproximación asintótica. Las aproximaciones más comunes usan los momentos de la estadística de prueba, su esperanza y varianza, y algunas veces momentos más altos. En el caso de las estadísticas de prueba anteriores, es fácil la obtención de los momentos necesarios.

La primera clase de estadísticas de prueba que se tratará es llamada la clase de las Estadísticas Suma y son de la forma  $t(Z, \mathbf{r}) = Z^T \mathbf{q}$  donde  $q$  es alguna función de  $\mathbf{r}$ . Una Estadística Suma, suma las puntuaciones  $q_{si}$  para las unidades del grupo tratado, entonces se puede observar que las estadísticas vistas anteriormente son Estadísticas Suma, donde  $q$  es simplemente  $\mathbf{r}$  (para las pruebas de Fisher, Mantel- Haenszel y McNemar), contiene los rangos de  $\mathbf{r}$  (en la prueba de suma de rangos de Wilcoxon) es el vector de unos y ceros identificando respuestas  $r_{si}$  que exceden el estrato de las medianas (en la prueba de la mediana) ó  $q_{si} = d_s c_{si}$  (en la estadística de rangos con signo).

Existen fórmulas simples para la obtención de los momentos de las Estadísticas Suma bajo la hipótesis nula de que el tratamiento no tiene efecto. En este caso se fija  $\mathbf{r}$ , por lo tanto,  $q$  queda fijo. Las fórmulas de momentos usan las propiedades del muestreo aleatorio simple sin reemplazo. Formar sin reemplazo una muestra aleatoria simple de tamaño  $m$  de una población de tamaño  $n$  es formar un subconjunto aleatorio de  $m$  elementos de un conjunto con  $n$  elementos donde cada uno de los  $\binom{n}{m}$  subconjuntos de tamaño  $m$  tiene la misma probabilidad  $\frac{1}{\binom{n}{m}}$ . Cochran discute el muestreo aleatorio simple sin reemplazo. En un experimento uniforme aleatorizado, las  $m_s$  unidades tratadas en el estrato son una muestra aleatoria simple sin reemplazo de las  $n_s$  unidades en el estrato  $s$ .

**Proposición 2.** *En un experimento uniforme aleatorizado, si el tratamiento no tiene efecto, la esperanza y la varianza de la Estadística Suma  $Z^T q$  son*

$$E(Z^T q) = \sum_{s=1}^S m_s \underline{q}_s \quad y$$

$$Var(Z^T q) = \sum_{s=1}^S \frac{m_s (n_s - m_s)}{n_s (n_s - 1)} \sum_{i=1}^{n_s} (q_{si} - \underline{q}_s)^2$$

donde  $\underline{q}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} q_{si}$ .

Los momentos son fáciles de determinar para las Estadísticas Suma, pero otras clases de estadísticas tienen otras propiedades usuales.

Dentro de las Estadísticas Suma encontramos el subconjunto de las Estadísticas Sign-Score (estadísticas de puntuaciones con signo) que son de la forma

$$t(Z, \mathbf{r}) = \sum_{s=1}^S d_s \sum_{i=1}^{n_s} C_{si} Z_{si}$$

donde  $c_{si}$  es binaria,  $c_{si} = 1$  ó  $c_{si} = 0$ , tanto  $d_s$  como  $c_{si}$  son funciones de  $\mathbf{r}$ . Las estadísticas de la prueba exacta de Fisher, la prueba de Mantel-Haenszel y la prueba de McNemar son Estadísticas Sign-Score con  $d_s = 1$  y  $c_{si} = r_{si}$ .

Una estadística Sign-Score es una Estadística Suma con  $q_{si} = d_s c_{si}$ , pero muchas Estadísticas Suma, incluyendo la estadística de la suma de rangos de Wilcoxon no son Estadísticas Sign-Score. Algunas veces es más fácil realizar cálculos con las estadísticas Sign-Score y algunas otras es con estadísticas Suma, por ello se hace la distinción.

Otra clase importante de estadísticas es la de arrangement-increasing functions (funciones de arreglos crecientes) de  $Z$  y de  $\mathbf{r}$ . De manera informal, una estadística es arrangement-increasing si incrementa su valor conforme las coordenadas de  $Z$  y  $\mathbf{r}$  se reorganizan en incrementos del mismo orden dentro de cada estrato. De hecho, las estadísticas revisadas con anterioridad son arrangement-increasing, por lo tanto, cualquier concepto aplicable a las funciones arrangement-increasing se les puede aplicar también.

Ahora se introducen algunos términos y convenciones que serán usuales. Los valores  $S$  y  $n_s$ ,  $s = 1, \dots, S$  con  $N = \sum_{s=1}^S n_s$  son dados. Una  $N$ -ada estratificada  $a$ , es una  $N$ -ada en la que las  $N$  coordenadas son divididas en  $S$  estratos con  $n_s$  coordenadas en el estrato  $s$ , donde  $a_{si}$  es la  $i$ -ésima coordenada de los  $n_s$  en el estrato  $s$ . Por ejemplo,  $Z$  y  $\mathbf{r}$  son cada una,  $N$ -adas estratificadas. Si  $a$  es una  $N$ -ada estratificada, y si  $i$  y  $j$  son dos enteros diferentes menores o iguales a  $n_s$ , entonces se denotará como  $a_{sij}$  a la  $N$ -ada estratificada formada por  $a$  pero intercambiando las posiciones de las coordenadas  $a_{si}$  y  $a_{sj}$ . Para evitar la repetición, siempre que el símbolo  $a_{sij}$  aparece, se supone sin que exista la mención, que los subíndices son adecuados, por lo tanto,  $s$  es un entero positivo entre 1 y  $S$  e  $i$  y  $j$  son dos enteros positivos diferentes menores o iguales que  $n_s$ . Una función  $f(a, b)$  de dos  $N$ -adas estratificadas es invariante si  $f(a, b) = f(a_{sij}, b_{sij})$  para toda  $s$ ,  $i$  y  $j$ , por lo tanto, renumerar las unidades en el mismo estrato no cambia el valor de  $f(a, b)$ . Por ejemplo, la función  $Z^T q$  es una función invariante de  $Z$  y  $q$ .

Definición: Una función invariante  $f(a, b)$  de dos  $N$ -adas estratificadas es arrangement-increasing (o AI) si  $f(a, b_{sij}) \geq f(a, b)$  siempre que  $(a_{si} - a_{sj})(b_{si} - b_{sj}) \leq 0$ .

Se puede notar que considerando la  $i$ -ésima y la  $j$ -ésima unidades en el estrato  $s$ ,  $(a_{si} - a_{sj})(b_{si} - b_{sj}) < 0$ , entonces de estas dos unidades, aquella que tiene el valor más grande estando en  $a$  es la misma unidad que tiene el valor más chico en  $b$ , por lo tanto, estas dos coordenadas están fuera de orden. Sin embargo, en  $a$  y  $b_{sij}$ , estas dos coordenadas están en el mismo orden, ya que  $b_{si}$  y  $b_{sj}$  se intercambiaron. La definición dice que una función arrangement-increasing será más grande, o al menos no menor, cuando estas dos coordenadas son conmutadas para dejarlas dentro de un mismo orden.

Nótese además que la definición dice que cuando  $(a_{si} - a_{sj})(b_{si} - b_{sj}) = 0$ , entonces,  $a_{si} = a_{sj}$ ,  $b_{si} = b_{sj}$  o ambas. En este caso  $f(a, b_{sij}) = f(a, b)$ .

Considerando como ejemplo, se puede decir que la  $Z^T q$  es arrangement-increasing como función de  $Z$  y  $q$ . Se tiene que,  $Z^T q_{sij} - Z^T q = (z_{si}q_{sj} + z_{sj}q_{si}) - (z_{si}q_{si} + z_{sj}q_{sj}) = -(z_{si} - z_{sj})(q_{si} - q_{sj})$ , por lo tanto, si  $(z_{si} - z_{sj})(q_{si} - q_{sj}) \leq 0$ , entonces  $Z^T q_{sij} - Z^T q \geq 0$ . Esto muestra como  $Z^T q$  es arrangement-increasing.

En la siguiente tabla se encuentra una pequeña ilustración de una estadística arrangement-increasing. Se trata de una estadística de suma de rangos con un solo estrato,  $S = 1$ ,  $N = 4$ , de las cuales,  $m = 2$  reciben el tratamiento:

<b>i</b>		<b>z</b>	<b>Q</b>	<b>q<sub>23</sub></b>
<b>1</b>	Tratado	1	4	4
<b>2</b>	Tratado	1	2	3
<b>3</b>	Control	0	3	2
<b>4</b>	Control	0	1	1
<b>Estadística de Suma de rangos</b>			6	7

Cuadro 2.5: Ejemplo de estadística de suma de rangos

Aquí  $(z_2 - z_3)(q_2 - q_3) = (1 - 0)(2 - 3) = -1 \leq 0$ , y la estadística de suma de rangos es  $Z^T q = 6$  se incrementa a  $Z^T q_{23} = 7$  por el intercambio de  $q_2$  y  $q_3$ . Como un segundo ejemplo, se considera la función  $t(Z, \mathbf{r}) = Z^T q$  cuando  $q$  es una función de  $\mathbf{r}$ , la cual puede ser escrita explícitamente como  $q(\mathbf{r})$ . Entonces,  $t(Z, \mathbf{r})$  puede o no, ser una función arrangement-increasing en  $Z$  y  $\mathbf{r}$  dependiendo de cómo varía  $q(\mathbf{r})$  con  $\mathbf{r}$ .

Todas las estadísticas usuales que se han revisado tienen las siguientes dos propiedades:

- Se permuta  $\mathbf{r}$  dentro del estrato y  $q$  es permutado de la misma forma, y

- Dentro de cada estrato, al  $r_{si}$  más grande le corresponde el  $q_{si}$  más grande.

Se puede notar que  $t(Z, \mathbf{r}) = Z^T q$  es arrangement-increasing si  $q(r)$  tiene estas dos propiedades, porque con la primera propiedad se asegura que  $t(Z, \mathbf{r})$  es invariante, y la segunda asegura que  $r_{si} - r_{sj} \geq 0$  implica  $q_{si} - q_{sj} \geq 0$ , por lo tanto,  $(z_{si} - z_{sj})(r_{si} - r_{sj}) \leq 0$  implica que  $(z_{si} - z_{sj})(q_{si} - q_{sj}) \leq 0$  y esto concuerda con lo que se ha mencionado en párrafos anteriores. La conclusión importante es que todas las estadísticas revisadas son arrangement-increasing.

Para describir el comportamiento de una estadística al adoptar una hipótesis nula de que el tratamiento no tiene efecto, existe una última clase de estadísticas muy usual, la clase de estadísticas order-preserving (estadísticas que preservan el orden). Muchas estadísticas que miden la diferencia entre los pacientes del grupo tratado y los del grupo control deberían tender a incrementarse en valor si las respuestas en el grupo tratado se incrementaran y las del grupo control disminuyeran. Expresando esto formalmente, una unidad tratada tiene  $2z_{si} - 1 = 1$  ya que  $z_{si} = 1$  y una unidad control tiene  $2z_{si} - 1 = -1$  ya que  $z_{si} = 0$ . Sea  $z \in \Omega$  una posible asignación de tratamientos y sean  $\mathbf{r}$  y  $\mathbf{r}^*$  las dos posibles  $N$ -adas de respuestas tales que  $(r_{si}^* - r_{si})(2z_{si} - 1) \geq 0$  para toda  $s, i$ . Con la asignación de tratamientos dada por  $z$ , se tiene que  $r_{si}^* \geq r_{si}$  para cada unidad tratada y  $r_{si}^* \leq r_{si}$  para cada unidad control. Esto es, si las respuestas más altas indican resultados favorables, entonces cada unidad tratada presenta mejores resultados con  $\mathbf{r}^*$  que con  $\mathbf{r}$  y cada unidad control muestra peores resultados con  $\mathbf{r}^*$  que con  $\mathbf{r}$ . Es decir, la diferencia entre las unidades de los grupos tratado y control parece ser mayor con  $\mathbf{r}^*$  que con  $\mathbf{r}$ . Una estadística de prueba es order-preserving si  $t(z, \mathbf{r}) \leq t(z, \mathbf{r}^*)$  siempre que  $\mathbf{r}^*$  y  $\mathbf{r}$  sean dos posibles valores de las respuestas tales que  $(r_{si}^* - r_{si})(2z_{si} - 1) \geq 0$  para toda  $s, i$ .

A continuación se presenta un ejemplo muy sencillo e hipotético para ilustrar la idea de estadísticas order-preserving:

Se supone un solo estrato,  $S = 1$ , cuatro sujetos,  $N = 4$ , de los cuales  $m = 2$  reciben el tratamiento.

$i$		$Z_i$	$2z_i - 1$	$r_i$	$r_i^*$	Rangos de $r_i$	Rangos de $r_i^*$
1	Tratado	1	1	5	6	4	4
2	Tratado	1	1	2	4	2	3
3	Control	0	-1	3	2	3	2
4	Control	0	-1	1	1	1	1
<b>Suma de Rangos de Wilcoxon</b>						6	7

Cuadro 2.6: Ejemplo de Rangos de Wilcoxon

Nótese que cuando se comparan  $r_i$  y  $r_{i^*}$ , los sujetos tratados tienen  $r_i \leq r_{i^*}$  mientras los sujetos control tienen  $r_{i^*} \leq r_i$ . Si las respuestas son asignadas a rangos del 1 al 4, y se suman los rangos de las unidades tratadas para obtener la estadística de prueba de suma de rangos de Wilcoxon, entonces la suma de rangos de  $\mathbf{r}^*$  es mayor que la de  $\mathbf{r}$ .

En resumen, se han considerado cuatro clases de estadísticas de prueba:

- Estadísticas Suma
- Estadísticas arrangement-increasing
- Estadísticas order-preserving; y
- Estadísticas sign-score

Las estadísticas que se revisaron anteriormente son miembros de las primeras tres clases y la mayoría son miembros de la cuarta clase a excepción de la estadística de suma de rangos de Wilcoxon, la estadística de suma de rangos estratificados y la estadística de la extensión de Mantel.

## 2.8. Modelos para Efecto del Tratamiento

### 2.8.1. Respuestas cuando el tratamiento tiene un efecto

Si el tratamiento tiene efecto, entonces la  $N$ -ada de respuestas observadas para las  $N$  unidades será diferente para asignaciones de tratamiento,  $Z$ , diferentes. En la sección anterior, al suponer la hipótesis nula de no efecto del tratamiento, se observaba que las respuestas eran fijas, ya que no variaban conforme variaba la asignación del tratamiento,  $Z$ , y el vector de respuestas era  $\mathbf{r}$ . Cuando no se tiene la hipótesis de No efecto del tratamiento, las respuestas varían conforme  $Z$  varía, por lo tanto, para cada  $z$  existe un vector de respuestas que será denotado como  $\mathbf{r}_z$ . Nótese que para cada  $z \in \Omega$ , la respuesta  $\mathbf{r}_z$  es una  $N$ -ada no aleatoria - todavía no se incluye la probabilidad. Se escribirá  $r_{siz}$  para la coordenada  $(s, i)$  de  $\mathbf{r}_z$ , esto es, la respuesta de la  $i$ -ésima unidad en el estrato  $s$  cuando las  $N$  unidades reciben la asignación de tratamientos  $z$ .

Comenzar con esto, implica retomar, por un momento, el problema abordado por Fisher de la Dama Catadora de Té. Si la Dama no pudiera discriminar del todo, entonces no importaría en qué orden son vertidos en la taza el té y la leche - esto es, no importa que  $z$  se elija- ella clasificará las tazas de la misma forma -es decir, ella dará el mismo vector de respuestas  $\mathbf{r}$ . Por otro lado, si ella discrimina perfectamente, siempre clasificará las tazas de manera correcta, variando su vector de respuestas  $\mathbf{r}$  conforme



varía la asignación  $z$  elegida, es más, las respuestas coincidirán perfectamente con la asignación de tratamientos correspondientes, es decir,  $\mathbf{r}_z = z$ .

Si los tratamientos son asignados aleatoriamente, entonces la asignación de tratamientos  $Z$  es una variable aleatoria, por lo tanto, las respuestas observadas son además variables aleatorias ya que dependen de  $Z$ . Específicamente, la respuesta observada es la variable aleatoria  $\mathbf{r}_z$ ,  $z \in \Omega$ , seleccionada al elegir una asignación  $z$  por el mecanismo aleatorio aplicado en el experimento. Se escribe  $R = \mathbf{r}_z$  para las respuestas observadas, donde tanto  $R$  como  $Z$  son variables aleatorias.

En principio, cada posible asignación de tratamientos  $z \in \Omega$  podría producir un patrón de respuestas  $\mathbf{r}_z$  que no tendría ninguna relación con el patrón observado con otra asignación  $z$ . Por ejemplo, en un experimento completamente aleatorizado con 50 unidades divididas en dos grupos de 25, debería haber  $|\Omega| = \binom{50}{25} = 1,3 \times 10^{14}$  50-adas de  $\mathbf{r}_z$  diferentes y ajenas. Considerando esto, es difícil entender un efecto del tratamiento en tales términos. Se observan por regularidad, patrones o modelos del comportamiento de  $\mathbf{r}_z$  conforme  $z$  varía sobre  $\Omega$ . A continuación se discuten algunos modelos para  $\mathbf{r}_z$  conforme  $z$  varía sobre  $\Omega$ .

## 2.8.2. Modelo de No Interferencia entre unidades

Un primer modelo a considerar es el modelo que adopta el supuesto de No Interferencia entre unidades, lo cual significa que la observación obtenida sobre una unidad no debería verse afectada por la asignación particular de los tratamientos a las otras unidades (Cox, 1958).

Por sus siglas en inglés, Rubin (1986) llama SUTVA a este supuesto al que denomina supuesto del valor estable de la unidad de tratamiento (stable unit treatment value assumption). De manera formal se puede decir que  $r_{zsi}$  varía conforme varía  $z_{si}$  sin importar la variación de las otras coordenadas de  $z$ . Es decir, la respuesta que tenga la unidad  $i$ -ésima del estrato  $s$  sólo depende del tratamiento asignado a esa unidad pero no de los tratamientos asignados a las otras unidades, por lo tanto, esta unidad tiene únicamente dos posibles valores en lugar de  $|\Omega|$ . Cuando se adopta el supuesto SUTVA, es decir, se aplica este modelo, se denotará  $r_{Tsi}$  y  $r_{Csi}$  a las respuestas de la  $i$ -ésima unidad del estrato  $s$  cuando es asignada al grupo tratamiento o al grupo control respectivamente, entonces  $r_{Tsi}$  es el valor común de  $r_{siz}$  para toda  $z \in \Omega$  con  $z_{si} = 1$ , y  $r_{Csi}$  es el valor común de  $r_{siz}$  para toda  $z \in \Omega$  con  $z_{si} = 0$ . Por lo tanto, la respuesta observada en la  $i$ -ésima unidad en el estrato  $s$  es  $R_{si} = r_{Tsi}$  si  $z_{si} = 1$  y  $R_{si} = r_{Csi}$  si  $z_{si} = 0$ , o de manera más formal

$$R_{si} = z_{si}r_{Tsi} + (1 - z_{si})r_{Csi}$$

Cuando no hay interferencia entre unidades, se denotará con  $r_T$  a  $(r_{T11}, \dots, r_{TSn_s})^T$  y con  $r_C$  a  $(r_{C11}, \dots, r_{CSn_s})^T$ .

La No Interferencia entre las unidades es sólo un modelo, que no implica que sea ajustable a todo tipo de experimentos. Existen experimentos en los cuales, el diseño o la metodología no permiten que se cumpla con el supuesto establecido, tal es el caso de la Dama Catadora de Té y las ocho tazas. La Dama debe clasificar las ocho tazas en dos grupos de cuatro, por lo tanto, cambiar la clasificación de una taza implicaría un cambio forzoso en la clasificación de otra. En el caso en el que el tratamiento no tiene efecto, habrá interferencia entre las ocho tazas.

En ensayos en los que las unidades son personas diferentes y el tratamiento es una intervención médica, el modelo parece ser suficientemente razonable. La interferencia surge cuando las unidades involucran múltiples mediciones sobre una misma persona en tiempos distintos bajo tratamientos diferentes.

### 2.8.3. Modelo de un Efecto Aditivo y modelos relacionados

El modelo para un tratamiento con un efecto aditivo supone que las unidades no interfieren unas con otras, y que la administración del tratamiento incrementa la respuesta de una unidad en una cantidad constante  $\lambda$ , tal que  $r_{Tsi} = r_{Csi} + \lambda$  para cada  $s, i$ . el principal atractivo de este modelo es que existe un parámetro definitivo para estimar el llamado efecto aditivo del tratamiento,  $\lambda$ .

Bajo el modelo de efecto aditivo, la respuesta observada en la  $i$ -ésima unidad en el estrato  $s$  es  $R_{si} = r_{Csi} + \lambda z_{si}$  o  $R = r_C + \lambda z$ . De esto, se sigue que las respuestas ajustadas,  $R - \lambda z = r_C$  son fijas, pues no varían con la asignación de tratamientos  $Z$ , por lo tanto, las respuestas ajustadas satisfacen la hipótesis nula de No Efecto. Este hecho será usado para poder hacer inferencia acerca de  $\lambda$ .

Hay varios modelos similares, incluyendo el modelo del efecto multiplicativo,  $r_{Tsi} = \eta r_{Csi}$ , el modelo de un efecto lineal,  $r_{Tsi} = \eta r_{Csi} + \lambda$  con un parámetro bidimensional  $(\lambda, \eta)$ , y modelos con parámetros ligados a la magnitud del efecto  $r_{Tsi} - r_{Csi}$  para los valores de las covariables.

