

# Contenido

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Principios Básicos del Diseño de Experimentos</b>                   | <b>1</b>  |
| 1.1      | Introducción . . . . .   | 1         |
| 1.2      | Tres Tipos de Variabilidad . . . . .                                   | 2         |
| 1.3      | Planificando un Experimento . . . . .                                  | 3         |
| 1.4      | Tres Principios Básicos . . . . .                                      | 9         |
| 1.5      | Un Ejemplo Real . . . . .  | 10        |
| 1.6      | Algunos Diseños Experimentales Estándar . . . . .                      | 13        |
| 1.6.1    | Diseños Completamente Aleatorizados . . . . .                          | 13        |
| 1.6.2    | Diseños en Bloques o con Un Factor Bloque . . . . .                    | 13        |
| 1.6.3    | Diseños con Dos o Más Factores Bloque . . . . .                        | 14        |
| 1.7      | Más Ejemplos Reales . . . . .  | 16        |
| <b>2</b> | <b>Diseños con Una Fuente de Variación</b>                             | <b>23</b> |
| 2.1      | Introducción . . . . .   | 23        |
| 2.2      | Aleatorización . . . . .   | 23        |
| 2.3      | Modelo Matemático para un Diseño Completamente Aleatorizado . . . . .  | 24        |
| 2.4      | Estimación de los Parámetros . . . . .                                 | 25        |
| 2.4.1    | Funciones Estimables de Parámetros . . . . .                           | 25        |
| 2.4.2    | Estimadores Puntuales Mínimo-Cuadráticos . . . . .                     | 26        |
| 2.4.3    | Estimación Puntual de $\sigma^2$ . . . . .                             | 27        |
| 2.5      | Análisis de la Varianza de Una Vía . . . . .                           | 29        |
| 2.5.1    | Idea General . . . . .   | 29        |
| 2.5.2    | Descomposición de la Variabilidad . . . . .                            | 30        |
| 2.6      | Determinación del Tamaño Muestral en base al Contraste Anova . . . . . | 31        |
| 2.7      | Inferencia de Contrastes y Medias de Tratamientos . . . . .            | 33        |
| 2.7.1    | Contrastes . . . . .   | 33        |
| 2.7.2    | Contrastes Individuales y Medias de Tratamientos . . . . .             | 36        |
| 2.8      | Métodos de Comparaciones Múltiples . . . . .                           | 39        |
| 2.8.1    | Método de Bonferroni . . . . .   | 41        |

|          |  |           |
|----------|--|-----------|
| 2.8.2    | Método de Scheffé . . . . .  | 42        |
| 2.8.3    | Método de Tukey para Todas las Comparaciones a Pares . . . . .                               | 43        |
| 2.8.4    | Método de Dunnett para Comparaciones entre Tratamientos y Control . . . . .                  | 44        |
| 2.8.5    | Otros Métodos . . . . .  | 45        |
| 2.9      | Tamaños Muestrales . . . . .   | 45        |
| 2.10     | Comprobación de la Idoneidad del Modelo . . . . .  | 46        |
| 2.10.1   | Bondad del Ajuste del Modelo . . . . .   | 47        |
| 2.10.2   | Existencia de Observaciones Atípicas . . . . .   | 48        |
| 2.10.3   | Independencia de los Errores . . . . .   | 50        |
| 2.10.4   | Homocedasticidad de los Errores . . . . .  | 55        |
| 2.10.5   | Normalidad . . . . .   | 56        |
| 2.11     | Alternativas al Análisis de la Varianza . . . . .  | 58        |
| 2.11.1   | Introducción . . . . .   | 58        |
| 2.11.2   | Alternativa No Paramétrica: Prueba de Kruskal-Wallis . . . . .                               | 59        |
| 2.11.3   | Alternativa para Variables Respuesta Nominales: Contraste $\chi^2$ de Homogeneidad . . . . . | 62        |
| 2.12     | Problemas Propuestos . . . . .   | 67        |
| <b>3</b> | <b>Diseños con Dos Factores Tratamiento Cruzados</b>   | <b>71</b> |
| 3.1      | Introducción . . . . .   | 71        |
| 3.2      | El Significado de la Interacción . . . . .   | 71        |
| 3.3      | Modelos Matemáticos Para Dos Factores Tratamiento . . . . .                                  | 73        |
| 3.4      | Contrastes . . . . .   | 76        |
| 3.4.1    | Contrastes de Efectos Principales e Interacciones . . . . .                                  | 76        |
| 3.4.2    | Contrastes como Listas de Coeficientes . . . . .   | 78        |
| 3.5      | Análisis del Modelo Completo de Dos Vías . . . . .   | 79        |
| 3.5.1    | Estimadores Mínimo-Cuadráticos . . . . .   | 79        |
| 3.5.2    | Estimación de la Varianza . . . . .  | 80        |
| 3.5.3    | Análisis de la Varianza para el Modelo Completo . . . . .                                    | 81        |
| 3.6      | Análisis del Modelo de Efectos Principales de Dos Vías . . . . .                             | 85        |
| 3.6.1    | Estimadores Mínimo-Cuadráticos . . . . .   | 85        |
| 3.6.2    | Estimación de la Varianza . . . . .  | 87        |
| 3.6.3    | Análisis de la Varianza para el Modelo de Dos Vías de Efectos Principales . . . . .          | 87        |
| 3.7      | Comparaciones Múltiples . . . . .  | 89        |
| 3.8      | Tamaños Muestrales . . . . .   | 90        |
| 3.9      | Experimentos Pequeños . . . . .  | 91        |
| 3.9.1    | Análisis Basado en Contrastes Ortogonales . . . . .  | 91        |
| 3.10     | Comprobación de la Idoneidad del Modelo . . . . .  | 92        |
| 3.11     | Problemas Propuestos . . . . .   | 93        |