### 2.8.4. Efectos positivos y efectos mayores

El modelo de un efecto aditivo supone un gran acuerdo sobre la relación que existe entre  $r_{Tsi}$  y  $r_{Csi}$ . A veces es deseable describir el comportamiento de los procedimientos estadísticos sin hacer muchos supuestos para idealizar la situación. Cuando no hay interferencia entre los tratamientos, un efecto es un par  $(r_T, r_C)$  dando las respuestas de cada unidad bajo cada tratamiento. Existen dos conceptos que son muy útiles:

efecto positivo y efectos mayores. A diferencia del modelo de un efecto aditivo del tratamiento, los modelos de efectos positivos y mayores son más significativos no sólo para respuestas continuas, sino incluso para respuestas binarias, respuestas ordinales, respuestas censuradas y respuestas multivariadas.

Se dice que un tratamiento tiene un efecto positivo si  $r_{Tsi} \geq r_{Csi}$  para todas las unidades  $(s, i)$  con la desigualdad estricta para al menos una unidad. Una forma más compacta de escribir esto es:  $(r_T, r_C)$  es un efecto positivo si  $r_T \geq r_C$  con  $r_T \neq r_C$ . Es decir, la aplicación del tratamiento nunca disminuye la respuesta de la unidad inclusive en algunos casos la incrementa. Por ejemplo, hay un efecto positivo si el efecto es aditivo y  $\lambda > 0$ .

Considerando dos posibles efectos,  $(r_T, r_C)$  y  $(r_{T*}, r_{C*})$ . Se tiene que  $(r_{T*}, r_{C*})$  es un efecto mayor que  $(r_T, r_C)$  si  $r_{Tsi*} \geq r_{Tsi}$  y  $r_{Csi*} \leq r_{Csi}$  para toda  $s, i$ . El ejemplo más sencillo que se puede ilustrar es el que ocurre cuando el efecto del tratamiento es aditivo,  $r_{C*} = r_C$ ,  $r_T = r_C + \lambda 1$ , y  $r_{T*} = r_C * + \lambda * 1$ , en este caso  $(r_{T*}, r_{C*})$  muestra un efecto mayor que  $(r_T, r_C)$  si  $\lambda * > \lambda$ . Se denotará por  $R$  y  $R^*$  a los vectores de respuestas observadas de los efectos  $(r_T, r_C)$  y  $(r_{T*}, r_{C*})$ , respectivamente, por lo tanto,  $R_{si*} = r_{Tsi*}$  si  $Z_{si} = 1$  y  $R_{si*} = r_{Csi*}$  si  $Z_{si} = 0$ .

Si una prueba estadística rechaza la hipótesis nula el 5% de las veces cuando esta es verdadera, uno esperaría que se rechazara al menos 5% de las veces cuando es falsa en la dirección anticipada. Es importante recordar que una prueba estadística es imparcial ante un conjunto de hipótesis alternativas si la prueba exhibe al menos la misma probabilidad de rechazar la hipótesis nula cuando una de las hipótesis alternativas es verdadera que cuando la hipótesis nula es verdadera.

La siguiente proposición sugiere que las pruebas comunes que se han revisado son pruebas imparciales ante efectos positivos de tratamientos, y que la prueba estadística es mejor cuando el efecto es mayor.

**Proposición 3.** *En un experimento aleatorizado, una prueba estadística que es order-preserving produce una prueba imparcial de no efecto ante la alternativa de un efecto positivo, y si  $(r_{T*}, r_{C*})$  es un efecto mayor que  $(r_T, r_C)$ , entonces  $t(Z, R^*) \geq t(Z, R)$ .*

## 2.9. Intervalos de Confianza

### 2.9.1. Pruebas de Hipótesis Generales

En los temas tratados con anterioridad se ha usado la estadística de prueba  $t(Z, \mathbf{r})$  para probar la hipótesis nula de No Efecto del Tratamiento, sin embargo, hay una

extensión a la prueba de hipótesis que especifica un efecto particular del tratamiento. Dicha extensión es utilizada en la construcción de intervalos de confianza que, como ya se sabe, representa el conjunto de hipótesis que no son rechazadas por una prueba.

Considérese probar la hipótesis  $H_0: \lambda = \lambda_0$  bajo el modelo de un efecto aditivo,  $R = r_C + \lambda Z$ . La idea es la siguiente: si la hipótesis nula  $H_0: \lambda = \lambda_0$  fuera cierta, entonces  $r_C = R - \lambda_0 Z$ , por lo tanto, probar que  $R - \lambda_0 Z$  satisface la hipótesis nula de No Efecto del Tratamiento.

De manera más precisa, si  $r_C$  fuera conocida, la probabilidad llamada  $\alpha$ , de que  $t(Z, \mathbf{r})$  es mayor o igual que algún número fijo  $T$  puede ser determinada usando

$$\Pr(t(Z, \mathbf{r}) \geq T) = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{|\Omega|} = \frac{|\{z \in \Omega : t(Z, \mathbf{r}) \geq T\}|}{K}$$

Si la hipótesis nula fuera cierta, entonces  $r_C$  igualaría las respuestas ajustadas,  $R - \lambda_0 Z$ , por lo tanto, bajo la hipótesis nula,  $r_C$  puede calcularse de  $\lambda_0$  y los datos observados. Si la hipótesis  $H_0: \lambda = \lambda_0$  es cierta, entonces la probabilidad de que  $t(Z, R - \lambda_0 Z) \geq T$  es  $\alpha$ , donde  $\alpha$  es calculada, como se describe anteriormente, con  $r_C = R - \lambda_0 Z$ .

Ahora, supóngase que la hipótesis nula no es cierta, se propone  $\lambda > \lambda_0$ , y se considera el comportamiento de la prueba anterior; en este caso,  $R = r_C + \lambda Z$  y las respuestas ajustadas igual a  $r_C + (\lambda - \lambda_0)Z$ , por lo tanto, las respuestas ajustadas variarán con la asignación de tratamientos  $Z$ . Si una unidad recibe el tratamiento, tendrá una respuesta ajustada que es  $\lambda - \lambda_0$  mayor que si esa unidad recibe el control. Si la estadística de prueba es order-preserving, entonces  $t(Z, R - \lambda_0 Z) = t[Z, r_C + (\lambda - \lambda_0)Z] \geq t(Z, r_C) = t(Z, R - \lambda Z)$ , donde la desigualdad se sigue de la definición de una estadística de prueba order-preserving. Es decir, si la hipótesis nula es falsa y no se cumple que  $\lambda > \lambda_0$ , entonces una estadística de prueba  $t(Z, R - \lambda_0 Z)$  será mayor con  $\lambda_0$  que si se hubiera probado con el valor correcto  $\lambda$ .

A continuación se presenta una tabla, en la que se ilustran estos cálculos con una Prueba de rangos. Se trata de un experimento hipotético aleatorizado uniforme con  $N = 8$  unidades, todas en un solo estrato  $S = 1$ , además  $m = 4$  unidades bajo tratamiento y un efecto aditivo del tratamiento  $\lambda = 7$ , aunque la hipótesis nula dice incorrectamente  $H_0: \lambda = \lambda_0 = 1$ .

Unidad	Respuesta bajo control	Tratamiento asignado	Respuesta observada	Respuesta ajustada	Rangos de respuestas ajustadas
$i$	$r_{Ci}$	$Z_i$	$R_i = r_{Ci} + \lambda Z_i$	$R_i - \lambda_0 Z_i$	$q_i$
1	2	1	9	8	7
2	1	0	1	1	1
3	3	0	3	3	2
4	4	0	4	4	3
5	0	1	7	6	5
6	4	1	11	10	8
7	1	1	8	7	6
8	5	0	5	5	4
			$\lambda = 7,$	$\lambda_0 = 1$	

Cuadro 2.7: Ejemplo del Cálculo de Intervalos de Confianza

Entonces, la suma de rangos calculada de las respuestas ajustadas  $R - 1Z$  es  $7 + 5 + 8 + 6 = 26$ , lo cual representa el mayor valor posible que puede tener la suma de rangos para  $N = 8$ ,  $m = 4$ ; el nivel de significancia unilateral es  $\binom{8}{4}^{-1} = \frac{1}{70} = 0,014$  y el nivel de significancia bilateral es  $2 \times 0,014 = 0,028$ .

Después de eliminar la hipótesis de que  $\lambda_0 = 1$ , las unidades tratadas continúan teniendo respuestas mayores que las de las unidades control. El mismo enfoque es usado para hacer otras pruebas de hipótesis bajo otros modelos.

### 2.9.2. Intervalos de Confianza

Bajo el modelo de un efecto aditivo del tratamiento,  $R = r_C + \lambda Z$ , un intervalo de confianza de  $1 - \alpha$  para  $\lambda$ , es obtenido de probar cada valor de  $\lambda$  y coleccionar todos los valores no rechazados al nivel  $\alpha$  en un conjunto  $A$ . Más precisamente,  $A$  es el conjunto de los valores de  $\lambda$  que al ser probados produjeron niveles de significancia o  $p$ -values mayores o iguales que  $\alpha$ . Por ejemplo, en la prueba de la tabla anterior el valor  $\lambda = 1$  no debería estar contenido en un intervalo de confianza al 95%. Cuando se prueba el verdadero valor de  $\lambda$ , éste es rechazado con una probabilidad no mayor que  $\alpha$ , por lo tanto, el conjunto aleatorio  $A$  contiene el verdadero valor de  $\lambda$  con probabilidad al menos de  $1 - \alpha$ . Esto se conoce como *invertir* una prueba y es la manera estándar de obtener un intervalo de confianza de una prueba. Para muchas pruebas estadísticas, una prueba de dos colas o bilateral produce un intervalo de confianza cuyos criterios de valoración pueden ser determinados por una línea de investigación.

## 2.10. Estimación puntual

### 2.10.1. Ausencia de sesgos en estimaciones del efecto promedio

El hecho que más se enfatiza al realizar experimentos aleatorizados es que dichos experimentos conducen a la ausencia de sesgos en el efecto promedio del tratamiento, es decir, conduce a estimadores insesgados. Si se toma el caso más simple: un experimento uniforme aleatorizado con un solo estrato, sin interferencia entre unidades, se tendrán  $m$  unidades tratadas,  $N - m$  unidades control,  $E(Z_i) = m/N$ ,  $R_i = r_{Ti}$  si  $Z_i = 1$ , y  $R_i = r_{Ci}$  si  $Z_i = 0$ . La diferencia entre la respuesta media en el grupo tratado, a saber,  $\frac{1}{m} \sum Z_i R_i$ , y la respuesta media en el grupo control, a saber,  $\frac{1}{N-m} \sum (1 - Z_i) R_i$ , tiene la esperanza

$$\begin{aligned} E \left\{ \sum \left( \frac{Z_i R_i}{m} - \frac{(1 - Z_i) R_i}{N - m} \right) \right\} &= E \left\{ \sum \left( \frac{Z_i r_{Ti}}{m} - \frac{(1 - Z_i) r_{Ci}}{N - m} \right) \right\} = \\ &= \sum \left( \frac{(m/N) r_{Ti}}{m} - \frac{(1 - m/N) r_{Ci}}{N - m} \right) = \frac{1}{N} \sum (r_{Ti} - r_{Ci}), \end{aligned}$$

Y el último término es el promedio de los  $N$  efectos de tratamiento  $r_{Ti} - r_{Ci}$  para las  $n$  unidades experimentales. Es decir, la diferencia entre las medias muestrales de los dos grupos es insesgada para el efecto promedio del tratamiento. Nótese cuidadosamente que esto es verdad sólo cuando se supone que no hay interferencia entre las unidades. No se supone que el efecto de tratamiento  $r_{Ti} - r_{Ci}$  sea constante entre unidades, ni hay supuestos sobre interacción de las unidades. El estimador es insesgado para el efecto promedio de las  $N$  unidades en este estudio, a saber,  $\frac{1}{N} \sum (r_{Ti} - r_{Ci})$ , pero no dice nada acerca del efecto sobre otras unidades que no están en el estudio. Campbell y Stanley (1963) afirman que un experimento aleatorizado tiene validez interna al permitir inferencias acerca de los efectos sobre las  $N$  unidades en el estudio, pero esto no tiene necesariamente, validez externa, es decir, no hay garantía de que el tratamiento será igualmente efectivo para otras unidades fuera del estudio.

La diferencia en las medias muestrales puede ser sesgada cuando hay dos o más estratos y el experimentador asigna desproporcionadamente más sujetos al tratamiento a un determinado estrato que en los otros. Sin embargo, hay un estimador insesgado que corrige el desequilibrio o desbalance. Esto consiste en calcular, dentro de cada estrato  $s$ , la diferencia entre la respuesta promedio en el grupo tratado, a saber  $\frac{1}{m_s} \sum Z_{si} R_{si}$ , y la respuesta promedio en el grupo control, a saber  $\frac{1}{n_s - m_s} \sum (1 - Z_{si}) R_{si}$ , y ponderando esta diferencia por la proporción de unidades en el estrato  $s$ , a saber  $n_s/N$ . El estimador, llamado *ajuste directo*, es entonces:

$$\sum_{s=1}^S \frac{n_s}{N} \left( \sum_{i=1}^{n_s} \left\{ \frac{Z_{si} R_{si}}{m_s} - \frac{(1 - Z_{si}) R_{si}}{n_s - m_s} \right\} \right)$$

Para probar que es insesgado, debe recordarse que en un experimento aleatorizado,  $Z_{si}$  tiene esperanza  $m_s/n_s$ . De lo que se sigue,

$$\begin{aligned} E \left[ \sum_{s=1}^S \frac{n_s}{N} \left( \sum_{i=1}^{n_s} \left\{ \frac{Z_{si} R_{si}}{m_s} - \frac{(1 - Z_{si}) R_{si}}{n_s - m_s} \right\} \right) \right] &= \\ E \left[ \sum_{s=1}^S \frac{n_s}{N} \left( \sum_{i=1}^{n_s} \left\{ \frac{Z_{si} r_{Tsi}}{m_s} - \frac{(1 - Z_{si}) r_{Csi}}{n_s - m_s} \right\} \right) \right] &= \\ \sum_{s=1}^S \frac{n_s}{N} \left( \sum_{i=1}^{n_s} \left\{ \frac{(m_s/n_s) r_{Tsi}}{m_s} - \frac{(1 - m_s/n_s) r_{Csi}}{n_s - m_s} \right\} \right) &= \frac{1}{N} \sum (r_{Ti} - r_{Ci}) \end{aligned}$$

por lo tanto, el ajuste directo ofrece, además, un estimador insesgado del efecto promedio.

En efecto, el ajuste directo trata a las unidades de los dos grupos (tratado y control) como dos muestras aleatorias estratificadas de las  $N$  unidades en el experimento. Entonces, el ajuste directo es el estimador estratificado usual de la media de las respuestas al tratamiento en la población de  $N$  unidades menos el estimador usual de la media de las respuestas control en una población de  $N$  unidades. Nótese de nuevo que el ajuste directo es insesgado para el efecto promedio incluso si el efecto varía de unidad a unidad o de estrato a estrato. Por otro lado, el efecto promedio no es más que un resumen de los efectos, y no una descripción, cuando los efectos varían de un estrato a otro.

### 2.10.2. Estimador Hodges-Lehmann de un efecto aditivo

Bajo el modelo de un efecto aditivo,  $R = r_C + \lambda Z$ , se pueden construir muchos estimadores de  $\lambda$ . Uno de ellos es el estimador de Hodges-Lehmann (1963) que está estrechamente ligado con las pruebas de hipótesis de No Efecto y los intervalos de confianza. Si se establece que  $H_0: \lambda = \lambda_0$  será probada usando  $t(Z, R - \lambda_0 Z)$ , esto es, restando el efecto del tratamiento hipotético,  $\lambda_0 Z$ , a la respuesta observada  $R$ , y preguntándose si la respuesta ajustada  $R - \lambda_0 Z$  parece estar libre del efecto del tratamiento. El estimador Hodges-Lehmann de  $\lambda$  es el valor  $\hat{\lambda}$  tal que las respuestas ajustadas  $R - \hat{\lambda} Z$  parezcan estar exactamente libres del efecto del tratamiento. A continuación se considerará esto en detalle tomando en cuenta que siempre tratamos un experimento uniforme aleatorizado.

Supóngase que se puede determinar la esperanza, es decir,  $\bar{t}$ , de la estadística  $t(Z, R - \lambda_0 Z)$  cuando se usa el  $\lambda$  correcto, esto es, cuando se calcula a partir de las respuestas  $R - \lambda Z$  que han sido ajustadas y que, por lo tanto, están libres del efecto del tratamiento. Por ejemplo, en un experimento con un solo estrato, la estadística de suma de rangos tiene esperanza  $\bar{t} = m(N+1)/2$  si el tratamiento no tiene ningún efecto. Esto se cumple porque, en ausencia de efecto del tratamiento, la estadística de suma de rangos es la suma de  $m$  puntuaciones seleccionadas aleatoriamente de  $N$  puntuaciones, de las cuales la media es  $(N+1)/2$ . De la misma forma, en un experimento estratificado, la estadística de suma de rangos estratificados tiene esperanza  $\bar{t} = \frac{1}{2} \sum m_s (n_s + 1)$  en ausencia de efecto del tratamiento. En un experimento pareado, también en ausencia de efecto del tratamiento, la esperanza de la estadística de rangos asignados es  $\bar{t} = (N+1)/4$ , que es la mitad de la suma de  $N$  puntuaciones con promedio  $(N+1)/2$ . En ausencia de efectos, en un experimento con un solo estrato, la diferencia en las medias muestrales tiene  $\bar{t} = 0$ . En cada uno de estos casos,  $\bar{t}$  puede ser determinada sin conocer  $\lambda$ , por lo tanto, se tiene un estimador Hodges-Lehmann.

Hablando rigurosamente, el estimador Hodges-Lehmann es la solución  $\hat{\lambda}$  de la ecuación  $\bar{t} = t(Z, R - \hat{\lambda}Z)$ . Dicho de otra forma,  $\hat{\lambda}$  es el valor tal que las respuestas ajustadas  $R - \hat{\lambda}Z$  parecen ser completamente libres de los efectos del tratamiento, en el sentido en que la estadística de prueba tiene su esperanza exactamente igual en ausencia de efectos.

Si  $t(\cdot, \cdot)$  es una estadística que preserva el orden, entonces  $t(Z, R - \hat{\lambda}Z)$  es monótona decreciente como función de  $\hat{\lambda}$  con  $Z$  y  $R$  fijos. Esto significa que: el mayor efecto del tratamiento  $\hat{\lambda}Z$  eliminado de las respuestas observadas  $R$ , se convierte en una estadística más pequeña. Esto es de mucha utilidad para resolver  $\bar{t} = t(Z, R - \hat{\lambda}Z)$ . Si un  $\hat{\lambda}$  ha sido probado tal que  $\bar{t} < t(Z, R - \hat{\lambda}Z)$ , entonces un  $\hat{\lambda}$  mayor tenderá a hacer  $t(Z, R - \hat{\lambda}Z)$  más pequeño, acercándose a  $\bar{t}$ . De manera similar, si  $\bar{t} > t(Z, R - \hat{\lambda}Z)$ , entonces se necesita un  $\hat{\lambda}$  menor.

Los problemas surgen inmediatamente. Para estadísticas de rangos, tales como la suma de rangos o la de rangos asignados,  $t(Z, R - \hat{\lambda}Z)$  varía en forma discreta conforme  $\hat{\lambda}$  varía, por lo tanto, puede no existir un valor de  $\hat{\lambda}$  tal que  $\bar{t} = t(Z, R - \hat{\lambda}Z)$ . Para verificar esto, se tomará un ejemplo sencillo de  $N = 2$  y  $m = 1$ . Entonces la estadística de suma de rangos es 0 o 1 ó 2 dependiendo de cual de las dos unidades recibe el tratamiento, pero  $\bar{t} = 3/2$ , por lo tanto, no es posible encontrar una  $\hat{\lambda}$  tal que  $\bar{t} = t(Z, R - \hat{\lambda}Z)$ .

La ecuación  $\bar{t} = t(Z, R - \hat{\lambda}Z)$  no puede sólo no tener solución, sino que puede tener infinitas soluciones. Si  $t(Z, R - \hat{\lambda}Z)$  varía de forma discreta, será constante para intervalos de valores de  $\hat{\lambda}$ . Hodges y Lehmann resolvieron este problema de la siguiente manera: Definieron la solución de una ecuación  $\bar{t} = t(Z, R - \hat{\lambda}Z)$  como



$$SOLUCION \{ \bar{t} = t(Z, R - \hat{\lambda}Z) \}$$

definida por

$$\hat{\lambda} = SOLUCION \{ \bar{t} = t(Z, R - \hat{\lambda}Z) \} = \frac{\inf \{ \lambda : \bar{t} > t(Z, R - \lambda Z) \} + \sup \{ \lambda : \bar{t} < t(Z, R - \lambda Z) \}}{2}$$

Esto define el estimador Hodges-Lehmann. Rigurosamente hablando, si no hay una solución exacta, entonces esta será el promedio de la menor  $\lambda$  que es demasiado grande y la mayor  $\lambda$  que es demasiado pequeña.

Considérese un ejemplo sencillo con base en la tabla 2.7. Bajo la hipótesis nula de No efecto, la estadística de suma de rangos tiene esperanza  $\bar{t} = m(N+1)/2 = 4(8+1)/2 = 18$ , esto es, la mitad de la suma de los ocho rangos,  $36 = 1+2+3+4+5+6+7+8$ . La siguiente tabla muestra algunos valores de  $t(Z, R - \lambda Z)$  para algunos valores de  $\lambda$ ,

$\lambda$	4,9999	5,0000	5,0001	5,9999	6,0000	6,0001
$t(Z, R - \lambda Z)$	20	19	18	18	17	15

Cuadro 2.8: Cálculo del estimador Hodges-Lehmann

Se desea que el estimador elegido sea un valor  $\hat{\lambda}$  tal que  $18 = t(Z, R - \hat{\lambda}Z)$ , pero la tabla indica que cualquier valor entre 5 y 6 lo cumplirá. Como la tabla sugiere,  $\inf \{ \lambda : \bar{t} > t(Z, R - \lambda Z) \} = 6$  y  $\sup \{ \lambda : \bar{t} < t(Z, R - \lambda Z) \} = 5$ , por lo tanto, el estimador Hodges-Lehmann es  $\hat{\lambda} = (6 + 5)/2 = 5,5$ .

Para pruebas estadísticas particulares, existen otras formas de calcular  $\hat{\lambda}$ . Esto se cumple, por ejemplo, para un solo estrato usando la prueba de suma de rangos. En este caso, debe mostrarse que  $\hat{\lambda}$  es la mediana de las  $m(N - m)$  diferencias por pares formadas tomando cada una de las  $m$  respuestas de los tratados y restándoles cada una de las  $N - m$  respuestas control.

El estimador Hodges-Lehmann,  $\hat{\lambda}$ , hereda las propiedades de la estadística de prueba, tal como la consistencia. Debe recordarse que una prueba es consistente si la probabilidad de rechazar cada hipótesis falsa tiende a 1 conforme el tamaño de la muestra incrementa; como es de esperarse, estas ideas están interconectadas. Una prueba que rechaza valores incorrectos de  $\lambda$  conduce a un estimador que se aleja de estos valores incorrectos. En otras palabras, bajo ligeras condiciones, las pruebas consistentes conducen a estimadores Hodges-Lehmann consistentes.

# Capítulo 3

## Introducción al análisis de datos longitudinales

En muchos experimentos en general, los investigadores se ven atraídos por indagar cómo se van dando cambios en las mediciones de las variables bajo estudio a través del tiempo. Esto pasa también con los investigadores clínicos, quienes muy frecuentemente están interesados en rastrear los cambios en las mediciones que son tomadas de los sujetos durante las visitas de seguimiento. A continuación se da una introducción a algunos modelos de regresión que pueden ser usados para ajustar y analizar los datos.

### 3.1. Definiciones básicas

El *análisis de datos longitudinales* o análisis longitudinal de datos es el análisis de los datos obtenidos a través de un estudio longitudinal. Un *estudio longitudinal* reúne las mediciones tomadas repetidamente a través del tiempo sobre los mismos sujetos. En una investigación médica o clínica, los datos que se obtienen de esta forma se conocen como datos longitudinales. Por lo tanto, un análisis longitudinal es el seguimiento de la evolución en el tiempo de un conjunto de líneas de vidas o cohortes (formadas por sujetos), y la incidencia de los diferentes fenómenos que las afectan. Un método alternativo para la toma de mediciones y que se contrapone al análisis longitudinal es un estudio seccional cruzado (cross-sectional study), en el cual las mediciones que se toman a los sujetos son realizadas sólo una vez en el tiempo. La ventaja de los estudios longitudinales es que se hace un seguimiento de los cambios en las mediciones, mientras que en los estudios seccionales-cruzados no existe dicho seguimiento.

Los ensayos clínicos son un caso especial de estudios longitudinales, ya que en dichos ensayos se realizan mediciones en cada una de las visitas de seguimiento. Por ejem-

plo, en un ensayo para probar una nueva droga que ayuda a disminuir los niveles de lipoproteínas de baja densidad (mejor conocidas como colesterol malo) en el cuerpo, se hacen mediciones periódicas de los niveles de colesterol en la sangre. Se considera benéfico tomar mediciones adicionales de algunas otras variables tales como peso o cantidad de ejercicio físico diario, así como hacer observaciones sobre si los sujetos llevan una dieta correcta.

## 3.2. Presentación gráfica

Es muy frecuente que, en la práctica, se haga una primera revisión de los datos longitudinales por medio de representaciones gráficas, para lo que se construyen distintos tipos de diagramas y gráficas, según lo prefiera o requiera el investigador. El siguiente ejemplo<sup>1</sup> ilustrará el uso de la presentación gráfica.

**Ejemplo 1.** Se realiza un ensayo clínico para comparar la efectividad de dos tratamientos distintos para atacar tumores cancerígenos. Se han registrado los datos del ensayo clínico que incluye a 24 sujetos, los cuales tienen un tumor cancerígeno en el cerebro (gliomas malignos). Dichos sujetos son divididos aleatoriamente en dos grupos de tratamiento. Los sujetos del grupo 1 reciben quimioterapia intravenosa mientras los sujetos del grupo 2 reciben quimio-radioterapia concurrente. En las visitas de seguimiento, que se programaron a los 0, 3, 6, 12, 18 y 24 meses, se hace la medición, en centímetros, del diámetro que muestra el tumor a través de la imagen de la resonancia magnética (IRM); a continuación, en el Cuadro 3.1 se muestran las mediciones registradas de los sujetos en cada visita de seguimiento.

En este ensayo clínico primero se grafica la variable respuesta (diámetro del tumor) registrada en cada visita de seguimiento, haciéndolo de manera individual para cada sujeto, como se muestra en la Figura 3.1<sup>2</sup>.

Cada una de las líneas que se obtuvieron en esta gráfica se llama Perfil Individual de Respuesta y por la forma en que se elaboró la gráfica los perfiles punteados corresponden a los sujetos del grupo 1 y los perfiles de líneas continuas a los sujetos del grupo 2.

Se puede observar que, a excepción de uno, todos los perfiles muestran una contracción del tumor al final del tiempo, o al menos no hay crecimiento, pues aunque a través del tiempo hay incremento en algunos casos luego vuelve a haber contracción, y al tiempo final casi todos logran una contracción del tumor. Además, la mayoría de los perfiles de los sujetos del grupo 2 tienden a estar más abajo que los perfiles de los

---

<sup>1</sup>Este ejemplo fue tomado del libro *Clinical Statistics* de Olga Korosteleva. Aunque los datos se obtuvieron directamente del libro, los gráficos y diagramas fueron elaborados y modificados por el autor.

<sup>2</sup>Las siguientes dos gráficas fueron realizadas en Excel

Sujeto	Grupo	Tamaño del tumor					
		0 meses	3 meses	6 meses	12 meses	18 meses	24 meses
1	1	3.1	3	2.7	2.3	2.1	1.8
2	1	3.3	2.9	2.4	1.8	1.7	0.2
3	1	2.9	2.4	2.3	2.1	2.1	1.6
4	1	3.2	2.7	2.7	2.2	2.1	1.3
5	1	3.5	3.2	3.2	3.2	3.1	0.8
6	1	3.6	3.5	1.7	1.6	1.5	1.1
7	1	2.2	2	2.7	2.4	2.4	1.6
8	1	3.8	3.7	3	2.8	2.6	1
9	1	3.4	3.1	3.3	4.3	4.1	3.6
10	1	4.6	4.4	2.6	2.5	2.4	1.5
11	1	2.7	2.6	3.1	3.1	3	1.6
12	1	4.9	4.7	2.9	2.9	2.8	2.6
13	2	4	3.5	3.1	2	1.1	1
14	2	3.8	3.3	2.7	1.6	1.2	0.6
15	2	3.7	3.2	2.9	2.4	1.5	1.4
16	2	3.2	3.1	2.9	1.8	0.6	0
17	2	2.5	2.1	1.5	0.3	0.1	0
18	2	2.9	2.2	3.1	1.9	1.7	1.5
19	2	2.8	2.6	2.8	2.6	2.6	2.3
20	2	3.1	2.7	2.3	2	1.9	1.3
21	2	4.3	4.2	2.4	2.3	1.2	0.4
22	2	2.9	2.4	2.1	2.1	1.7	1.7
23	2	4	3.4	2.3	1.4	0.9	0
24	2	2.5	2.4	2	1	0.3	0

Cuadro 3.1: Tamaño del tumor en cms en cada visita de seguimiento

sujetos del grupo 1.

Una manera de hacer más rigurosa esta última afirmación es obteniendo el Perfil Medio de Respuesta para cada uno de los grupos y graficarlos de manera análoga a como se procedió en los perfiles individuales. De esta forma se obtiene la gráfica de la Figura 3.2 de Perfiles Medios.

Ahora se pueden observar las medias de la variable respuesta registradas en cada visita de seguimiento para cada uno de los grupos y con ello se puede corroborar que la media del tamaño del tumor para los sujetos del grupo 2 en cada visita es más pequeña que la media de los sujetos del grupo 1. Se observa que al inicio del estudio, la media de ambos grupos era cercana y se percibe como conforme transcurre el tiempo, en las últimas visitas de seguimiento las medias comienzan a alejarse mostrando que la mejora es mayor para los sujetos del grupo 2.

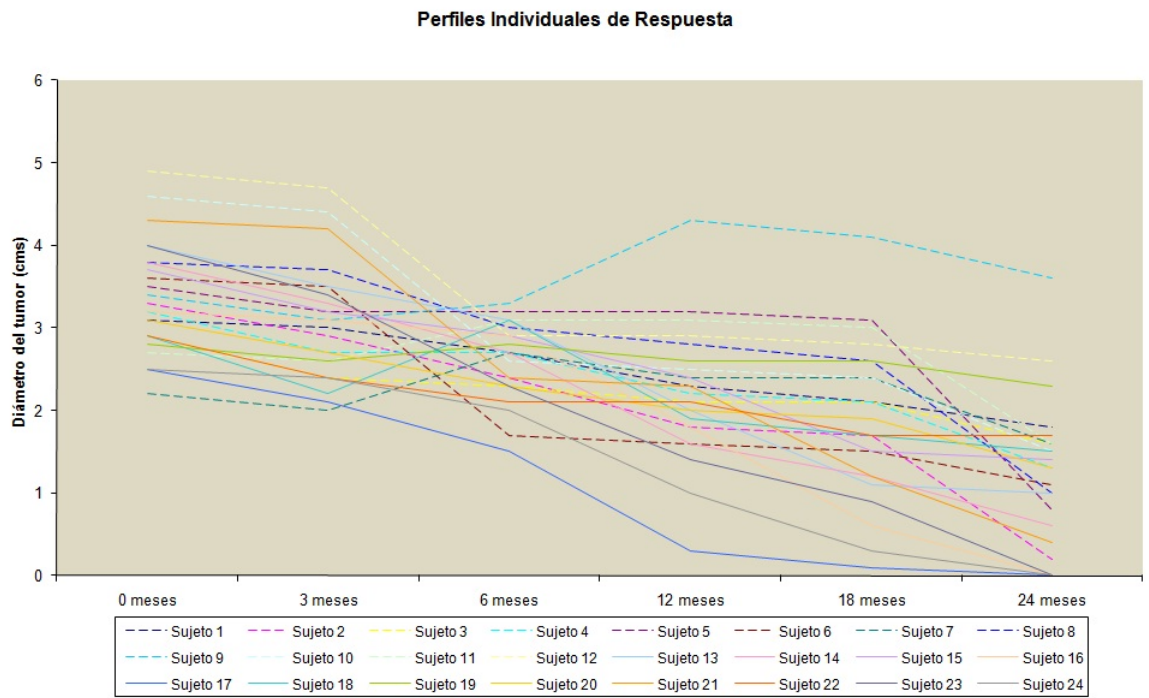


Figura 3.1: Perfiles individuales de respuesta. Líneas punteadas = grupo 1, Líneas continuas = grupo 2

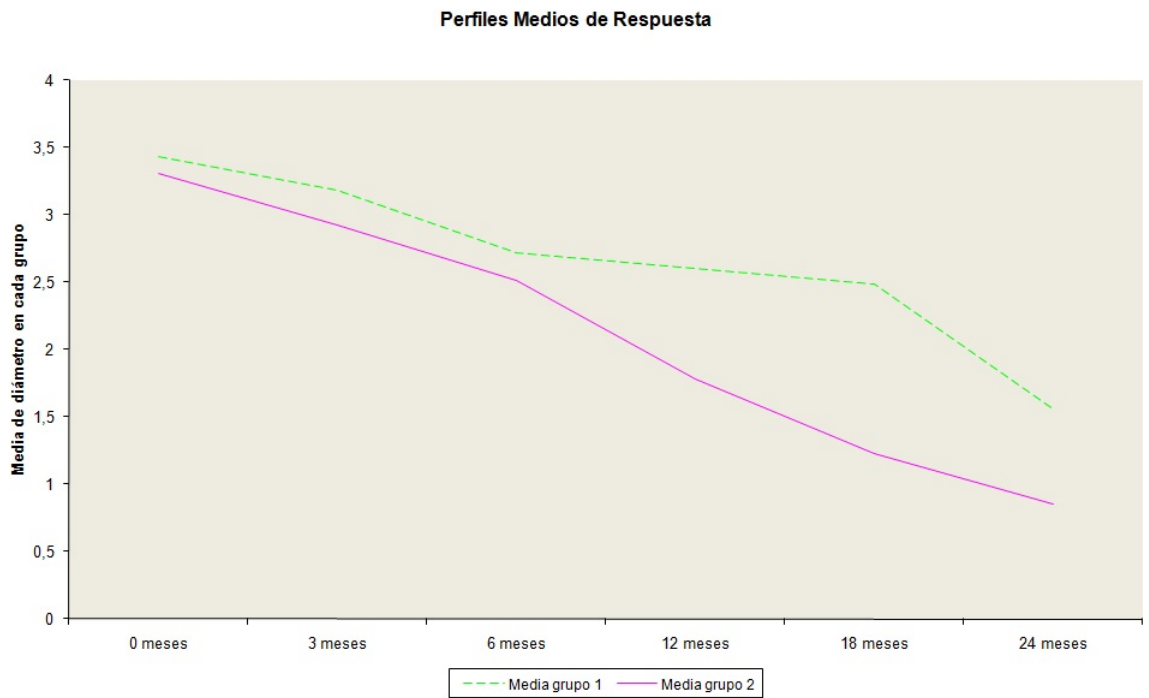


Figura 3.2: Perfiles de Respuesta Media. Línea punteada = grupo 1, Línea continua = grupo 2

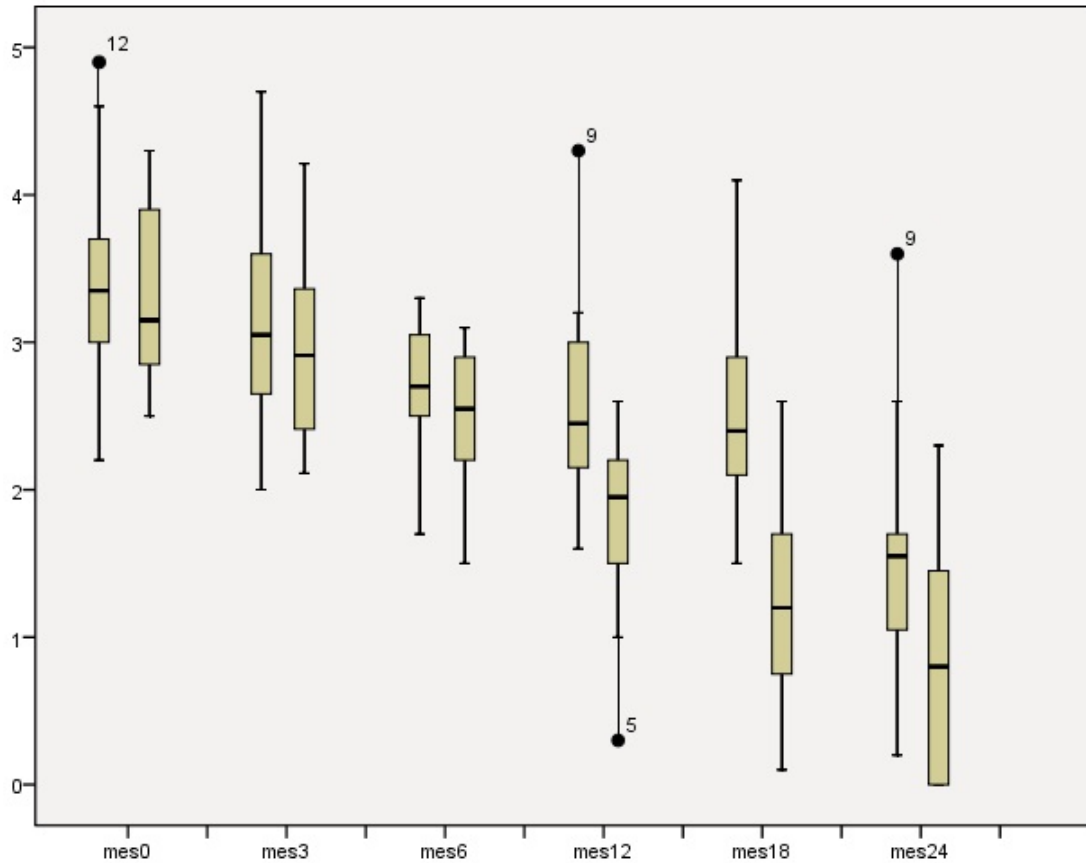


Figura 3.3: Diagramas de caja para cada grupo en cada visita de seguimiento

Otro recurso disponible para el análisis gráfico de los datos es dibujar diagramas de caja para la variable respuesta en cada grupo por cada una de las visitas. Un diagrama de caja es una forma conveniente de obtener de manera resumida cinco datos importantes del conjunto de datos: los cuartiles primero, segundo y tercero (el segundo es la mediana), y las observaciones máxima y mínima en la variable respuesta. Existen varios paquetes estadísticos que son auxiliares en el diseño de estos diagramas.

Se incluyen en la Figura 3.3 los diagramas de caja para cada uno de los grupos en cada visita de seguimiento con los mismos datos tratados anteriormente<sup>3</sup>.

Para cada visita se encuentra un par de diagramas de caja. En cada una de las visitas, el diagrama de caja de la izquierda corresponde a los sujetos en el grupo 1 y el de la derecha a los sujetos del grupo 2. Nótese que los diagramas para el grupo 1 tienden a ser más altos que los del grupo 2, y como la variable respuesta es el diámetro

<sup>3</sup>Los diagramas de caja presentados en este trabajo fueron realizados con ayuda del paquete estadístico SPSS

del tumor significa que en el grupo 1 el tamaño del tumor de los sujetos disminuye más lento que en los sujetos del grupo 2. Con los gráficos anteriores se advierte que la respuesta media del grupo 2 es menor que la respuesta media del grupo 1 en todas las visitas. En este último gráfico se puede corroborar que pasa algo análogo con la mediana de ambos grupos. La mediana en el grupo 2 siempre es menor que la mediana en el grupo 1, lo cual refleja la superioridad del tratamiento aplicado al grupo 2.

En las tres primeras visitas de seguimiento los diagramas de ambos grupos están más alineados que en las tres últimas visitas, en las cuales, los diagramas del grupo 2 se separan más de los del grupo 1 siendo más bajos cada vez, lo cual sugiere, que en las últimas visitas, el tratamiento aplicado al grupo 2 tuvo mucho mejores resultados que los del grupo 1. No obstante, al comparar la última visita con la penúltima, se observa que en el tiempo transcurrido entre ambas, el tratamiento aplicado al grupo 1 también provoca una gran mejora en los pacientes.

Así, lo que sugiere la ayuda gráfica es que el segundo tratamiento -quimio-radioterapia concurrente- es superior al primero -quimioterapia-.

### 3.3. Algunos modelos para el análisis de datos longitudinales

Los estudios longitudinales son diseñados para medir los cambios que se producen en las variables respuesta individuales (de cada sujeto) en relación a un conjunto de **covariables**, que también pueden llamarse variables predictor. Los modelos lineales para observaciones longitudinales deben tomar en cuenta la covariable *estructura* de los datos. En los modelos que aquí son considerados, se adopta el supuesto de que, para diferentes sujetos, los datos son independientes - es decir, existe una correlación intra-sujeto (dentro de cada sujeto) de la variable respuesta a través del tiempo, mientras que la correlación entre sujetos es insignificante.

Los modelos de efectos combinados son usados para modelos de datos longitudinales. En estos modelos, algunas de las covariables tienen efectos aleatorios (niveles aleatorios) mientras las otras tienen efectos fijos (niveles fijos). A continuación, se discutirá el modelo con intercepto aleatorio que es un caso especial simple de los modelos de efectos combinados, además de otros modelos con interesantes estructuras de correlación en términos del error.