|          |  |            |
|----------|--|------------|
| <b>4</b> | <b>Análisis de la Covarianza</b>   | <b>97</b>  |
| 4.1      | Introducción . . . . .   | 97         |
| 4.2      | Modelos Matemáticos . . . . .  | 98         |
| 4.3      | Estimadores Mínimo-Cuadráticos . . . . .   | 99         |
| 4.4      | Análisis de la Covarianza . . . . .  | 101        |
| 4.5      | Comprobación de la Idoneidad del Modelo . . . . .  | 103        |
| 4.6      | Contrastes Individuales y Comparaciones Múltiples . . . . .                                  | 104        |
| 4.7      | Problemas Propuestos . . . . .   | 106        |
| <br>     |  |            |
| <b>5</b> | <b>Diseños en Bloques</b>  | <b>109</b> |
| 5.1      | Generalidades . . . . .  | 109        |
| 5.1.1    | Introducción . . . . .   | 109        |
| 5.1.2    | Acerca del Tamaño de los Bloques . . . . .   | 110        |
| 5.1.3    | Definiciones . . . . .   | 111        |
| 5.2      | Diseños en Bloques Completos . . . . .   | 112        |
| 5.2.1    | Diseño en Bloques Completamente Aleatorizados . . . . .                                      | 112        |
| 5.2.2    | Diseño General en Bloques Completos . . . . .  | 113        |
| 5.2.3    | Análisis del Diseño en Bloques Completamente Aleatorizados . . . . .                         | 114        |
| 5.2.4    | Análisis del Diseño General en Bloques Completos . . . . .                                   | 116        |
| 5.2.5    | Tamaños Muestrales . . . . .   | 119        |
| 5.3      | Diseños en Bloques Incompletos . . . . .   | 119        |
| 5.3.1    | Definiciones y Consideraciones Generales . . . . .   | 119        |
| 5.3.2    | Diseños en Bloques Incompletos Balanceados . . . . .   | 124        |
| 5.3.3    | Diseños Divisibles en Grupos . . . . .   | 125        |
| 5.3.4    | Diseños Cíclicos . . . . .   | 126        |
| 5.3.5    | Análisis de un Diseño General en Bloques Incompletos . . . . .                               | 127        |
| 5.3.6    | Análisis de un Diseño en Bloques Incompletos Balanceados . . . . .                           | 132        |
| 5.3.7    | Análisis de un Diseño Divisible en Grupos . . . . .  | 133        |
| 5.3.8    | Análisis de un Diseño Cíclico . . . . .  | 134        |
| 5.3.9    | Tamaños Muestrales . . . . .   | 134        |
| 5.4      | Diseños con Dos Factores Bloque sin Interacción . . . . .                                    | 136        |
| 5.4.1    | Definiciones y Consideraciones Generales: Diseños Fila-Columna . . . . .                     | 136        |
| 5.4.2    | Diseños en Cuadrado Latino . . . . .   | 137        |
| 5.4.3    | Diseños Youden . . . . .   | 139        |
| 5.4.4    | Diseños Cíclicos y otros Diseños Fila-Columna . . . . .                                      | 140        |
| 5.4.5    | Modelo Matemático para un Diseño Fila-Columna . . . . .                                      | 141        |
| 5.4.6    | Estimadores Mínimo-Cuadráticos en un Diseño Fila-Columna . . . . .                           | 142        |
| 5.4.7    | Análisis de la Varianza en un Diseño Fila-Columna con Bloques Columna<br>Completos . . . . . | 144        |