### 3.3.1. Modelo con intercepto aleatorio

#### Definición del modelo

Supóngase que se tienen disponibles observaciones longitudinales de  $N$  sujetos. Los datos son obtenidos en  $k$  tiempos establecidos  $t_1, t_2, \dots, t_k$ . Sea  $y_{ij}$  la respuesta observada sobre el  $i$ -ésimo sujeto para  $i = 1, \dots, N$  en el tiempo  $t_j$  para  $j = 1, \dots, k$  y sean  $x_{1ij}, \dots, x_{pij}$  los valores observados de  $p$  covariables de efectos fijos sobre el  $i$ -ésimo sujeto al tiempo  $t_j$ . Un modelo con intercepto aleatorio es de la siguiente forma

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i + \epsilon_{ij}$$

donde las  $u_i \sim N(0, \sigma_u^2)$ , son independientes e idénticamente distribuidas para  $i = 1, \dots, N$  y son, cada una, un *intercepto aleatorio*, y los  $\epsilon_{ij} \sim N(0, \sigma^2)$  (también independientes e idénticamente distribuidos) son *los errores aleatorios*. Se agrega la hipótesis de que  $u_i$  y  $\epsilon_{ij}$  son independientes. En este modelo, la varianza de las observaciones es constante. En efecto,  $Var(y_{ij}) = Var(u_i + \epsilon_{ij}) = \sigma_u^2 + \sigma^2$ . Del mismo modo, la covarianza entre las observaciones en los tiempos  $t_j$  y  $t_{j'}$ , en el  $i$ -ésimo sujeto es constante:

$$Cov(y_{ij}, y_{ij'}) = Cov(u_i + \epsilon_{ij}, u_i + \epsilon_{ij'}) = Var(u_i) = \sigma_u^2$$

Las observaciones para sujetos diferentes  $i$ -ésimo e  $i'$ -ésimo en los tiempos  $t_j$  y  $t_{j'}$ , no están correlacionados porque

$$Cov(y_{ij}, y_{i'j'}) = Cov(u_i + \epsilon_{ij}, u_{i'} + \epsilon_{i'j'}) = 0.$$

Usando la notación matricial, el modelo con intercepto aleatorio queda representado como sigue:

$$\underline{y} = X\underline{\beta} + \underline{u} + \underline{\epsilon}$$

donde

$$\underline{y}_{(Nk,1)} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1k} \\ y_{21} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{Nk} \end{bmatrix}, X_{(Nk,p+2)} = \begin{bmatrix} 1 & x_{111} & \dots & x_{p11} & t_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{11k} & \dots & x_{p1k} & t_k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1N1} & \dots & x_{pN1} & t_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1Nk} & \dots & x_{pNk} & t_k \end{bmatrix}$$



$$\underline{\beta}_{(p+2,1)} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{p+1} \end{bmatrix}, \underline{u}_{(Nk,1)} = \begin{bmatrix} u_1 \\ \vdots \\ u_1 \\ \vdots \\ u_N \\ \vdots \\ u_N \end{bmatrix}, \underline{\epsilon}_{(Nk,1)} = \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1k} \\ \vdots \\ \epsilon_{N1} \\ \vdots \\ \epsilon_{Nk} \end{bmatrix}$$

La matriz de covarianzas  $\mathbf{V}$  para el vector respuesta  $\underline{y}$  es una matriz diagonal de bloques, con  $N$  bloques no nulos de tamaño  $k \times k$  construidos, cada uno, como el siguiente

$$\mathbf{V}_{(k,k)} = \begin{bmatrix} \sigma^2 + \sigma_u^2 & \sigma_u^2 & \vdots & \sigma_u^2 \\ \sigma_u^2 & \sigma^2 + \sigma_u^2 & \vdots & \sigma_u^2 \\ \dots & \dots & \dots & \dots \\ \sigma_u^2 & \sigma_u^2 & \vdots & \sigma^2 + \sigma_u^2 \end{bmatrix} = \sigma^2 I_k + \sigma_u^2 J_k$$

donde  $I_k$  es la matriz identidad de tamaño  $k \times k$ , y  $J_k$  denota la matriz de tamaño  $k \times k$  con todas las entradas iguales a 1.

## Interpretación de Coeficientes

La interpretación de los coeficientes  $\beta_0, \dots, \beta_{p+1}$  del modelo es la misma que en un modelo de regresión lineal ordinario. La respuesta media  $E(y_{ij}) = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j$  y, por lo tanto, cada coeficiente  $\beta_i$  representa el cambio en la respuesta media para un incremento unitario en la variable respectiva, siempre que las otras variables estén fijas.

Para una variable categórica con  $l$  niveles, los coeficientes  $(\beta_1, \dots, \beta_{l-1})$  corresponden a las variables ficticias, que son indicadores de los niveles  $1, \dots, l-1$ . Entonces  $\beta_a - \beta_b$ , donde  $a, b = 1, \dots, l-1$ , es interpretada como la diferencia en la respuesta media para los sujetos en el nivel  $a$  y los sujetos en el nivel  $b$  de la covariable, mientras las otras variables permanecen iguales. El coeficiente  $\beta_a$  es la diferencia en la respuesta media para los sujetos en el nivel  $a$  y los sujetos en el nivel  $l$  de la covariable.

## Estimación de parámetros

A continuación se tratan dos métodos para la estimación de los parámetros de un modelo y son el método de máxima verosimilitud y el método de máxima verosimilitud restringida.

**a) Método de Máxima Verosimilitud.** Un enfoque común para la estimación de los parámetros del modelo con intercepto aleatorio  $\beta_0, \dots, \beta_{p+1}, \sigma^2$  y  $\sigma_u^2$  es el método de máxima verosimilitud. La distribución del vector respuesta  $\underline{y}$  es una normal multivariada con media  $X\underline{\beta}$  y matriz de covarianzas  $\mathbf{V}$ . La matriz  $\mathbf{V}$  de tamaño  $Nk \times Nk$  está formada diagonalmente con  $N$  bloques  $\mathbf{V}'$  de dimensiones  $k \times k$  cada uno. Por lo tanto, la función de log-verosimilitud es proporcional a

$$\ln L(\underline{\beta}, \mathbf{V}') \propto -\frac{n}{2} \ln |\mathbf{V}'| - \frac{1}{2} (\underline{y} - X\underline{\beta})^T \mathbf{V}'^{-1} (\underline{y} - X\underline{\beta})$$

donde  $|\mathbf{V}'|$  es el determinante de  $\mathbf{V}'$ , y  $\mathbf{V}'^{-1}$  la inversa de la matriz  $\mathbf{V}'$  es una matriz diagonal formada con los bloques  $\mathbf{V}'^{-1}$ . Se puede demostrar que en el modelo con intercepto aleatorio

$$|\mathbf{V}'| = \sigma^{2k} + k\sigma^{2(k-1)}\sigma_u^2$$

$$\mathbf{V}'^{-1} = \frac{1}{\sigma^4 + k\sigma^2\sigma_u^2} [(\sigma^2 + k\sigma_u^2)I_k - \sigma_u^2 J_k]$$

Si  $\mathbf{V}'$  es conocido, entonces el estimador máximo verosímil de  $\underline{\beta}$  es

$$\hat{\beta}(\mathbf{V}') = (X^T \mathbf{V}'^{-1} X)^{-1} X^T \mathbf{V}'^{-1} \underline{y}$$

Sustituyendo esta expresión en la primera ecuación se obtiene

$$\ln L(\hat{\beta}(\mathbf{V}'), \mathbf{V}') \propto -\frac{n}{2} \ln |\mathbf{V}'| - \frac{1}{2} RSS(\mathbf{V}')$$

donde  $RSS(\mathbf{V}') = (\underline{y} - X\hat{\beta}(\mathbf{V}'))^T \mathbf{V}'^{-1} (\underline{y} - X\hat{\beta}(\mathbf{V}'))$  denota la suma de los cuadrados residuales. De esto se puede deducir que la función de log-verosimilitud depende únicamente de  $\sigma^2$  y  $\sigma_u^2$ . Maximizando con respecto a estas dos variables se producen los estimadores máximo verosímiles  $\hat{\sigma}^2$  y  $\hat{\sigma}_u^2$  (en consecuencia  $\hat{\mathbf{V}}'$ ), los cuales se relacionan en la ecuación  $\hat{\beta}(\mathbf{V}') = (X^T \mathbf{V}'^{-1} X)^{-1} X^T \mathbf{V}'^{-1} \underline{y}$  para obtener  $\hat{\beta}(\hat{\mathbf{V}}')$ .

**b) Método de Máxima Verosimilitud Restringida.** El método de máxima verosimilitud produce un estimador sesgado para la varianza. Entonces un método alternativo de encontrar los estimadores de los parámetros es el método de Máxima Verosimilitud Restringida.

Los estimadores que son obtenidos por este método para  $\sigma^2$  y  $\sigma_u^2$  maximizan la función de log-verosimilitud de cierta transformación lineal de  $\underline{y}$ . La transformación es elegida de tal manera que la función resultante, a diferencia de lo que pasa en el método de Máxima Verosimilitud, no depende de  $\underline{\beta}$ . Los estimadores que proporciona este segundo método para las varianzas, es decir,  $\hat{\sigma}^2$  y  $\hat{\sigma}_u^2$ , maximizan esta función de log-verosimilitud y son estimadores insesgados para los parámetros.

**Proposición 4.** *Considérese la matriz de tamaño  $Nk \times Nk$ ,*

$$I_{Nk} - X (X^T X)^{-1} X^T$$

*que convierte a  $\underline{y}$  en los residuales mínimos cuadrados ordinarios. Aquí  $I_{Nk}$  denota la matriz identidad de tamaño  $Nk \times Nk$ . De la teoría general de algebra lineal, existe una matriz  $A$  de tamaño  $(Nk - p - 2) \times Nk$  con las siguientes dos propiedades  $A^T A = I_{Nk} - X (X^T X)^{-1} X^T$  y  $AA^T = I_{Nk-p-2}$  donde  $I_{Nk-p-2}$  es la matriz identidad de tamaño  $(Nk - p - 2) \times (Nk - p - 2)$ . Al introducir la transformación  $Z = A\underline{y}$  que propone el método de máxima verosimilitud restringida, se obtiene un vector aleatorio de longitud  $Nk - p - 2$ . Entonces la correspondiente función de log-verosimilitud restringida tiene la forma*

$$\ln L_r(\mathbf{V}') \propto -\frac{n}{2} \ln |\mathbf{V}'| - \frac{1}{2} \ln |X^T \mathbf{V}^{-1} X| - \frac{1}{2} RSS(\mathbf{V}')$$

*donde la suma residual de los cuadrados  $RSS(\mathbf{V}') = (\underline{y} - X\hat{\beta}(\mathbf{V}'))^T \mathbf{V}^{-1} (\underline{y} - X\hat{\beta}(\mathbf{V}'))$  es igual que en el método anterior.*

Demostración. Considérese el estimador por mínimos cuadrados ponderado de  $\hat{\beta}$ :

$$\hat{\beta} = (X^T \mathbf{V}^{-1} X)^{-1} X^T \mathbf{V}^{-1} \underline{y} = B\underline{y}$$

donde  $B = (X^T \mathbf{V}^{-1} X)^{-1} X^T \mathbf{V}^{-1}$  es una matriz de tamaño  $(p+2) \times Nk$ . Es bien sabido que  $\hat{\beta}$  tiene una distribución normal multivariada con media  $\underline{\beta}$  y matriz de covarianzas  $(X^T \mathbf{V}^{-1} X)^{-1}$ . Así la función de densidad de  $\hat{\beta}$  es

$$f_{\hat{\beta}}(\hat{\beta}) = (2\pi)^{\frac{p+2}{2}} |X^T \mathbf{V}^{-1} X|^{1/2} \exp \left\{ -\frac{1}{2} (\hat{\beta} - \underline{\beta})^T (X^T \mathbf{V}^{-1} X) (\hat{\beta} - \underline{\beta}) \right\} \dots \dots \textcircled{1}$$

donde  $|X^T \mathbf{V}^{-1} X|$  denota el determinante de la matriz  $X^T \mathbf{V}^{-1} X$ . Además, usando las propiedades de  $A$ , se llega a que

$$AX = I_{Nk-p-2} AX = (AA^T) AX = A(A^T A) X = A(I_{Nk} - X(X^T X)^{-1} X^T) X = A(X - X) = 0 \dots \dots \dots \textcircled{2}$$

Ahora debe recordarse que el vector respuesta  $\underline{y}$  tiene una distribución normal multivariada con media  $X\underline{\beta}$  y matriz de covarianzas  $\mathbf{V}$ . Ya que  $Z$  es una transformación lineal de  $\underline{y}$ , su distribución es también normal multivariada. La media de  $Z$  es  $E(Z) = E(A\underline{y}) = AX\underline{\beta}$  y por el último resultado obtenido, se tiene que  $E(Z) = 0$ . Además,  $Z$  y  $\hat{\underline{\beta}}$  no están correlacionadas (y a partir de eso son independientes porque ambas son normalmente distribuidas). Además,

$$\begin{aligned} Cov(Z, \hat{\underline{\beta}}) &= E[Z(\hat{\underline{\beta}} - \underline{\beta})^T] = E[A\underline{y}(\underline{y}^T B^T - \underline{\beta}^T)^T] = \\ &A\{E[\underline{y}\underline{y}^T] B^T\} - A\{E[\underline{y}]\} \underline{\beta}^T = A\{Var(\underline{y}) + E(\underline{y})E(\underline{y}^T)\} B^T - AX\underline{\beta}\underline{\beta}^T = \\ &A\{\mathbf{V} + X\underline{\beta}\underline{\beta}^T X^T\} B^T - AX\underline{\beta}\underline{\beta}^T = A\mathbf{V}B^T = A\mathbf{V}\mathbf{V}^{-1}X(X^T \mathbf{V}^{-1} X)^{-1} = 0 \end{aligned}$$

donde las últimas dos igualdades se dan por la ecuación  $\textcircled{2}$ .

A continuación, se escriben en forma matricial las transformaciones introducidas quedando

$$\begin{bmatrix} Z \\ \hat{\underline{\beta}} \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} \underline{y}$$

Por algebra lineal, la ecuación de igualación del diferencial respectivo es

$$dZd\hat{\underline{\beta}} = |J(A, B)| d\underline{y} \dots \dots \dots \textcircled{3}$$

donde  $|J(A, B)| = \left| \begin{bmatrix} A \\ B \end{bmatrix} \right|$  denota al determinante del Jacobiano que se calcula como sigue:

$$\begin{aligned} |J(A, B)| &= \left| \begin{bmatrix} A \\ B \end{bmatrix} \right| = \left( \left| \begin{bmatrix} A \\ B \end{bmatrix} \right| \left| \begin{bmatrix} A^T & B^T \end{bmatrix} \right| \right)^{\frac{1}{2}} = \left| \begin{bmatrix} A \\ B \end{bmatrix} \begin{bmatrix} A^T & B^T \end{bmatrix} \right|^{\frac{1}{2}} = \\ &\left| \begin{bmatrix} AA^T & AB^T \\ BA^T & BB^T \end{bmatrix} \right|^{\frac{1}{2}} \end{aligned}$$

Ahora se hace uso del siguiente resultado sobre determinantes: para matrices cualesquiera  $F$ ,  $G$  y  $H$ , se cumple

$$\left| \begin{bmatrix} F & G \\ G^T & H \end{bmatrix} \right| = |FH - FG^T F^{-1}G|$$

Aplicando este resultado y las definiciones de  $A$  y  $B$ , se obtiene

$$|J(A, B)| = \left| AA^T (BB^T) - (AA^T) BA^T (AA^T)^{-1} AB^T \right|^{\frac{1}{2}} = |X^T X|^{-\frac{1}{2}} \dots \dots \dots \textcircled{4}$$

Finalmente, todo se reduce a expresar la función de densidad de  $Z$  en términos de  $\underline{y}$ . Reescribiendo

$$f(Z, \hat{\beta}) dZ d\hat{\beta} = f_Z(Z) f_{\hat{\beta}}(\hat{\beta}) dZ d\hat{\beta}$$

ahora, por independencia

$$= f_{\underline{y}}(\underline{y}) |X^T X|^{-\frac{1}{2}} d\underline{y}$$

por las ecuaciones  $\textcircled{3}$  y  $\textcircled{4}$ . Consecuentemente, usando la ecuación  $\textcircled{1}$ , se obtiene

$$f_Z(Z) = |X^T X|^{-\frac{1}{2}} \frac{f_{\underline{y}}(\underline{y})}{f_{\hat{\beta}}(\hat{\beta})} = (2\pi)^{-\frac{Nk-p-2}{2}} |X^T X|^{-\frac{1}{2}} |\mathbf{V}|^{-\frac{1}{2}} |X^T \mathbf{V}^{-1} X|^{-\frac{1}{2}} \times$$

$$\exp \left\{ -\frac{1}{2} \left[ (\underline{y} - X\hat{\beta})^T \mathbf{V}^{-1} (\underline{y} - X\hat{\beta}) - (\hat{\beta} - \beta)^T (X^T \mathbf{V}^{-1} X) (\hat{\beta} - \beta) \right] \right\} \propto$$

$$|X^T X|^{-\frac{1}{2}} |\mathbf{V}|^{-\frac{1}{2}} |X^T \mathbf{V}^{-1} X|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (\underline{y} - X\hat{\beta})^T \mathbf{V}^{-1} (\underline{y} - X\hat{\beta}) \right\}$$

Entonces aplicando logaritmos se obtiene la ecuación que describe la función de log-verosimilitud restringida que se quería demostrar.

## Ejemplo de aplicación del modelo

**Ejemplo 2.**<sup>4</sup> Si se consideran los datos del ejemplo 1 de la presentación gráfica (del estudio de personas con tumores cancerígenos cerebrales), el modelo con intercepto aleatorio parece ser un modelo apropiado para su análisis por las siguientes razones: la dispersión del tamaño de los gliomas no incrementa drásticamente a través del tiempo, por lo tanto, puede decirse que la respuesta varía de manera constante en el tiempo; además, para cada sujeto, el historial de los diámetros del tumor influye de la misma forma en el diámetro actual, por ello puede adoptarse el supuesto de que la covarianza es constante.

Dado que los sujetos en el grupo 1 (pacientes sometidos a la quimioterapia) están para ser comparados con los sujetos del grupo 2 (pacientes sometidos a la quimiorradioterapia), la variable *grupo* debería ser considerada como una covariable en el modelo. Los valores individuales para esta covariable se denotan por  $x_i$ . Nótese que  $x_i = 1$  si el  $i$ -ésimo sujeto está en el grupo 1 y  $x_i = 2$  si está en el grupo 2. Sea  $y_{ij}$  el tamaño del tumor del  $i$ -ésimo sujeto al tiempo  $t_j$ , donde  $t_1 = 0$ ,  $t_2 = 3$ ,  $t_3 = 6$ ,  $t_4 = 12$ ,  $t_5 = 18$  y  $t_6 = 24$  (todos en meses).

Entonces el modelo es,

$$y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 t_j + u_i + \epsilon_{ij}$$

para  $i = 1, 2, \dots, 24$  donde los interceptos aleatorios  $u_i \sim N(0, \sigma_u^2)$ , son independientes e idénticamente distribuidas y son independientes de los errores aleatorios  $\epsilon_{ij} \sim N(0, \sigma^2)$  que también son independientes e idénticamente distribuidos.

De los paquetes estadísticos disponibles, a continuación, para la aplicación de los modelos, se utiliza el paquete SAS, mediante el cual se obtiene el valor de los estimadores de los parámetros del modelo por medio de los métodos de Máxima Verosimilitud y de Máxima Verosimilitud Restringida<sup>5</sup>.

De los resultados que se obtuvieron en SAS, los que se presentan en los siguientes cuadros son los más relevantes.

Para el Método de estimación de Máxima Verosimilitud

---

<sup>4</sup>Los ejemplos que se introducen en cada uno de los siguientes modelos fueron adaptados tomando los datos de los ejemplos del libro de Clinical Statistics de Olga Korosteleva pero los resultados fueron obtenidos por el autor con ayuda del paquete estadístico SAS

<sup>5</sup>El código en SAS correspondiente a este ejemplo se encuentra en el Anexo 3

Estimación de los parámetros del modelo			
Estimador	Efecto	Estimación	Pr > t  (p-value)
$\hat{\beta}_0$	Intercepto	4,1240	<0,0001
$\hat{\beta}_1$	Grupo	-0,5625	0,0038
$\hat{\beta}_2$	Tiempo	-0,08548	<0,0001

Estimación de los parámetros de la varianza		
Estimador	Parámetro	Estimación
$\hat{\sigma}_u^2$	Intercepto	0,1636
$\hat{\sigma}^2$	Residual	0,3237

Y para el Método de estimación de Máxima Verosimilitud Restringida

Estimación de los parámetros del modelo			
Estimador	Efecto	Estimación	Pr > t  (p-value)
$\hat{\beta}_0$	Intercepto	4,1240	<0,0001
$\hat{\beta}_1$	Grupo	-0,5625	0,0055
$\hat{\beta}_2$	Tiempo	-0,08548	<0,0001

Estimación de los parámetros de la varianza		
Estimador	Parámetro	Estimación
$\hat{\sigma}_u^2$	Intercepto	0,1829
$\hat{\sigma}^2$	Residual	0,3264

A partir de estos resultados se pueden obtener algunas conclusiones. En ambos casos,

se determina que los coeficientes del modelo son estadísticamente significativos al nivel de significancia del 1 %.

Se sabe que  $\hat{\beta}_1$  es estimador de  $\beta_1$  y que este último es el coeficiente de la covariable grupo. Partiendo de que se está comparando al grupo 1 con el grupo 2 y haciéndose notar que el estimador  $\hat{\beta}_1$  es negativo, se puede afirmar que el tratamiento 1 (quimioterapia) es inferior que el tratamiento 2 (quimio-radioterapia) y, de esta forma, se confirma lo que había mostrado ya la representación gráfica. De hecho, se estima que, en cada visita, el promedio del tamaño del tumor en los sujetos del grupo 2 es 0,5625 cm menor que en los sujetos del grupo 1.

Por otro lado,  $\hat{\beta}_2$  es estimador de  $\beta_2$  y al ser éste el coeficiente de las visitas, se concluye que, para cada grupo, una estimación del promedio de decrecimiento en el diámetro del tumor es de 0,0855 cm en cada visita.