|          |   |            |
|----------|---|------------|
| 5.4.8    | Comparaciones Múltiples y Tamaños Muestrales . . . . .                      | 147        |
| 5.5      | Alternativas No Paramétricas . . . . .                                      | 148        |
| 5.5.1    | Diseño en Bloques Completamente Aleatorizados: Prueba de Friedman . . . . . | 148        |
| 5.5.2    | Diseño en Bloques Incompletos Balanceados: Prueba de Durbin . . . . .       | 151        |
| 5.6      | Problemas Propuestos . . . . .  | 153        |
| <b>6</b> | <b>Efectos Aleatorios y Componentes de la Varianza</b>                      | <b>161</b> |
| 6.1      | Introducción . . . . .  | 161        |
| 6.2      | Algunos Ejemplos . . . . .  | 162        |
| 6.3      | Un Efecto Aleatorio . . . . .   | 164        |
| 6.3.1    | El Modelo de Una Vía de Efectos Aleatorios . . . . .                        | 164        |
| 6.3.2    | Estimación de $\sigma^2$ . . . . .  | 165        |
| 6.3.3    | Estimación de $\sigma_T^2$ . . . . .  | 166        |
| 6.3.4    | Contraste de Igualdad de los Efectos Tratamiento . . . . .                  | 167        |
| 6.3.5    | Intervalos de Confianza para las Componentes de la Varianza . . . . .       | 169        |
| 6.4      | Tamaños Muestrales para un Experimento con Un Efecto Aleatorio . . . . .    | 171        |
| 6.5      | Dos o Más Efectos Aleatorios . . . . .                                      | 173        |
| 6.5.1    | Algunos Modelos . . . . .   | 173        |
| 6.5.2    | Estimación de $\sigma^2$ . . . . .  | 174        |
| 6.5.3    | Estimación de las Componentes de la Varianza . . . . .                      | 174        |
| 6.5.4    | Intervalos de Confianza para las Componentes de la Varianza . . . . .       | 177        |
| 6.5.5    | Pruebas de Hipótesis para las Componentes de la Varianza . . . . .          | 178        |
| 6.6      | Problemas Propuestos . . . . .  | 180        |
|          | <b>Apéndice: Tablas</b>   | <b>183</b> |
|          | <b>Referencias Bibliográficas</b>   | <b>197</b> |

# Capítulo 1

## Principios Básicos del Diseño de Experimentos

### 1.1 Introducción

La adquisición de conocimientos nuevos viene condicionada por dos elementos esenciales: la ocurrencia de algún hecho diferente de lo habitual, cosa por tanto poco frecuente, y la circunstancia de que este hecho se produzca en presencia de una persona capaz de identificarlo como extraño, reflexionar sobre él y, lo que es más difícil, extraer consecuencias.

Sin duda, algunos ejemplos muy conocidos y relevantes son:

- la manzana, Newton y la ley de gravitación universal;
- la tapadera de la olla, Watt y la máquina de vapor;
- el incensario, Galileo y la ley del péndulo;
- la aparición de hongos, Fleming y la penicilina; . . .

La experimentación no consiste más que en reproducir artificialmente estas dos circunstancias. Se trata de forzar la aparición de circunstancias “extrañas” en presencia de personas especialmente preparadas para interpretar y extraer conclusiones de lo que ocurra.

En una investigación de tipo empírico lo natural es que, repitiendo un experimento en condiciones indistinguibles para el experimentador, los resultados presenten **variabilidad**.

*La metodología del Diseño de Experimentos estudia cómo variar deliberadamente las condiciones habituales de un proceso empírico para aumentar la probabilidad de detectar cambios significativos en la respuesta y obtener así un conocimiento más profundo sobre el comportamiento del proceso.*

Típicamente, un experimento se realiza por una o varias de las siguientes razones:

## 2 Principios Básicos del Diseño de Experimentos

- determinar las principales causas de variación en la respuesta,
- encontrar las condiciones experimentales que permiten alcanzar un valor extremo en la respuesta,
- comparar las respuestas en diferentes niveles de observación de variables controladas,
- obtener un modelo matemático que permita predecir respuestas futuras.

Pero, ¿es realmente necesario planificar experimentos? ¿no se podrían obtener las mismas conclusiones analizando convenientemente los **datos disponibles**?

Sin con ello pretender afirmar que no se pueda aprender de los estudios *observacionales*, es claro que analizar datos que no proceden de una adecuada planificación experimental puede resultar peligroso. Las siguientes cuatro posibles situaciones, derivadas de la simple observación muestral, implican un alto riesgo de cometer errores en el análisis estadístico:

- inconsistencia de los datos;
- rango de variables limitado;
- variables altamente correlacionadas: confusión de efectos;
- variables altamente correlacionadas: relación no causal (variable oculta).