### 3.3.2. Modelo con intercepto y pendiente aleatorios

#### Definición del modelo

El modelo con intercepto aleatorio, que fue introducido con anterioridad, respalda la hipótesis de que las varianzas y covarianzas permanecen constantes a través del tiempo. Sin embargo, en la práctica, es muy raro que estas hipótesis se cumplan. Un modelo más realista es llamado *Modelo con intercepto y pendiente aleatorios*. En este modelo,  $y_{ij}$ , la observación de la variable respuesta para el  $i$ -ésimo sujeto al tiempo  $t_j$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, k$ , tiene la siguiente forma

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_{i1} + u_{i2} t_j + \epsilon_{ij}$$

donde las  $u_{i1} \sim N(0, \sigma_{u1}^2)$ , son independientes e idénticamente distribuidas para  $i = 1, \dots, N$  y son los *interceptos aleatorios*, además las  $u_{i2} \sim N(0, \sigma_{u2}^2)$ , son independientes e idénticamente distribuidas para  $i = 1, \dots, N$  y representan las *pendientes aleatorias*. Debe agregarse que,  $Cov(u_{i1}, u_{i2}) = \sigma_{u1u2}$ ,  $i = 1, \dots, N$ , y  $Cov(u_{i1}, u_{i'2}) = 0$  para  $i \neq i'$ . Los términos del error,  $\epsilon_{ij} \sim N(0, \sigma^2)$ , son también independientes e idénticamente distribuidos y son independientes de  $u_1$  y  $u_2$ . En este modelo,

$$Var(y_{ij}) = Var(u_{i1} + u_{i2} t_j + \epsilon_{ij}) = \sigma_{u1}^2 + 2\sigma_{u1u2} t_j + \sigma_{u2}^2 t_j^2 + \sigma^2 \quad y$$

$$Cov(y_{ij}, y_{ij'}) = Cov(u_{i1} + u_{i2} t_j + \epsilon_{ij}, u_{i1} + u_{i2} t_{j'} + \epsilon_{ij'}) = \\ \sigma_{u1}^2 + \sigma_{u1u2} (t_j + t_{j'}) + \sigma_{u2}^2 t_j t_{j'} \quad \text{para } j \neq j'$$



$$Cov(y_{ij}, y_{i'j'}) = Cov(u_{i1} + u_{i2}t_j + \epsilon_{ij}, u_{i'1} + u_{i'2}t_{j'} + \epsilon_{i'j'}) = 0$$

para  $j \neq j'$  y para  $i \neq i'$

Nótese que en este caso las varianzas y covarianzas dependen del tiempo. La matriz de covarianzas  $\mathbf{V}$  sigue siendo diagonal de bloques, pero ahora el bloque  $\mathbf{V}_0$  tiene una estructura asimétrica a diferencia del método con intercepto aleatorio. La función de log-verosimilitud y la función de log-verosimilitud restringida siguen siendo válidas y los dos métodos de estimación de parámetros pueden ser aplicados.

En el modelo con intercepto y pendientes aleatorios, los coeficientes son interpretados de la misma forma que en el modelo con intercepto aleatorio.

### Ejemplo de aplicación del modelo

**Ejemplo 3.** Con frecuencia los estudios clínicos incluyen sub-estudios que se centran en la calidad de vida relacionada con la salud de los sujetos. Incluso cuando un tratamiento prolongue el tiempo de vida de un sujeto, podría reducir la calidad de su vida -por ejemplo, causarle depresión, limitaciones en la movilidad y dificultades con las actividades diarias. Durante las visitas de seguimiento, los sujetos llenan cuestionarios que respaldan su satisfacción con la terapia. Las puntuaciones sobre estos cuestionarios son registradas, y se define como variable respuesta al porcentaje de satisfacción con la terapia. Supóngase que para los sujetos del ejemplo anterior, en el Cuadro 3.2, se han registrado los siguientes porcentajes de satisfacción en cada una de las visitas.

Sujeto	Grupo	Porcentaje de satisfacción					
		0 meses	3 meses	6 meses	12 meses	18 meses	24 meses
1	1	90	90	85	70	67	63
2	1	87	87	90	95	90	90
3	1	78	78	67	65	63	60
4	1	77	77	65	61	57	56
5	1	78	78	77	77	67	67
6	1	82	82	80	80	73	55
7	1	88	88	83	68	58	58
8	1	95	95	95	90	87	85
9	1	84	84	74	64	60	27
10	1	78	78	71	63	60	60
11	1	91	91	90	84	84	78
12	1	83	83	81	80	81	80
13	2	78	78	67	63	70	70
14	2	85	85	78	63	61	40
15	2	85	85	70	70	70	70
16	2	85	85	78	63	53	13
17	2	85	85	79	44	41	30
18	2	96	96	66	58	40	32
19	2	89	89	63	60	52	45
20	2	95	95	65	55	40	33
21	2	73	73	68	41	52	26
22	2	85	85	75	41	37	33
23	2	74	74	70	64	63	42
24	2	96	96	67	78	74	55

Cuadro 3.2: Porcentajes de satisfacción en cada visita

A partir de estos registros se pueden obtener los perfiles individuales de respuesta, los perfiles medios de respuesta y los diagramas de caja de las Figuras 3.4, 3.5 y 3.6, respectivamente, necesarios para comenzar con el análisis de los datos.

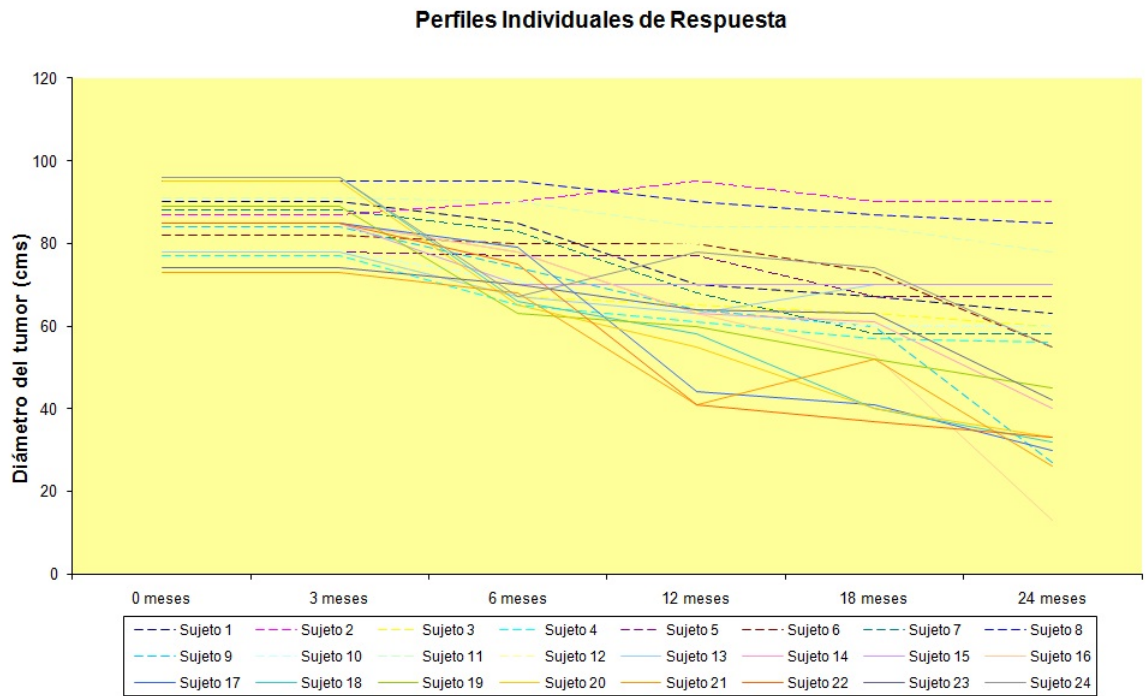


Figura 3.4: Perfiles individuales de respuesta. Líneas punteadas = grupo 1, Líneas continuas = grupo 2

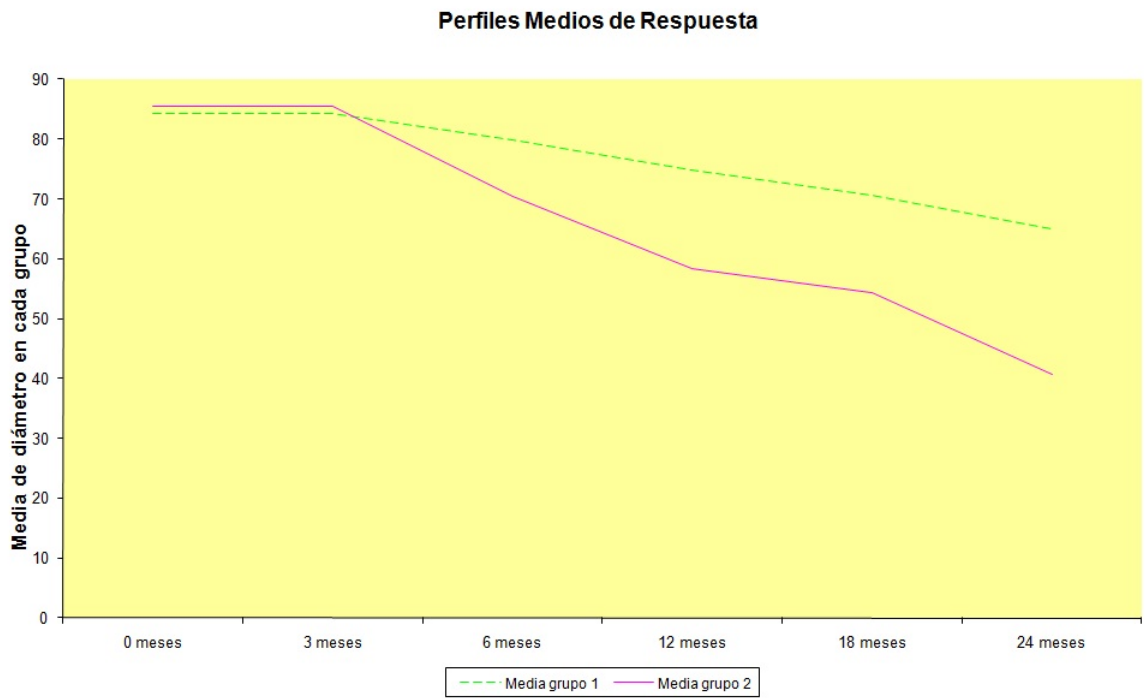


Figura 3.5: Perfiles de Respuesta Media. Línea punteada = grupo 1, Línea continua = grupo 2

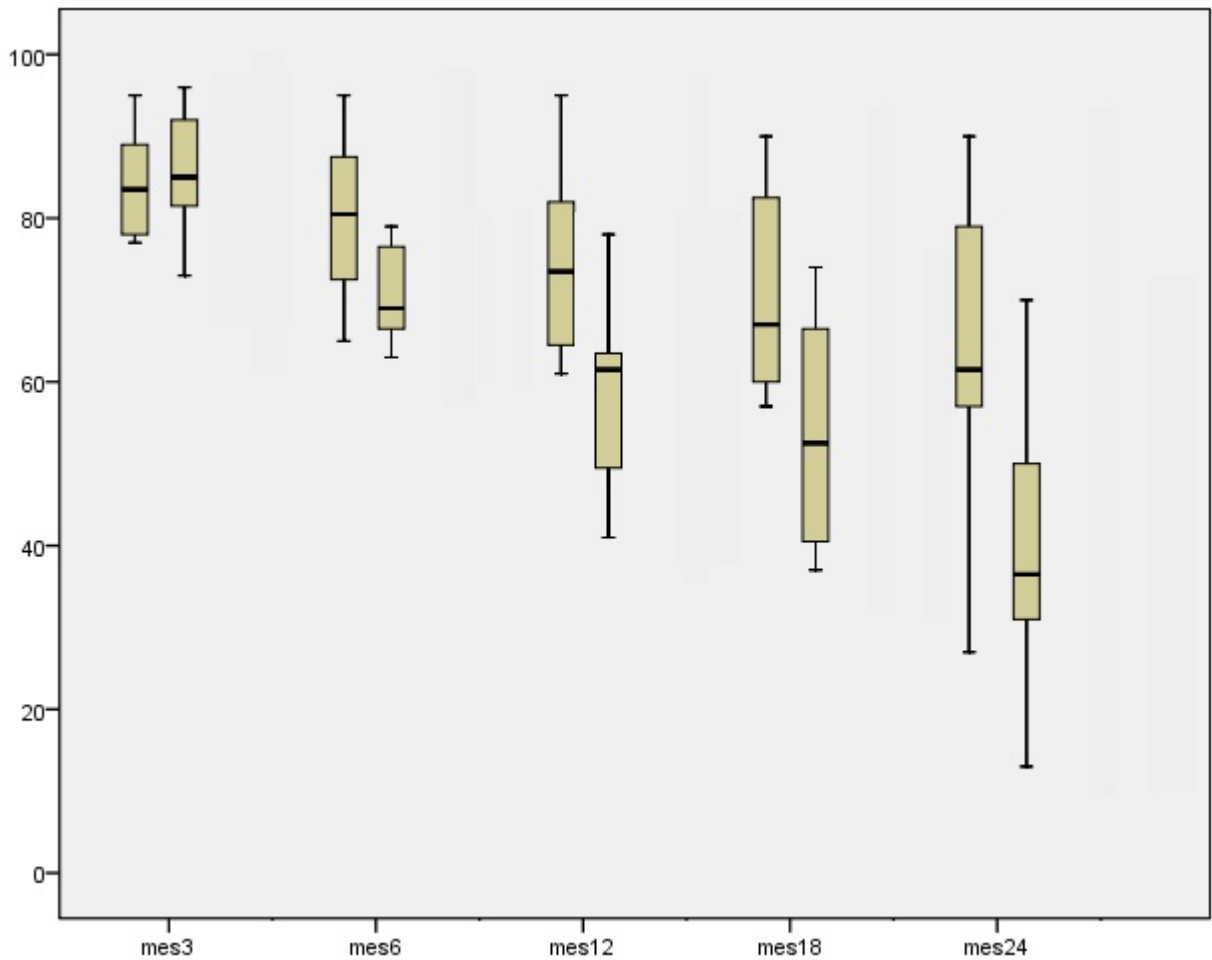


Figura 3.6: Diagramas de caja para cada grupo en cada visita de seguimiento

Recuérdese, del ejemplo 2, que la conclusión, según los reportes gráficos de este estudio, fue que, en promedio, de los dos tratamientos, la quimio-radioterapia reduce de forma más rápida el tamaño de los tumores cancerígenos que la quimioterapia, por lo tanto, se consideró que la quimio-radioterapia era un tratamiento superior. Sin embargo, como se puede ver en las gráficas y diagramas de caja, las tasas de satisfacción de los sujetos en el tratamiento 2 (quimio-radioterapia) son, en promedio, menores que las tasas de los sujetos en el grupo 1 (quimioterapia).

El punto que se quiere hacer notar es que, aunque un tratamiento pueda ser más eficaz, potente o rápido que otro, éste puede provocar una disminución en la calidad de vida de una persona que quizá sea mayor a la que le provocarían otros tratamientos alternos. No obstante, la elección de un tratamiento u otro puede considerarse como una decisión muy personal del paciente.

Para el conjunto de datos presente, la varianza de la respuesta no es constante a través del tiempo, es decir, las observaciones comienzan a estar más dispersas al final del estudio. De esta manera, el modelo con intercepto y pendiente aleatorios será ajustado a estos datos. Como se indicó en la definición del modelo, éste tendrá la siguiente forma

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 t_j + u_{i1} + u_{i2} t_j + \epsilon_{ij}$$

donde  $y_{ij}$  es el porcentaje de satisfacción con la terapia para el  $i$ -ésimo sujeto al tiempo  $t_j$ ,  $i = 1, \dots, 24$ ,  $j = 1, \dots, 5$ ;  $x_{1ij} = x_{1i}$  es el número de grupo del  $i$ -ésimo sujeto (es el mismo para todos los tiempos  $t_j$ );  $x_{2ij}$  es el tamaño del tumor del  $i$ -ésimo sujeto al tiempo  $t_j$ ; y  $t_1 = 3$ ,  $t_2 = 6$ ,  $t_3 = 12$ ,  $t_4 = 18$  y  $t_5 = 24$  (todos en meses). En adición a los coeficientes de regresión, los otros parámetros del modelo son  $Var(u_{i1}) = \sigma_{u_1}^2$ ,  $Var(u_{i2}) = \sigma_{u_2}^2$ ,  $Var(\epsilon_{ij}) = \sigma^2$  y  $Cov(u_{i1}, u_{i2}) = \sigma_{u_1 u_2}$ .

Nuevamente se hará uso del paquete estadístico SAS para obtener las estimaciones de los parámetros por medio de los dos métodos de estimación aquí tratados. A continuación se presentan los resultados más destacados <sup>6</sup>.

Los coeficientes estimados para este modelo son los siguientes:

<b>Método de Máxima Verosimilitud</b>			
<b>Estimador</b>	<b>Efecto</b>	<b>Estimación</b>	<b>Pr &gt; t  (p-value)</b>
$\hat{\beta}_0$	Intercepto	97,0411	<0,0001
$\hat{\beta}_1$	Grupo	-5,6404	0,0297
$\hat{\beta}_2$	Tamaño del tumor	-0,8258	0,5700
$\hat{\beta}_3$	Tiempo	-1,4598	<0,0001

<b>Método de Máxima Verosimilitud Restringida</b>			
<b>Estimador</b>	<b>Efecto</b>	<b>Estimación</b>	<b>Pr &gt; t  (p-value)</b>
$\hat{\beta}_0$	Intercepto	97,0087	<0,0001
$\hat{\beta}_1$	Grupo	-5,6549	0,0376
$\hat{\beta}_2$	Tamaño del tumor	-0,8089	0,5852
$\hat{\beta}_3$	Tiempo	-1,4584	<0,0001

<sup>6</sup>El código en SAS correspondiente a este ejemplo se encuentra en el anexo 3

Se observa como con ambos métodos el  $p$ -value para el estimador  $\hat{\beta}_2$  es mayor que 0.05, lo que indica que la covariable *tamaño del tumor* tiene un efecto insignificante en lo que respecta al nivel de satisfacción con la terapia, por lo tanto, no es de gran importancia en el modelo. Si se remueve esta covariable se obtienen nuevos valores para los estimadores.

Eliminando la covariable tamaño del tumor se vuelven a obtener los estimadores de los coeficientes y de los parámetros. Para el caso de los coeficientes se tiene lo siguiente:

<b>Método de Máxima Verosimilitud</b>			
<b>Estimador</b>	<b>Efecto</b>	<b>Estimación</b>	<b>Pr &gt; t  (p-value)</b>
$\hat{\beta}_0$	Intercepto	94,0481	<0,0001
$\hat{\beta}_1$	Grupo	-5,4140	0,0329
$\hat{\beta}_2$	Tiempo	-1,3924	<0,0001

<b>Método de Máxima Verosimilitud Restringida</b>			
<b>Estimador</b>	<b>Efecto</b>	<b>Estimación</b>	<b>Pr &gt; t  (p-value)</b>
$\hat{\beta}_0$	Intercepto	94,0480	<0,0001
$\hat{\beta}_1$	Grupo	-5,4140	0,0409
$\hat{\beta}_2$	Tiempo	-1,3924	<0,0001

Para la estimación de los parámetros de la covarianza y varianza se obtienen los resultados que a continuación se muestran:

<b>Método de Máxima Verosimilitud</b>		
<b>Estimador</b>	<b>Parámetro</b>	<b>Estimación</b>
$\hat{\sigma}_{u_1}^2$	Intercepto	26,7464
$\hat{\sigma}_{u_1u_2}$	Intercepto/Pendiente	-2,5705
$\hat{\sigma}_{u_2}^2$	Pendiente	0,7105
$\hat{\sigma}^2$	Residual	46,5772

Método de Máxima Verosimilitud Restringida		
Estimador	Parámetro	Estimación
$\hat{\sigma}_{u_1}^2$	Intercepto	31,1667
$\hat{\sigma}_{u_1u_2}$	Intercepto/Pendiente	-2,7686
$\hat{\sigma}_{u_2}^2$	Pendiente	0,7482
$\hat{\sigma}^2$	Residual	46,5770

Con este modelo reducido se puede concluir, ya que todas las estimaciones de  $\underline{\beta}$  son valores significativos. Con ambos métodos se observa que el coeficiente de la covariable *grupo* es negativo y esto prueba que los sujetos que se someten a la quimio-radioterapia tienen una menor calidad de vida comparados con los sujetos que se someten a la quimioterapia. De hecho, el porcentaje aproximado de satisfacción en los sujetos bajo quimio-radioterapia es menor un 5.41 % que los porcentajes de satisfacción del otro grupo en cada visita. También se puede señalar que, en ambos grupos, los porcentajes de satisfacción de los pacientes decrecen, en promedio, un 1.39 % en cada visita.

### 3.3.3. Modelo con Estructura Covarianza Espacial de Potencia para el Error

#### Definición del modelo

Otros modelos usuales para el análisis en estudios con datos longitudinales incluyen modelos de efectos combinados con procesos con Matrices de covarianzas para los errores que tienen estructura espacial de potencia. La observación,  $y_{ij}$ , de la variable respuesta sobre el  $i$ -ésimo sujeto al tiempo  $t_j$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, k$ , tiene la siguiente forma

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_{i1} + u_{i2} t_j + w_i(t_j)$$

donde  $x_{1ij}, \dots, x_{pij}$  son los valores observados de  $p$  covariables de efectos fijos sobre el  $i$ -ésimo sujeto al tiempo  $t_j$ . Además  $u_{i1}$  y  $u_{i2}$  son los coeficientes que corresponden a los términos de *efectos aleatorios*. Se adopta la hipótesis de que las  $u_{i1} \sim N(0, \sigma_{u_1}^2)$ , son independientes e idénticamente distribuidas para  $i = 1, \dots, N$  y las  $u_{i2} \sim N(0, \sigma_{u_2}^2)$ , son independientes e idénticamente distribuidas para  $i = 1, \dots, N$ . Debe agregarse que,  $Cov(u_{i1}, u_{i2}) = \sigma_{u_1u_2}$  para  $i = 1, \dots, N$ , y  $Cov(u_{i1}, u_{i'2}) = 0$  para  $i \neq i'$ . El componente  $w_i$  representa el *error* y es un proceso en tiempo discreto con media y varianza constantes. Este proceso satisface la fórmula recursiva

$$w_i(t_j) = \rho w_i(t_j - 1) + z_i(t_j)$$

Aquí  $\rho$  es un número fijo que satisface  $|\rho| < 1$ , y  $z_i(t_j) \sim N(0, (1 - \rho^2)\sigma^2)$  (también independientes e idénticamente distribuidas) son independientes de  $w_i(t_1)$ , la observación inicial del  $i$ -ésimo sujeto. Se sigue que para todo  $j' > j$ ,  $w_i(t_j)$  y  $z_i(t_{j'})$  son independientes. En efecto, considerando la fórmula recursiva del proceso  $w_i$ , se tiene

$$\begin{aligned} w_i(t_j) &= \rho w_i(t_j - 1) + z_i(t_j) = \rho(\rho w_i(t_j - 2) + z_i(t_j - 1)) + z_i(t_j) = \\ &= \rho^2(\rho w_i(t_j - 3) + z_i(t_j - 2)) + \rho z_i(t_j - 1) + z_i(t_j) = \dots = \\ &= \rho^{t_j - t_1} w_i(t_1) + \rho^{t_j - t_1 - 1} z_i(t_1 + 1) + \dots + \rho z_i(t_j - 1) + z_i(t_j) \dots \dots \dots \textcircled{1} \end{aligned}$$

Y cada término en esta suma es independiente de  $z_i(t_{j'})$ .

Se puede mostrar que el proceso  $w_i$  tiene media 0 y covarianza  $\sigma^2$ ; es decir,  $E(w_i(t_j)) = 0$  y  $E(\{w_i(t_j)\}^2) = \sigma^2$ . Para calcular la matriz covarianza de este proceso, nótese que  $Cov(w_i(t_j), w_i(t_{j'})) = E(w_i(t_j) w_i(t_{j'}))$  para todo  $j' > j$ . Por la ecuación  $\textcircled{1}$ , podemos escribir  $w_i(t_{j'})$  como

$$w_i(t_{j'}) = \rho^{t_{j'} - t_j} w_i(t_j) + \rho^{t_{j'} - t_j - 1} z_i(t_j + 1) + \dots + \rho z_i(t_{j'} - 1) + z_i(t_{j'}) \dots \dots \dots \textcircled{2}$$

Por la independencia de  $w_i(t_j)$  y  $z_i(t_{j''})$  para cualquier  $j'' > j$ ,  $E(w_i(t_j) z_i(t_{j''})) = E(w_i(t_j)) E(z_i(t_{j''})) = 0$ . Por consiguiente, tomando en cuenta la ecuación  $\textcircled{2}$ , la covarianza entre  $w_i(t_j)$  y  $w_i(t_{j'})$  es

$$E[w_i(t_j) w_i(t_{j'})] = \rho^{t_{j'} - t_j} E[(w_i(t_j))^2] = \rho^{t_{j'} - t_j} \sigma^2$$

En consecuencia, la matriz de covarianzas para los errores es una matriz diagonal de bloques con  $k \times k$  bloques como el siguiente

$$\sigma^2 \begin{bmatrix} 1 & \rho^{t_2 - t_1} & \rho^{t_3 - t_1} & \dots & \rho^{t_k - t_1} \\ \rho^{t_2 - t_1} & 1 & \rho^{t_3 - t_2} & \dots & \rho^{t_k - t_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{t_k - t_1} & \rho^{t_k - t_2} & \rho^{t_k - t_3} & \dots & 1 \end{bmatrix}$$

Una matriz de esta forma es llamada Matriz con estructura espacial de potencia, y se dice que el proceso correspondiente tiene una estructura covarianza espacial de potencia.



La matriz de covarianzas  $\mathbf{V}$  de la variable respuesta es una matriz diagonal de bloques de tamaño  $Nk \times Nk$ , formada por  $N$  bloques  $\mathbf{V}_0$  de tamaño  $k \times k$  que tienen entradas diagonales de la siguiente forma

$$Var(y_{ij}) = Var(u_{i1} + u_{i2}t_j + w_i(t_j)) = \sigma_{u_1}^2 + 2\sigma_{u_1u_2}t_j + \sigma_{u_2}^2t_j^2 + \sigma^2$$

donde  $j = 1, \dots, k$ , y las entradas fuera de la diagonal de  $\mathbf{V}_0$  son  $Cov(y_{ij}, y_{ij'}) = Cov[u_{i1} + u_{i2}t_j + w_i(t_j), u_{i1} + u_{i2}t_{j'} + w_i(t_{j'})] = \sigma_{u_1}^2 + \sigma_{u_1u_2}(t_j + t_{j'}) + \sigma_{u_2}^2t_jt_{j'} + \rho^{t_{j'}-t_j}\sigma^2$  para  $j \neq j'$ .

De la misma forma que en el modelo con intercepto y pendientes aleatorios, la varianza y la covarianza de la variable respuesta dependen del tiempo. Sin embargo, a diferencia de los modelos discutidos con anterioridad, en los que no se supone que exista correlación entre los términos del error, en el Modelo con Estructura Covarianza Espacial de Potencia para el Error existe una débil dependencia entre los términos de error para el mismo individuo. En valor absoluto, la covarianza entre los términos del error decae exponencialmente conforme el lapso de tiempo incrementa.

Para este modelo, se conserva la validez de los métodos de verosimilitud y de verosimilitud restringida para la estimación de parámetros y los coeficientes del modelo admiten la misma interpretación que en los modelos anteriores.

## Ejemplo de aplicación del modelo

**Ejemplo 4.** Considérense los datos del ejemplo 3. El objetivo es ajustar el modelo de efectos combinados con Estructura Covarianza Espacial de Potencia para el Error. El modelo es de la forma

$$y_{ij} = \beta_0 + \beta_1x_{1ij} + \beta_2x_{2ij} + \beta_3t_j + u_{i1} + u_{i2}t_j + w_i(t_j)$$

donde  $y_{ij}$  denota el porcentaje de satisfacción con la terapia para el  $i$ -ésimo sujeto al tiempo  $t_j$ ;  $x_{1ij} = x_{1i}$  es el grupo (1 ó 2) del  $i$ -ésimo sujeto;  $x_{2ij}$  es el tamaño del tumor del  $i$ -ésimo sujeto en la  $j$ -ésima medición;  $t_1 = 3, t_2 = 6, t_3 = 12, t_4 = 18$  y  $t_5 = 24$  (todos en meses),  $i = 1, \dots, 24$ , y  $j = 1, \dots, 5$ . Las variables aleatorias  $u_{i1}$  y  $u_{i2}$  son el intercepto y la pendiente aleatorios respectivamente. Los términos del error  $w_i(t_j)$  tienen media cero y una matriz de covarianzas de tamaño  $120 \times 120$  conformada por 24 bloques en la diagonal y cada uno de tamaño  $5 \times 5$  como el que se presenta a continuación:

$$\sigma^2 \begin{bmatrix} 1 & \rho^3 & \rho^9 & \rho^{15} & \rho^{21} \\ \rho^3 & 1 & \rho^6 & \rho^{12} & \rho^{18} \\ \rho^9 & \rho^6 & 1 & \rho^6 & \rho^{12} \\ \rho^{15} & \rho^{12} & \rho^6 & 1 & \rho^6 \\ \rho^{21} & \rho^{18} & \rho^{12} & \rho^6 & 1 \end{bmatrix}$$

Los parámetros de este modelo son  $\beta_0, \beta_1, \beta_2, \beta_3, \sigma_{u_1}^2, \sigma_{u_2}^2, \sigma_{u_1 u_2}, \sigma^2$  y  $\rho$ .

Para obtener las estimaciones correspondientes a dichos parámetros se procede de nuevo en SAS. Debe recordarse que, para este modelo, es factible utilizar los mismos métodos de estimación que se utilizaron en los ejemplos anteriores, por lo tanto se procederá de manera análoga <sup>7</sup>.

Los valores estimados que proporcionan los dos métodos para los coeficientes de la regresión (parámetros  $\beta_0, \beta_1, \beta_2, \beta_3$ ) son los siguientes:

<b>Método de Máxima Verosimilitud</b>			
<b>Estimador</b>	<b>Efecto</b>	<b>Estimación</b>	<b>Pr &gt; t  (p-value)</b>
$\hat{\beta}_0$	Intercepto	97,0409	<0,0001
$\hat{\beta}_1$	Grupo	-5,6402	0,0297
$\hat{\beta}_2$	Tamaño del tumor	-0,8260	0,5700
$\hat{\beta}_3$	Tiempo	-1,4599	<0,0001

<b>Método de Máxima Verosimilitud Restringida</b>			
<b>Estimador</b>	<b>Efecto</b>	<b>Estimación</b>	<b>Pr &gt; t  (p-value)</b>
$\hat{\beta}_0$	Intercepto	97,0087	<0,0001
$\hat{\beta}_1$	Grupo	-5,6549	0,0376
$\hat{\beta}_2$	Tamaño del tumor	-0,8089	0,5872
$\hat{\beta}_3$	Tiempo	-1,4584	<0,0001

Como en el Modelo con intercepto y pendiente aleatorios, se observa que la covariable *tamaño del tumor* no tiene un nivel de significancia ni del 5 %, por lo tanto, puede decirse que es insignificante y retirarse del modelo para obtener un modelo reducido.

Con el modelo reducido se obtiene, para los coeficientes:

<sup>7</sup>El código en SAS correspondiente a este ejemplo se encuentra en el anexo 3

Método de Máxima Verosimilitud			
Estimador	Efecto	Estimación	Pr > t  (p-value)
$\hat{\beta}_0$	Intercepto	93,7928	<0,0001
$\hat{\beta}_1$	Grupo	-5,0491	0,0380
$\hat{\beta}_2$	Tiempo	-1,3979	<0,0001

Método de Máxima Verosimilitud Restringida			
Estimador	Efecto	Estimación	Pr > t  (p-value)
$\hat{\beta}_0$	Intercepto	93,7946	<0,0001
$\hat{\beta}_1$	Grupo	-5,0514	0,0467
$\hat{\beta}_2$	Tiempo	-1,3980	<0,0001

Y para los parámetros de la covarianza y varianza se obtienen las siguientes estimaciones:

Método de Máxima Verosimilitud		
Estimador	Parámetro	Estimación
$\hat{\sigma}_{u_1}^2$	Intercepto	38,0795
$\hat{\sigma}_{u_1u_2}$	Intercepto/Pendiente	-3,4003
$\hat{\sigma}_{u_2}^2$	Pendiente	0,7555
$\hat{\rho}$	Número $ p  < 1$	-0,7726
$\hat{\sigma}^2$	Residual	47,0917

Método de Máxima Verosimilitud Restringida		
Estimador	Parámetro	Estimación
$\hat{\sigma}_{u_1}^2$	Intercepto	42,2323
$\hat{\sigma}_{u_1u_2}$	Intercepto/Pendiente	-3,6026
$\hat{\sigma}_{u_2}^2$	Pendiente	0,7944
$\hat{\rho}$	Número $ p  < 1$	-0,7710
$\hat{\sigma}^2$	Residual	47,0152

Nótese que la estimación de los parámetros en este modelo es muy parecida a la que se produjo con el modelo anterior. Y la interpretación de los coeficientes ajustados es la misma para ambos modelos.

### 3.3.4. Modelo de regresión logística con intercepto aleatorio

#### Definición del modelo

Si la respuesta  $y_{ij}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, k$  para el  $i$ -ésimo sujeto al tiempo  $t_j$  es binaria (por lo tanto, puede ser representada como 0 y 1) puede utilizarse un modelo de regresión logística con intercepto aleatorio. Se denotará con  $\pi_{ij}(u) = \Pr(y_{ij} = 1 | u)$  la probabilidad condicional de que  $y_{ij}$  sea igual a 1 dada una variable aleatoria  $u$ . El cociente

$$\frac{\pi_{ij}(u)}{1 - \pi_{ij}(u)} = \frac{\Pr(y_{ij} = 1 | u)}{\Pr(y_{ij} = 0 | u)}$$

es llamado *la oportunidad condicional a favor de que  $y_{ij} = 1$ , dado  $u$* . Una transformación *logit* de  $\pi_{ij}(u)$  es el logaritmo natural de las oportunidades a favor de  $y_{ij} = 1$  condicionadas a  $u$  -esto es,

$$\log \text{it}(\pi_{ij}(u)) = \ln \left( \frac{\pi_{ij}(u)}{1 - \pi_{ij}(u)} \right)$$

El modelo de regresión logística con intercepto aleatorio tiene la forma

$$\log \text{it}(\pi_{ij}(u)) = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i$$

donde las  $u_i \sim N(0, \sigma_u^2)$ , son independientes e idénticamente distribuidas para  $i = 1, \dots, N$  y son los *interceptos aleatorios*. De manera equivalente, se puede escribir el modelo como sigue

$$\pi_{ij}(u_i) = \frac{\exp\{\beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i\}}{1 + \exp\{\beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i\}}$$

## Estimación de parámetros

Los parámetros del modelo son  $\beta_0, \dots, \beta_{p+1}$  y  $\sigma_u^2$  y es muy común usar el método de máxima verosimilitud para hacer su estimación. Para una  $u_i$  dada,  $i = 1, \dots, N$ , la distribución de  $y_{ij}$  es Bernoulli con parámetro  $\pi_{ij}(u_i)$  definido anteriormente. Por lo tanto, la función de verosimilitud condicional es

$$\begin{aligned} L(\beta_0, \dots, \beta_{p+1} | u_1, \dots, u_N) &= \prod_{i=1}^N \prod_{j=1}^k (\pi_{ij}(u_i))^{y_{ij}} (1 - \pi_{ij}(u_i))^{1-y_{ij}} = \\ &= \prod_{i=1}^N \prod_{j=1}^k \left[ \frac{\exp\{\beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i\}}{1 + \exp\{\beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i\}} \right]^{y_{ij}} \times \\ &= \prod_{i=1}^N \prod_{j=1}^k \left[ \frac{1}{1 + \exp\{\beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i\}} \right]^{1-y_{ij}} = \\ &= \prod_{i=1}^N \prod_{j=1}^k \left[ \frac{\exp\{(\beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i) y_{ij}\}}{1 + \exp\{\beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} t_j + u_i\}} \right] \end{aligned}$$

Para obtener la función de verosimilitud convencional, se integra esta función sobre todos los posibles valores de  $u_1, \dots, u_N$ , los cuales son variables aleatorias independientes distribuidas  $N(0, \sigma_u^2)$ :

$$\begin{aligned} L(\beta_0, \dots, \beta_{p+1}, \sigma_u^2) &= (2\pi\sigma_u^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma_u^2} \sum_{i=1}^N \sum_{j=1}^k (\beta_0 + \dots + \beta_{p+1} t_j) y_{ij}\right\} \times \\ &= \prod_{i=1}^N \int_{-\infty}^{\infty} \left[ \frac{\exp\left\{\sum_{j=1}^k (u_i y_{ij}) - \frac{u_i^2}{2\sigma_u^2}\right\}}{\prod_{j=1}^k (1 + \exp\{\beta_0 + \dots + \beta_{p+1} t_j + u_i\})} \right] du_i \end{aligned}$$

Los estimadores  $\hat{\beta}_0, \dots, \hat{\beta}_{p+1}$  y  $\hat{\sigma}_u^2$  son la solución numérica al problema de maximización para esta función.

## Interpretación de los coeficientes

La interpretación de los coeficientes de la regresión  $\beta_0, \dots, \beta_{p+1}$  es análoga a la que se hace en el modelo de regresión logística ordinaria. Por conveniencia, se hace la siguiente notación

$$\pi(x_m) = \frac{\exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \dots + \beta_p x_p + \beta_{p+1} t + u\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \dots + \beta_p x_p + \beta_{p+1} t + u\}}$$

donde  $m = 1, \dots, p+1$ . Entonces el coeficiente  $\beta_m$  de la regresión con exponentes representa el cociente de oportunidad condicionada por  $u$ , cuando  $x_m$  incrementa unitariamente y se mantienen fijas las demás covariables. Es decir,

$$\exp\{\beta_m\} = \frac{\pi(x_m + 1) / (1 - \pi(x_m + 1))}{\pi(x_m) / (1 - \pi(x_m))}$$

Por lo tanto,  $100(\exp\{\beta_m\} - 1)\%$  es el porcentaje de cambio correspondiente a la oportunidad.

En el caso de una covariable categórica con  $l$  niveles, los coeficientes  $\beta_1, \dots, \beta_{l-1}$ , corresponden a las variables ficticias para los niveles del 1 al  $l-1$  y, entonces la cantidad  $100(\exp\{\beta_a - \beta_b\})\%$ ,  $a, b = 1, \dots, l-1$ , representa el porcentaje de oportunidad condicional para sujetos con la covariable dicha en el nivel  $a$  comparados con los sujetos en el nivel  $b$ , conservando fijas las otras variables. El porcentaje de oportunidad condicional para sujetos en el nivel  $a$  comparados con sujetos en el nivel  $l$  es  $100(\exp\{\beta_a\})\%$ .

### Ejemplo de aplicación del modelo

**Ejemplo 5.** La osteoporosis es una enfermedad metabólica ósea en la cual todos los huesos del cuerpo se van haciendo paulatinamente porosos y pueden fracturarse muy fácilmente. Esta enfermedad es más común en mujeres mayores de 50 años de edad. Se está probando un medicamento que supuestamente ayuda a la reconstrucción del hueso y para ello se realiza un ensayo con 20 mujeres que padecen la enfermedad. Para cada sujeto en el estudio se registran los siguientes datos: edad hasta el momento de la primera visita (en años), consumo de suplemento de calcio (si = 1, no = 0) e historial familiar con osteoporosis (si = 1, no = 0). Durante cada visita de seguimiento se aplica al sujeto una prueba de densidad del hueso para saber si todavía hay presencia de osteoporosis o si la densidad aumentó tanto que la hizo desaparecer. La variable respuesta es la presencia de osteoporosis (si = 1, no = 0), y los datos recogidos se presentan a continuación:

Sujeto	Edad	Calcio	Historia	Presencia de osteoporosis			
				3 meses	9 meses	12 meses	18 meses
1	76	0	1	1	1	1	1
2	57	1	1	1	1	0	0
3	58	0	1	1	1	1	0
4	62	1	0	1	0	0	0
5	60	1	0	0	0	0	0
6	58	0	1	1	0	1	1
7	52	1	0	0	1	0	0
8	74	0	1	1	0	1	0
9	51	0	0	0	1	0	0
10	56	1	0	0	1	0	0
11	75	0	1	1	1	1	1
12	63	1	0	1	0	0	0
13	67	1	1	0	1	0	0
14	68	0	0	1	1	0	0
15	56	1	0	1	0	1	0
16	62	1	0	1	0	1	1
17	60	1	1	0	0	0	0
18	61	1	1	1	1	0	0
19	54	1	0	1	0	0	0
20	53	0	0	1	1	0	0

El modelo de regresión logística con intercepto aleatorio para estos datos es

$$\pi_{ij}(u_i) = \frac{\exp\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 t_j + u_i\}}{1 + \exp\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 t_j + u_i\}}$$

donde  $\pi_{ij}(u_i)$  es la probabilidad condicional de que el  $i$ -ésimo sujeto tenga osteoporosis en la  $j$ -ésima visita dado  $u_i$ ,  $i = 1, \dots, 20$ ,  $j = 1, \dots, 4$ ;  $x_{1i}$ ,  $x_{2i}$  y  $x_{3i}$  son, respectivamente, la edad, la ingesta de calcio, y el historial familiar de osteoporosis para el  $i$ -ésimo sujeto;  $t_1 = 3$ ,  $t_2 = 9$ ,  $t_3 = 12$  y  $t_4 = 18$  (todos en meses); y  $u_1$  es el intercepto aleatorio.

Los parámetros de este modelo son  $\beta_0, \dots, \beta_4$ , y  $\sigma_u^2$ , y al estimarlos en SAS con el método de máxima verosimilitud<sup>8</sup> se obtienen los siguientes datos:

<sup>8</sup>El código en SAS se encuentra en el anexo 3

Método de Máxima Verosimilitud			
Estimador	Efecto	Estimación	Pr > t  (p-value)
$\hat{\beta}_0$	Intercepto	-0,9445	0,7544
$\hat{\beta}_1$	Edad	0.0545	0,2692
$\hat{\beta}_2$	Calcio	-1,4343	0,0309
$\hat{\beta}_3$	Historial	0.5401	0,4273
$\hat{\beta}_4$	Tiempo	-0,1926	0,0012

Con estos datos, puede observarse que únicamente la ingesta de Calcio y el Tiempo son variables significativas (considerando un nivel de significancia del 5%). Entonces podemos eliminar del análisis las variables Edad e Historial, procediendo de esta forma se obtiene el siguiente modelo reducido:

Método de Máxima Verosimilitud			
Estimador	Efecto	Estimación	Pr > t  (p-value)
$\hat{\beta}_0$	Intercepto	2,7129	0,0031
$\hat{\beta}_1$	Calcio	-1,7626	0,0060
$\hat{\beta}_2$	Tiempo	-0,1825	0,0014

Y para la covarianza se obtiene que

Método de Máxima Verosimilitud		
Estimador	Parámetro	Estimación
$\hat{\sigma}_u^2$	Intercepto	0,3225

Entonces se puede deducir que las probabilidades en favor de la osteoporosis, condicionadas por  $u$ , para los sujetos que ingieren suplemento de Calcio es únicamente del  $100 \exp \{-1,7626\} = 17\%$  del total de sujetos que no lo ingieren. Además, las probabilidades en favor de la osteoporosis cambian cada mes en un  $100 (\exp \{-0,1825\} - 1) = -16,68\%$ , es decir, que disminuye aproximadamente un 17%.



## 3.4. Manejo de Observaciones Faltantes

Ocurre con frecuencia, en la práctica, que dentro de un conjunto de datos longitudinales haya observaciones faltantes. Estas observaciones faltantes pueden deberse a razones de diversa índole, por ejemplo, si por algún motivo, un sujeto no asiste a una de las visitas de seguimiento esto provocará la observación faltante y, en este caso, será llamada *valor intermitente faltante*; por otro lado, si un sujeto abandona el estudio, entonces, a partir de la última visita de seguimiento a la que asistió, habrá observaciones faltantes para este sujeto y el caso será llamado *deserción*. Un conjunto de datos con observaciones faltantes es llamado un conjunto de datos *desbalanceado o incompleto*.

### 3.4.1. Mecanismo para Observaciones Faltantes

En los modelos de datos longitudinales que carecen de algunas observaciones pueden usarse varios enfoques al momento del análisis, pero cuál de ellos elegir depende del mecanismo del valor faltante. Se denotará por  $y_{i(obs)}$  a las mediciones respuesta observadas en el  $i$ -ésimo sujeto, y por  $y_{i(fal)}$  a los valores faltantes -no las observaciones desaparecidas, sino las observaciones faltantes de realizar.

Sea  $r_i$  un conjunto de variables aleatorias indicadoras binarias que indican si  $y_{ij}$  es un valor observado o un valor faltante,  $j = 1, \dots, k$ . Entonces los valores faltantes caen dentro una de las siguientes tres categorías:

- Faltante completamente por azar, si  $r_i$  es independiente de  $y_{i(obs)}$  y de  $y_{i(fal)}$ .
- Faltante por azar, si  $r_i$  depende de  $y_{i(obs)}$  pero no de  $y_{i(fal)}$ .
- Faltante no por azar, si  $r_i$  depende de  $y_{i(fal)}$ , y puede depender de  $y_{i(obs)}$ .

**a) Faltante completamente por azar.** Para los valores que pertenecen a esta categoría, la ausencia de la observación no está relacionada con la respuesta que el sujeto está teniendo con el tratamiento. En otras palabras, podría suceder que cualquier sujeto en el estudio falte a una visita de seguimiento o abandone el estudio por completo. Por ejemplo, si debido a una emergencia familiar un sujeto perdió una visita que estaba programada, y en ese momento no pudo ser localizado por los investigadores para reprogramar la cita, la medición prevista para esta fecha no será registrada. Como el sujeto permanecerá en el estudio, la observación faltante será un valor intermitente faltante que ocurre completamente por azar. En cambio, si el sujeto sufre muerte accidental o necesita mudarse de ciudad por cuestiones laborales, por ejemplo, es evidente que abandonará el estudio y en este caso se trata de una deserción completamente por azar.

**b) Faltante por azar.** Un valor intermitente o una deserción son faltantes al azar si dependen del historial de mediciones observadas (dependen de las observaciones). Por ejemplo, supóngase que un sujeto entra a un ensayo clínico en el que se está probando un medicamento nuevo para adelgazar, pero después de dos visitas de seguimiento, la reducción en el paciente es mínima y el sujeto falta a la siguiente visita debido a la depresión o a la vergüenza, en este caso estaría tratándose de un valor intermitente faltante por azar. Ahora supóngase que una joven participa en un ensayo en el que se está probando un nuevo medicamento para tratar el desorden psicológico que conlleva a la anorexia. El medicamento no ayuda a que disminuya su falta de deseos por comer debido al temor a ser obesa, y consecuentemente continua perdiendo peso. Después de varias visitas de seguimiento, los investigadores consideran que es antiético y peligroso permitir que ella permanezca en el estudio y prefieren que ella lo abandone. Esta es una situación en la que su retiro se considera una deserción que conlleva a observaciones faltantes por azar, pues está relacionada con el historial de observaciones.

**c) Faltante no por azar.** Los valores intermitentes y las deserciones que no se deben al azar surgen cuando la ausencia está relacionada con un cambio no registrado en la variable respuesta que causa que ésta sea demasiado alta o demasiado baja. Por ejemplo, en el caso del sujeto que participa en un estudio para probar un medicamento nuevo que, de ser aprobado, ayudará a eliminar el sobrepeso. El sujeto muestra una mejora importante en las dos primeras visitas de seguimiento, pero repentinamente, antes de la tercera visita de seguimiento, engorda rápidamente y a causa de su depresión decide no asistir a su visita. Entonces la medición de la respuesta para esta visita será faltante y no se deberá al azar. En otro estudio que prueba un tratamiento para la anorexia, un sujeto comienza a ganar peso al inicio, pero entre dos visitas recae y comienza a perder peso nuevamente de manera tan rápida que debe ser hospitalizada y es preferible, que por su salud, se retire del estudio. En este caso el resultado es una deserción que no se debe al azar.

### **3.4.2. Distinción entre faltantes completamente al azar, al azar y no debido al azar**

Discernir a que categoría pertenece una observación faltante puede convertirse en un problema en algunos casos. Cuando se trata de valores intermitentes, el problema no es tan difícil de resolver, ya que en la posterior visita de seguimiento puede cuestionarse al sujeto cuál fue la razón por la que faltó a la anterior visita y esto daría la pauta para poder clasificar la observación faltante (claro que esto funciona sólo suponiendo que el sujeto dice la verdad).

En el caso de las deserciones no es tan fácil conocer la razón exacta para el abandono

si la persona no puede ser contactada por los investigadores, a menos que se trate de una deserción que ha sido sugerida por los mismos investigadores como pasa en el caso de una deserción que genera observaciones faltantes por azar.

Para comprender de manera intuitiva como clasificar las deserciones, considérese primero una situación hipotética en la que un sujeto progresa bien dentro del ensayo y de repente abandona el estudio. Es poco probable que la decisión de abandono se haya producido completamente por azar, a pesar de que es posible que lo sea. Es más viable pensar que el sujeto desertó porque continuaba mejorando y se sentía bien (una deserción por azar) o por otro lado, porque de repente se sintió mucho peor (una deserción no debida al azar). Sin embargo, no es posible afirmar que se puede aplicar cualquiera de estos casos porque la razón real de la deserción es desconocida.

Ahora supóngase que al sujeto que participaba en el estudio, no le había ido tan bien a lo largo de las visitas de seguimiento como se esperaba y en determinado momento deserta. Es más probable que la deserción no se deba completamente al azar, pero es difícil determinar si es una deserción por azar (la persona decide abandonar el tratamiento porque su salud se deterioraba gradualmente) o una deserción que no se debió al azar (repentinamente el sujeto se sintió mucho mejor).

Si de forma continua, un sujeto no mejora ni empeora durante el tiempo que lleva el estudio y éste deserta, es mucho más probable pensar que se trata de una deserción completamente por azar.

En consecuencia, basándose en el historial de respuestas observadas de un desertor, es posible distinguir entre una deserción completamente por azar y los otros dos casos. Lo que no es posible es clasificar una deserción debida al azar de una que no lo fue. No obstante, en ambos casos la decisión de desertar está basada en mediciones no disponibles para los investigadores.

### **3.4.3. Elección del modelo de efectos combinados adecuado para el conjunto de datos desbalanceados**

#### **Observaciones faltantes prescindibles e imprescindibles**

Las observaciones de las categorías completamente por azar y por azar pueden ser ignoradas y por ello son llamadas *prescindibles*, mientras los valores que no se deben al azar no pueden ser ignorados y se conocen como *imprescindibles* o *informativos*. La razón de esta terminología es la siguiente: Cuando se quiere determinar qué modelo de efectos combinados se debe emplear para datos desbalanceados, los métodos de máxima verosimilitud y máxima verosimilitud restringida son aplicados para estimar los parámetros. Los valores faltantes debidos completamente al azar y faltantes al azar

no aparecen en la función de verosimilitud, por lo tanto, pueden ser ignorados. En cambio, la función de verosimilitud sí depende de las observaciones faltantes que no se deben al azar, y por ende, no pueden ser ignoradas. Para confirmar esto, nótese que la contribución del  $i$ -ésimo sujeto a la función de verosimilitud es

$$f(r_i, y_{i(obs)}) = \int f(r_i, y_{i(obs)}, y_{i(fal)}) dy_{i(fal)} = \int f(r_i | y_{i(obs)}, y_{i(fal)}) f(y_{i(obs)}, y_{i(fal)}) dy_{i(fal)}$$

Ahora se analiza que sucede con esta contribución en cada una de las tres categorías:

- Si los valores están en la categoría faltantes completamente por azar, entonces

$$f(r_i | y_{i(obs)}, y_{i(fal)}) = f(r_i)$$

Por lo tanto,  $f(r_i, y_{i(obs)}) = f(r_i) \int f(y_{i(obs)}, y_{i(fal)}) dy_{i(fal)} = f(r_i) f(y_{i(obs)}) \propto f(y_{i(obs)})$

Ya que  $f(r_i)$  no depende de los parámetros del modelo, puede ser ignorado para efectos de resolver el problema de maximización de la función de verosimilitud.

- Si los valores están en la categoría faltantes por azar, entonces

$$f(r_i | y_{i(obs)}, y_{i(fal)}) = f(r_i | y_{i(obs)})$$

Por lo tanto,  $f(r_i, y_{i(obs)}) = f(r_i | y_{i(obs)}) \int f(y_{i(obs)}, y_{i(fal)}) dy_{i(fal)} = f(r_i | y_{i(obs)}) f(y_{i(obs)}) \propto f(y_{i(obs)})$

Ya que  $f(r_i | y_{i(obs)})$  es constante con respecto a los parámetros del modelo.

- Si los valores faltantes no se deben al azar, entonces  $f(r_i, y_{i(obs)})$  no se puede reducir más. De este modo, la función de verosimilitud depende de las observaciones faltantes, con lo cual no pueden ser ignoradas para propósitos de su maximización.

## Gestión de Conjuntos de Datos Desbalanceados

Existen varias formas relativamente simples de manejar las observaciones faltantes una vez que se ha elegido el modelo de efectos combinados adecuado para el conjunto de datos desequilibrado:

- El mejor enfoque que se puede adoptar si todos los valores faltantes son prescindibles es trabajar con estos conjuntos de datos desbalanceados y usar los métodos de máxima verosimilitud y máxima verosimilitud restringida para la estimación de los parámetros. Sin embargo, para valores faltantes no debidos al azar (imprescindibles), esta estrategia conlleva a la obtención de estimadores sesgados.
- Se puede considerar la eliminación de todos los casos *incompletos*, es decir, la eliminación de los datos de todos los sujetos que tengan al menos una observación faltante en su historial, y de esta forma analizar los resultados de los datos *ya balanceados*. Este tratamiento de los datos es llamado *método de eliminación de casos*, y puede ser el más viable cuando el número de casos incompletos no es muy grande y no hay ninguna estructura de los datos que faltan (son observaciones faltantes completamente por azar). De lo contrario, el método origina estimadores sesgados.
- Otra opción es imputar los valores faltantes y trabajar con el conjunto de datos *completo o balanceado* resultante. La *imputación* es la sustitución de los valores faltantes en un conjunto de datos incompleto o desbalanceado por otros. Con este método se obtienen estimadores sesgados para los parámetros.

## Procedimientos de Imputación

Existen dos tipos de imputación comúnmente usados para valores intermitentes faltantes. Si  $y_{ij}$  es faltante, puede ser reemplazada por un valor  $\hat{y}_{ij}$  que es

- La respuesta media para el  $i$ -ésimo sujeto:

$$\hat{y}_{ij} = \frac{\sum_{j=1}^k r_{ij} y_{ij}}{\sum_{j=1}^k r_{ij}}$$

donde  $r_{ij} = 1$  si  $y_{ij}$  es observada, y  $r_{ij} = 0$  si  $y_{ij}$  es faltante. Este tipo de imputación es llamada *Imputación con media-sujeto*.

- La respuesta media para la  $j'$ -ésima ocasión

$$\hat{y}_{ij'} = \frac{\sum_{j=1}^k r_{ij'} y_{ij'}}{\sum_{j=1}^k r_{ij'}}$$

donde  $r_{ij'} = 1$  si  $y_{ij'}$  es observada, y  $r_{ij'} = 0$  si  $y_{ij'}$  es faltante. Este tipo de imputación es llamada *Imputación con media-ocasión*.

La forma más simple de imputar los valores faltantes de un conjunto de datos debidos a la deserción de los sujetos es reemplazar todos los valores faltantes por el valor de la última observación registrada para el sujeto correspondiente. Es decir, si un sujeto deserta después de la  $j'$ -ésima visita, se reemplazarán las respuestas faltantes  $y_{i,j'+1}, y_{i,j'+2}, \dots, y_{i,k}$  por  $y_{ij'}$ . Este método es conocido como *Imputación con la última observación*.

**Observación 1.** Hay varias formas de manejar las observaciones faltantes de una variable binaria:

- Ignorar los valores faltantes y ajustar un modelo de regresión logística con intercepto aleatorio para el conjunto de datos incompletos.
- Remover todos los casos incompletos y ajustar un modelo logístico para conjuntos de datos completos o balanceados.
- Imputar los valores faltantes y ajustar el modelo logístico a los datos imputados. Dos formas simples de imputar los valores intermitentes faltantes son:
  1. Reemplazar un valor faltante por la *moda del sujeto*, que es la observación más frecuente (0 ó 1) para el sujeto.
  2. Reemplazar un valor faltante por la *moda de la ocasión*, que es la observación más frecuente (0 ó 1) para la ocasión.

Para una deserción, es más fácil usar la técnica de Imputación con la última observación.

**Observación 2.** En algunos ensayos clínicos, faltar a una visita o desertar resulta en valores faltantes para la variable respuesta y para las covariables. En estos casos puede ser aplicado el mismo enfoque descrito previamente para manejar los valores faltantes. En el método de imputación, las observaciones faltantes de la variable respuesta y las de las covariables son imputadas.

## Ejemplo del manejo de las observaciones faltantes

**Ejemplo 6.** Catorce personas sordas que fueron beneficiadas con implantes cocleares (dispositivos internos para mejorar la audición) fueron monitoreadas por 3 años. La variable respuesta registrada fue un porcentaje de entendimiento de la palabra evaluado por un panel de investigadores. Tres de los sujetos beneficiados representan observaciones faltantes porque abandonaron el estudio en algún momento a partir del cual

dejaron de asistir a las visitas de seguimiento y otros dos perdieron sólo una de las visitas de seguimiento. Los datos desbalanceados o incompletos se presentan en la siguiente tabla:

Sujeto	Porcentaje de entendimiento de la palabra				
	3 meses	6 meses	12 meses	24 meses	36 meses
1	46	57	67	68	75
2	24	35	-	-	-
3	30	45	58	72	82
4	22	-	48	66	76
5	18	15	14	-	-
6	35	67	77	80	86
7	27	47	50	55	67
8	12	31	41	47	52
9	55	76	-	-	-
10	18	39	50	67	78
11	10	25	28	33	33
12	35	44	67	73	84
13	22	45	-	68	78
14	35	55	70	75	83

Para ilustrar como se lleva a cabo el manejo de las observaciones faltantes en un ensayo clínico, se aplicarán los tres métodos que se discutieron con anterioridad a manera de que, al manipular las observaciones, se pueda lograr ajustar un modelo de efectos combinados. Por simplicidad, el procedimiento que se utilizará, permitirá ajustar la versión más sencilla de un modelo con intercepto aleatorio y es la siguiente:

$$y_{ij} = \beta_0 + \beta_1 t_j + u_i + \epsilon_{ij}$$

para  $i = 1, \dots, 14$ ,  $j = 1, \dots, 5$ , donde  $y_{ij}$  es el porcentaje registrado para el  $i$ -ésimo sujeto en la  $j$ -ésima visita,  $t_1 = 3$ ,  $t_2 = 6$ ,  $t_3 = 12$ ,  $t_4 = 24$  y  $t_5 = 36$  (todos en meses) y los interceptos aleatorios  $u_i \sim N(0, \sigma_u^2)$  (i.i.d) y son independientes de los errores aleatorios  $\epsilon_{ij} \sim N(0, \sigma^2)$  (i.i.d.).

El primer método que se revisó para el tratamiento de datos faltantes indica que se puede trabajar con los datos desbalanceados y usar los métodos de máxima verosimilitud y máxima verosimilitud restringida para la estimación de los parámetros. Además se declara que este enfoque funciona mejor cuando las observaciones faltantes son imprescindibles (debidas al azar o completamente al azar), de lo contrario se obtendrán estimadores sesgados.

Supóngase que al cuestionar a los dos sujetos a los que les hace falta una observación, acerca de los motivos que los llevaron a no asistir a su visita se determina que sus observaciones faltantes se debieron completamente al azar. Además, sin adoptar este

supuesto, nótese en la tabla de los datos, que las observaciones de uno de los sujetos muestran que en sus dos visitas anteriores ha habido mejoría, lo cual haría suponer que su falta fue debida al azar (debida a su historial) y la del otro sujeto completamente al azar ya que su historial clínico aún es muy corto en el momento de la falta. De cualquier modo se permite determinar que ambas observaciones faltantes son imprescindibles.

El caso de los tres sujetos que desertaron del ensayo es distinto. El segundo de ellos no muestra mejoría, al contrario, su entendimiento de la palabra disminuye un poco en el primer año, aunque no en gran medida y es factible que debido a la estabilidad que tenía haya desertado. Entonces su deserción se debió al azar. Para los otros dos sujetos, es difícil determinar si debido a que vieron mejoría y seguían sintiéndose bien decidieron desertar (debido al azar) o si se sintieron decepcionados o temerosos al ver que la mejoría desapareció de una visita a otra y prefirieron abandonar el ensayo (no debida al azar). Este tipo de cuestiones representan grandes problemas en la práctica ya que las decisiones que se toman en cuanto a la clasificación de las observaciones faltantes suele ser muy subjetiva. Pero, para fines de ilustración de la aplicación del método, se procederá dejando abierta la cuestión sólo poniendo en claro que se debe ser cuidadoso al momento de determinar si las estimaciones de los parámetros serán sesgadas o no.

Entonces, utilizando el paquete SAS se obtendrá la estimación de los parámetros por medio de los dos métodos permitidos por este enfoque<sup>9</sup>. Dichas estimaciones se encuentran más adelante.

La segunda propuesta para manejar los datos desequilibrados es la eliminación de los casos incompletos y trabajar sólo con los casos equilibrados o completos y de esta forma se ajusta el modelo propuesto. En este caso se trabaja con la siguiente tabla de datos:

Sujeto	Porcentaje de entendimiento de la palabra				
	3 meses	6 meses	12 meses	24 meses	36 meses
1	46	57	67	68	75
3	30	45	58	72	82
6	35	67	77	80	86
7	27	47	50	55	67
8	12	31	41	47	52
10	18	39	50	67	78
11	10	25	28	33	33
12	35	44	67	73	84
14	35	55	70	75	83

<sup>9</sup>El códigos en SAS correspondientes a este ejemplo se encuentran en el anexo 3



Después de hacer la eliminación de los casos incompletos en SAS y de aplicar el modelo correspondiente, se obtienen los nuevos valores para los estimadores de los coeficientes. Estos nuevos valores de los estimadores se encuentran también posteriormente.

La última opción disponible es imputar los valores faltantes y así trabajar con el conjunto de datos completo que resulte. Partiendo de eso, los valores faltantes se pueden imputar por la respuesta media para el  $i$ -ésimo sujeto, la respuesta media para la  $j'$ -ésima ocasión o por el valor de la última observación registrada para el sujeto correspondiente.

En este caso, se imputarán los valores faltantes a causa de deserción de los sujetos, por la última observación de dichos sujetos de manera respectiva. Y los datos intermitentes se imputarán por la respuesta media para dicho sujeto. De tal manera que las imputaciones serán las que se indican en la siguiente tabla:

Sujeto	Porcentaje de entendimiento de la palabra				
	3 meses	6 meses	12 meses	24 meses	36 meses
2	24	35	35	35	35
4	22	$(22+48+66+76)/4$ <b>=53</b>	48	66	76
5	18	15	14	14	14
9	55	76	76	76	76
13	22	45	$(22+45+68+78)/4$ <b>=53.25</b>	68	78

Y así se trabajará con un conjunto de datos balanceado resultante que es el siguiente:

Sujeto	Porcentaje de entendimiento de la palabra				
	3 meses	6 meses	12 meses	24 meses	36 meses
1	46	57	67	68	75
2	24	35	35	35	35
3	30	45	58	72	82
4	22	53	48	66	76
5	18	15	14	14	14
6	35	67	77	80	86
7	27	47	50	55	67
8	12	31	41	47	52
9	55	76	76	76	76
10	18	39	50	67	78
11	10	25	28	33	33
12	35	44	67	73	84
13	22	45	53.25	68	78
14	35	55	70	75	83

Finalmente haciendo las imputaciones correspondientes en SAS, se obtiene una tercera estimación de los parámetros de interés en el modelo.

A continuación se presenta el resumen de estimaciones de los coeficientes del modelo que se obtienen por ambos métodos (Máxima Verosimilitud y Máxima Verosimilitud Restringida) para cada uno de los tres procedimientos que se aplicaron para los valores faltantes:

<b>Método de Máxima Verosimilitud</b>					
Estimador	Efecto	Estimación para los datos			(p-value)
		Desbalanceados	Completos	Imputados	
$\hat{\beta}_0$	Intercepto	32,6164	34,7019	34,6573	<0,0001
$\hat{\beta}_1$	Tiempo	1,1868	1,1172	0,9517	<0,0001

<b>Método de Máxima Verosimilitud Restringida</b>					
Estimador	Efecto	Estimación para los datos			(p-value)
		Desbalanceados	Completos	Imputados	
$\hat{\beta}_0$	Intercepto	32,6197	34,7019	34,6573	<0,0001
$\hat{\beta}_1$	Tiempo	1,1867	1,1172	0,9517	<0,0001

Y también se presentan las estimaciones para los parámetros de la covarianza

<b>Método de Máxima Verosimilitud</b>					
Estimador	Efecto	Estimación para los datos			
		Desbalanceados	Completos	Imputados	
$\hat{\sigma}_u^2$	Intercepto	188,8000	157,5100	239,7800	
$\hat{\sigma}^2$	Residual	83,2608	83,1635	105,8900	

<b>Método de Máxima Verosimilitud Restringida</b>					
Estimador	Efecto	Estimación para los datos			
		Desbalanceados	Completos	Imputados	
$\hat{\sigma}_u^2$	Intercepto	205,0300	178,8000	259,4700	
$\hat{\sigma}^2$	Residual	85,0518	85,5396	107,8200	

Con estos resultados se puede observar que las variaciones en las estimaciones de los parámetros de covarianzas son notables, mientras que para los coeficientes del modelo se encuentra un comportamiento relativamente estable.

# Conclusiones

A lo largo de este trabajo pudo hacerse notar que la Estadística Clínica debe entenderse como la herramienta básica del investigador clínico, y además, que un buen científico estadístico comprende que esta herramienta no serviría de nada sin su correcta implementación, por lo tanto, debe reconocerse la competencia intelectual del investigador que logra una exitosa investigación y la correcta práctica de esta disciplina en las Ciencias Médicas y de la Salud.

Se pudo comprobar que el enfoque estadístico puede ayudar en la comprensión, organización y/o estructuración de los datos que son obtenidos de la realidad en diversos estudios, gracias a las bases que motivan la inferencia (como las llamó Fisher) y que son fundamentales para la forma de proceder de la Estadística, ya que, a partir del estudio de pequeñas muestras de poblaciones en las que claramente se aplican las leyes de los grandes números y de las fluctuaciones aleatorias se pueden hacer generalizaciones a toda la población.

En este tipo de experimentación es necesario poder interpretar correctamente y de una manera crítica los resultados obtenidos, comenzando por seleccionar correctamente el tipo de diseño del experimento y describir detalladamente los métodos de análisis de datos que serán empleados; complementado por el perfecto conocimiento de las variables de estudio.

La estadística indica el nivel de probabilidad que un estudio ofrece de que el efecto encontrado en un tratamiento se haya producido por casualidad, pero el objetivo de la investigación no es sólo verificar si los tratamientos provocan alguna reacción favorable, es decir, no se trata sólo de encontrar que tienen efectos estadísticamente significativos, sino además de determinar la magnitud de dicho efecto, para poder discernir de forma clara lo que es un efecto estadísticamente significativo y lo que es un efecto clínicamente relevante, ya que la asociación de ambos conceptos es lo que hace exitosa la investigación. Un efecto estadísticamente significativo pero sin relevancia clínica al igual que un tratamiento con relevancia clínica sin significación estadística no sirve para aprobar un nuevo tratamiento o terapia.

Si un tratamiento tiene efectos suficientemente significativos, significa que pudo

comprobarse científicamente su relevancia y que la magnitud del efecto es suficientemente grande como para considerar benéfico sustituir tratamientos existentes o aplicar los encontrados.

El análisis de datos longitudinales es una buena opción para implementar, ya que se hace un seguimiento, a través del tiempo, de cómo se han visto beneficiados o afectados los sujetos de un ensayo, lo que da pauta a que se tenga un mejor control del estado físico y mental de los pacientes, permitiendo así, una investigación más ética.

Es importante darse cuenta de que existen muchos tipos de análisis para los datos y con ello, muchos tipos de modelos que pueden ser aplicados al momento de realizar ensayos clínicos sobre nuevos tratamientos y terapias que en el futuro mejoren la calidad de vida de las personas, pero es imprescindible que el investigador cuente con los conocimientos suficientes para hacer una buena interpretación y selección de modelos.

# Bibliografía

Olga Korosteleva, Clinical Statistics: Introducing Clinical Trials, Survival Analysis and Longitudinal Data Analysis

Paul R. Rosenbaum, Observational Studies, Springer Series in Statistics

MC, Dr en C., Eduardo Lazcano-Ponce, Ensayos clínicos aleatorizados: variantes, métodos de aleatorización, análisis, consideraciones éticas y regulación

Méd. Tinoco Mora, Zahira y Méd. Sáenz Campos, Desirée; Investigación científica: Protocolos de investigación

Martín Andrés, Juan de Dios Luna del Castillo, Bioestadística para Ciencias de la Salud

Fernando Tato, Bases metodológicas del ensayo clínico

# Anexo 1

## Título Quinto de la Ley General de Salud: Materia de Investigación

**Artículo 96.** La investigación para la salud comprende el desarrollo de acciones que contribuyan:

1. Al conocimiento de los procesos biológicos y psicológicos en los seres humanos;
2. Al conocimiento de los vínculos entre las causas de enfermedad, la práctica médica y la estructura social;
3. A la prevención y control de los problemas de salud que se consideren prioritarios para la población;
4. Al conocimiento y control de los efectos nocivos del ambiente en la salud;
5. Al estudio de las técnicas y métodos que se recomienden o empleen para la prestación de servicios de salud, y
6. A la producción nacional de insumos para la salud.

**Artículo 97.** La Secretaría de Educación Pública, en coordinación con la Secretaría de Salud y con la participación que corresponda al Consejo Nacional de Ciencia y Tecnología orientará al desarrollo de la investigación científica y tecnológica destinada a la salud.

La Secretaría de Salud y los gobiernos de las entidades federativas, en el ámbito de sus respectivas competencias, apoyarán y estimularán el funcionamiento de establecimientos públicos destinados a la investigación para la salud.

**Artículo 98.** En las instituciones de salud, bajo la responsabilidad de los directores o titulares respectivos y de conformidad con las disposiciones aplicables, se constituirán: una comisión de investigación; una comisión de ética, en el caso de que se realicen investigaciones en seres humanos, y una comisión de bioseguridad, encargada de regular el uso de radiaciones ionizantes o de técnicas de ingeniería genética.

El Consejo de Salubridad General emitirá las disposiciones complementarias sobre áreas o modalidades de la investigación en las que considere que es necesario.

**Artículo 99.** La Secretaría de Salud, en coordinación con la Secretaría de Educación Pública, y con la colaboración del Consejo Nacional de Ciencia y Tecnología y de las instituciones de educación superior, realizará y mantendrá actualizando un inventario de la investigación en el área de salud del país.

**Artículo 100.** La investigación en seres humanos se desarrollará conforme a las siguientes bases:

1. Deberá adaptarse a los principios científicos y éticos que justifican la investigación médica, especialmente en lo que se refiere a su posible contribución a la solución de problemas de salud y al desarrollo de nuevos campos de la ciencia médica;
2. Podrá realizarse sólo cuando el conocimiento que se pretenda producir no pueda obtenerse por otro método idóneo;
3. Podrá efectuarse sólo cuando exista una razonable seguridad de que no expone a riesgos ni daños innecesarios al sujeto en experimentación;
4. Se deberá contar con el consentimiento por escrito del sujeto en quien se realizará la investigación, o de su representante legal en caso de incapacidad legal de aquél, una vez enterado de los objetivos de la experimentación y de las posibles consecuencias positivas o negativas para su salud;
5. Sólo podrá realizarse por profesionales de la salud en instituciones médicas que actúen bajo la vigilancia de las autoridades sanitarias competentes;
6. El profesional responsable suspenderá la investigación en cualquier momento, si sobreviene el riesgo de lesiones graves, invalidez o muerte del sujeto en quien se realice la investigación, y
7. Las demás que establezca la correspondiente reglamentación.

**Artículo 101.** Quien realice investigación en seres humanos en contravención a lo dispuesto en esta Ley y demás disposiciones aplicables, se hará acreedor de las sanciones correspondientes.



**Artículo 102.** La Secretaría de Salud podrá autorizar con fines preventivos, terapéuticos, rehabilitatorios o de investigación, el empleo en seres humanos de medicamentos o materiales respecto de los cuales aún no se tenga evidencia científica suficiente de su eficacia terapéutica o se pretenda la modificación de las indicaciones terapéuticas de productos ya conocidos. Al efecto, los interesados deberán presentar la documentación siguiente:

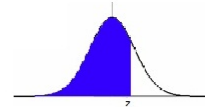
1. Solicitud por escrito;
2. Información básica farmacológica y preclínica del producto;
3. Estudios previos de investigación clínica, cuando los hubiere;
4. Protocolo de investigación, y
5. Carta de aceptación de la institución donde se efectúe la investigación y del responsable de la misma.

**Artículo 103.** En el tratamiento de una persona enferma, el médico, podrá utilizar nuevos recursos terapéuticos o de diagnóstico, cuando exista posibilidad fundada de salvar la vida, restablecer la salud o disminuir el sufrimiento del paciente, siempre que cuente con el consentimiento por escrito de éste, de su representante legal, en su caso, o del familiar más cercano en vínculo, y sin perjuicio de cumplir con los demás requisitos que determine esta Ley y otras disposiciones aplicables.

# Anexo 2

## Distribución Normal Estándar

**P [ Z ≤ z ]**



z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983

## Anexo 3

# Códigos SAS para los ejemplos del Capítulo 3

### Ejemplo de Modelo con intercepto aleatorio

```
data glioma;
```

```
input individuo grupo visita0 visita3 visita6 visita12 visita18 visita24;
```

```
cards;
```

```
1 1 3,1 3 2,7 2,3 2,1 1,8
2 1 3,3 2,9 2,4 1,8 1,7 0,2
3 1 2,9 2,4 2,3 2,1 2,1 1,6
4 1 3,2 2,7 2,7 2,2 2,1 1,3
5 1 3,5 3,2 3,2 3,2 3,1 0,8
6 1 3,6 3,5 1,7 1,6 1,5 1,1
7 1 2,2 2 2,7 2,4 2,4 1,6
8 1 3,8 3,7 3 2,8 2,6 1
9 1 3,4 3,1 3,3 4,3 4,1 3,6
10 1 4,6 4,4 2,6 2,5 2,4 1,5
11 1 2,7 2,6 3,1 3,1 3 1,6
12 1 4,9 4,7 2,9 2,9 2,8 2,6
13 2 4 3,5 3,1 2 1,1 1
14 2 3,8 3,3 2,7 1,6 1,2 0,6
15 2 3,7 3,2 2,9 2,4 1,5 1,4
16 2 3,2 3,1 2,9 1,8 0,6 0
17 2 2,5 2,1 1,5 0,3 0,1 0
18 2 2,9 2,2 3,1 1,9 1,7 1,5
19 2 2,8 2,6 2,8 2,6 2,6 2,3
20 2 3,1 2,7 2,3 2 1,9 1,3
```

```

21 2 4,3 4,2 2,4 2,3 1,2 0,4
22 2 2,9 2,4 2,1 2,1 1,7 1,7
23 2 4 3,4 2,3 1,4 0,9 0
24 2 2,5 2,4 2 1 0,3 0
;
data new;
set glioma;
array x{6} visita0 visita3 visita6 visita12 visita18 visita24;
array t{6} t1-t6 (0 3 6 12 18 24);
do visitas = 1 to 6;
tamanotumor = x{visitas};
tiempo = t{visitas};
output;
end;
keep individuo grupo tamanotumor tiempo;
run;

proc mixed data = new method =ml;
model tamanotumor = grupo tiempo / solution;
random intercept / subject = individuo;
run;

proc mixed data = new method =reml;
model tamanotumor = grupo tiempo / solution;
random intercept / subject = individuo;
run;

```

### **Ejemplo de modelo con intercepto y pendiente aleatorios**

```

data glioma;
input individuo grupo visita0 visita3 visita6 visita12 visita18 visita24;
cards;

```

```

1 1 3,1 3 2,7 2,3 2,1 1,8
2 1 3,3 2,9 2,4 1,8 1,7 0,2
3 1 2,9 2,4 2,3 2,1 2,1 1,6
4 1 3,2 2,7 2,7 2,2 2,1 1,3
5 1 3,5 3,2 3,2 3,2 3,1 0,8
6 1 3,6 3,5 1,7 1,6 1,5 1,1
7 1 2,2 2 2,7 2,4 2,4 1,6
8 1 3,8 3,7 3 2,8 2,6 1
9 1 3,4 3,1 3,3 4,3 4,1 3,6
10 1 4,6 4,4 2,6 2,5 2,4 1,5

```

```

11 1 2,7 2,6 3,1 3,1 3 1,6
12 1 4,9 4,7 2,9 2,9 2,8 2,6
13 2 4 3,5 3,1 2 1,1 1
14 2 3,8 3,3 2,7 1,6 1,2 0,6
15 2 3,7 3,2 2,9 2,4 1,5 1,4
16 2 3,2 3,1 2,9 1,8 0,6 0
17 2 2,5 2,1 1,5 0,3 0,1 0
18 2 2,9 2,2 3,1 1,9 1,7 1,5
19 2 2,8 2,6 2,8 2,6 2,6 2,3
20 2 3,1 2,7 2,3 2 1,9 1,3
21 2 4,3 4,2 2,4 2,3 1,2 0,4
22 2 2,9 2,4 2,1 2,1 1,7 1,7
23 2 4 3,4 2,3 1,4 0,9 0
24 2 2,5 2,4 2 1 0,3 0
;
data porcentajes;
input individuo grupo mes3 mes6 mes12 mes18 mes24;
cards;

```

```

1 1 90 85 70 67 90
3 1 78 67 65 63 60
4 1 77 65 61 57 56
5 1 78 77 77 67 67
6 1 82 80 80 73 55
7 1 88 83 68 58 58
8 1 95 95 90 87 85
9 1 84 74 64 60 27
10 1 78 71 63 60 60
11 1 91 90 84 84 78
12 1 83 81 80 81 80
13 2 78 67 63 70 70
14 2 85 78 63 61 40
15 2 85 70 70 70 70
16 2 85 78 63 53 13
17 2 85 79 44 41 30
18 2 96 66 58 40 32
19 2 89 63 60 52 45
20 2 95 65 55 40 33
21 2 73 68 41 52 26
22 2 85 75 41 37 33

23 2 74 70 64 63 42
24 2 96 67 78 74 55

```

```

;
data combined;
merge glioma porcentajes;
by individuo;
run;

data new;
set combined;
array x{5} visita3 visita6 visita12 visita18 visita24;
array z{5} mes3 mes6 mes12 mes18 mes24;
array t{5} (3 6 12 18 24);
do visitas = 1 to 5;
tamanotumor = x{visitas};
tiempo = t{visitas};
porcentaje = z{visitas};
output;
end;
keep individuo grupo tamanotumor tiempo porcentaje;
run;

proc mixed data = new method = reml;
model porcentaje = grupo tamanotumor tiempo / solution;
random intercept tiempo / subject = individuo type = un;
run;

proc mixed data = new method = ml;
model porcentaje = grupo tamanotumor tiempo / solution;
random intercept tiempo / subject = individuo type = un;
run;

```

### **Ejemplo de Modelo con Estructura Covarianza Espacial de Potencia para el Error**

```

data glioma;
input individuo grupo visita0 visita3 visita6 visita12 visita18 visita24;
cards;

```

```

1 1 3,1 3 2,7 2,3 2,1 1,8
2 1 3,3 2,9 2,4 1,8 1,7 0,2
3 1 2,9 2,4 2,3 2,1 2,1 1,6

```

```

4  1  3,2  2,7  2,7  2,2  2,1  1,3
5  1  3,5  3,2  3,2  3,2  3,1  0,8
6  1  3,6  3,5  1,7  1,6  1,5  1,1
7  1  2,2  2    2,7  2,4  2,4  1,6
8  1  3,8  3,7  3    2,8  2,6  1
9  1  3,4  3,1  3,3  4,3  4,1  3,6
10 1  4,6  4,4  2,6  2,5  2,4  1,5
11 1  2,7  2,6  3,1  3,1  3    1,6
12 1  4,9  4,7  2,9  2,9  2,8  2,6
13 2  4    3,5  3,1  2    1,1  1
14 2  3,8  3,3  2,7  1,6  1,2  0,6
15 2  3,7  3,2  2,9  2,4  1,5  1,4
16 2  3,2  3,1  2,9  1,8  0,6  0
17 2  2,5  2,1  1,5  0,3  0,1  0
18 2  2,9  2,2  3,1  1,9  1,7  1,5
19 2  2,8  2,6  2,8  2,6  2,6  2,3
20 2  3,1  2,7  2,3  2    1,9  1,3
21 2  4,3  4,2  2,4  2,3  1,2  0,4
22 2  2,9  2,4  2,1  2,1  1,7  1,7
23 2  4    3,4  2,3  1,4  0,9  0
24 2  2,5  2,4  2    1    0,3  0
;
data porcentajes;
input individuo grupo mes3 mes6 mes12 mes18 mes24;
cards;

```

```

1  1  90  85  70  67  90
3  1  78  67  65  63  60
4  1  77  65  61  57  56
5  1  78  77  77  67  67
6  1  82  80  80  73  55
7  1  88  83  68  58  58
8  1  95  95  90  87  85
9  1  84  74  64  60  27
10 1  78  71  63  60  60
11 1  91  90  84  84  78
12 1  83  81  80  81  80
13 2  78  67  63  70  70
14 2  85  78  63  61  40
15 2  85  70  70  70  70

```

```

16 2 85 78 63 53 13
17 2 85 79 44 41 30
18 2 96 66 58 40 32
19 2 89 63 60 52 45
20 2 95 65 55 40 33
21 2 73 68 41 52 26
22 2 85 75 41 37 33
23 2 74 70 64 63 42
24 2 96 67 78 74 55
;
data combined;
merge glioma porcentajes;
by individuo;
run;

data new;
set combined;
array x{5} visita3 visita6 visita12 visita18 visita24;
array z{5} mes3 mes6 mes12 mes18 mes24;
array t{5} (3 6 12 18 24);
do visitas = 1 to 5;
tamanotumor = x{visitas};
tiempo = t{visitas};
porcentaje = z{visitas};
output;
end;
keep individuo grupo tamanotumor tiempo porcentaje;
run;

proc mixed data = new method = reml;
model porcentaje = grupo tamanotumor tiempo / solution;
random intercept tiempo / subject = individuo type = un;
repeated / subject = individuo type = sp(pow)(tiempo);
run;

proc mixed data = new method = ml;
model porcentaje = grupo tamanotumor tiempo / solution;
random intercept tiempo / subject = individuo type = un;
repeated / subject = individuo type = sp(pow)(tiempo);
run;

```

### **Ejemplo de Modelo de regresión logística con intercepto aleatorio**



```

data osteoporosis;
input individual age calcium history mos3 mos9 mos12 mos18;
datalines;

  1  76  0  1  1  1  1  1
  2  57  1  1  1  1  0  0
  3  58  0  1  1  1  1  0
  4  62  1  0  1  0  0  0
  5  60  1  0  0  0  0  0
  6  58  0  1  1  0  1  1
  7  52  1  0  0  1  0  0
  8  74  0  1  1  0  1  0
  9  51  0  0  0  1  0  0
 10  56  1  0  0  1  0  0
 11  75  0  1  1  1  1  1
 12  63  1  0  1  0  0  0
 13  67  1  1  0  1  0  0
 14  68  0  0  1  1  0  0
 15  56  1  0  1  0  1  0
 16  62  1  0  1  0  1  1
 17  60  1  1  0  0  0  0
 18  61  1  1  1  1  0  0
 19  54  1  0  1  0  0  0
 20  53  0  0  1  1  0  0

data new;
set osteoporosis;
array x{4} mos3 mos9 mos12 mos18;
array t{4} (3 9 12 18);
do visists = 1 to 4;
disease = x{visits};
time = t{visits};
output;
end;
keep individual age calcium history disease time;
run;

proc glimmix data = new;
model disease(event = '1') = age calcium history disease time / solution
dist = binary link = logit;
random intercept / subject = individual type = un;
run;

```

## Ejemplo con Observaciones Faltantes

```
/*Parte 1*/
data recorded;
input individual mos3 mos6 mos12 mos24 mos36;
datalines;

    1  46  57  67  68  75
    2  24  35  .   .   .
    3  30  45  58  72  82
    4  22  .   48  66  76
    5  18  15  14  .   .
    6  35  67  77  80  86
    7  27  47  50  55  67
    8  12  31  41  47  52
    9  55  76  .   .   .
   10  18  39  50  67  78
   11  10  25  28  33  33
   12  35  44  67  73  84
   13  22  45  .   68  78
   14  35  55  70  75  83
;
data unbalanced;
set recorded;
array x{5} mos3 mos6 mos12 mos24 mos36;
array t{5} t1-t5 (3 6 12 24 36);
do visits = 1 to 5;
score = x{visits};
time = t{visits};
output;
end;
keep individual score time;
run;

proc mixed data = unbalanced method = ml;
model score = time / solution;
random intercept / subject = individual;
run;

proc mixed data = unbalanced method = reml;
model score = time / solution;
random intercept / subject = individual;
run;
```

```

/*Parte 2*/ data complete;
set unbalanced;
if individual = 2 then delete;
if individual = 4 then delete;
if individual = 5 then delete;
if individual = 9 then delete;
if individual = 13 then delete;
run;

proc mixed data = complete method = ml;
model score = time / solution;
random intercept / subject = individual;
run;

proc mixed data = complete method = reml;
model score = time / solution;
random intercept / subject = individual;
run;

/*Parte 3*/ data imputed;
set unbalanced;
if individual = 2 then do;
if score = . then score = 35;
end;
if individual = 4 then do;
if score = . then score = 53;
end;
if individual = 5 then do;
if score = . then score = 14;
end;
if individual = 9 then do;
if score = . then score = 76;
end;
if individual = 13 then do;
if score = . then score = 53.25;
end;
run;

proc mixed data = imputed method = ml;
model score = time / solution;
random intercept / subject = individual;

```

```
run;  
  
proc mixed data = imputed method = reml;  
model score = time / solution;  
random intercept / subject = individual;  
run;
```