### 1.2 Tres Tipos de Variabilidad

Los resultados de cualquier experimento estarán, probablemente, sometidos a tres tipos de variabilidad:

#### 1. Variabilidad sistemática y planificada. (“DESEABLE”)

Es el tipo de variabilidad que tratamos de identificar y provocamos con nuestro diseño. Incluye la posible dispersión de los resultados achacable a diferencias sistemáticas entre las distintas condiciones experimentales impuestas en el diseño por expreso deseo del experimentador.

Cuando este tipo de variabilidad está presente y tiene un tamaño importante, esperamos que las respuestas tiendan a agruparse formando clusters.

#### 2. Variabilidad típica de la naturaleza del problema y del experimento. (“TOLERABLE”)

Nos referimos al llamado *ruido aleatorio* (e incluye, entre otras, a la componente de variabilidad no planificada denominada *error de medida*).

Es impredecible e inevitable. Su comportamiento es similar a la variabilidad que pueden presentar en un laboratorio las medidas repetidas de un mismo objeto. Probablemente la segunda medida no sea igual a la primera y, más aún, no se pueda predecir el valor de la tercera. Sin embargo, bajo el aparente caos, existe un patrón regular de comportamiento en esas medidas: todas ellas tenderán a fluctuar en torno a algún valor central.

Como ya se ha dicho, esta variabilidad es inevitable pero, si el experimento ha sido bien planificado, será posible estimar su tamaño, lo cual resultará de extrema utilidad para establecer conclusiones apropiadas.

### 3. Variabilidad sistemática y no planificada. (“AMENAZA CON EL DESASTRE”)

Supone la principal causa de conclusiones erróneas y estudios ruinosos. Tiene lugar cuando, debido a causas desconocidas y no planificadas en el experimento, se produce una variación sistemática en los resultados. En otras palabras, los resultados están siendo *sesgados* sistemáticamente por causas desconocidas.

Como veremos a continuación existen dos estrategias principales para manipular este tipo de variabilidad: *aleatorización* y *bloqueo*.

## 1.3 Planificando un Experimento

El siguiente listado propone una serie de acciones que deben ser ejecutadas secuencialmente en un buen proceso de planificación de un diseño experimental.

1. Definir los objetivos del experimento.
2. Identificar todas las posibles fuentes de variación, incluyendo:
  - *factores tratamiento y sus niveles,*
  - *unidades experimentales,*
  - *factores “nuisance”: bloques, factores ruido y covariables.*
3. Elegir una regla de asignación de las unidades experimentales a las condiciones de estudio (*tratamientos*).
4. Especificar las medidas que se realizarán (la *respuesta*), el procedimiento experimental y anticiparse a las posibles dificultades.
5. Ejecutar un experimento piloto.
6. Especificar el modelo.
7. Esquematisar los pasos del análisis.
8. Determinar el tamaño muestral.
9. Revisar las decisiones anteriores. Modificar si es necesario.

## 4 Principios Básicos del Diseño de Experimentos

Los pasos del anterior listado no son independientes de modo que en cualquier paso puede ser necesario volver atrás y modificar decisiones tomadas en algún paso previo.

A continuación se da una breve descripción de las decisiones que hay que tomar en cada uno de los pasos del listado anterior. Sólo después de haber tomado estas decisiones se procederá con la ejecución del experimento.

### 1.- Definir los objetivos del experimento.

Confeccionar una lista con las cuestiones concretas a las que debe dar respuesta nuestro experimento. Sólo las cuestiones esenciales ya que problemáticas colaterales suelen complicar innecesariamente el experimento, incrementando el coste y la verosimilitud de los errores.

Suele ser útil, una vez finalizada la lista de objetivos, esquematizar las conclusiones a las que esperamos llegue el posterior análisis de los datos.

Es muy probable que la lista tenga que ser refinada a medida que desarrollamos los siguientes pasos en la planificación de la experiencia.

### 2.- Identificar todas las posibles fuentes de variación.

Una *fente de variación* es cualquier “cosa” que pueda generar variabilidad en la respuesta. Una práctica muy recomendable es confeccionar una lista con todas las fuentes de variación concebibles en nuestro problema, distinguiendo aquellas que, a priori, generarán una mayor variabilidad. De entre éstas últimas se distinguen dos tipos: aquellas cuyo efecto sobre la respuesta es de particular interés para el experimentador (*factores tratamiento*) y aquellas que no son de interés directo pero que se contemplarán en el diseño en aras de reducir la variabilidad no planificada (*factores “nuisance”*).

#### (i) Factores tratamiento y sus niveles.

Se denomina *factor tratamiento* a cualquier variable cuyo posible efecto sobre los datos desea ser investigado. Los *niveles* de un factor tratamiento serán los tipos o grados específicos del factor que se utilizarán en el experimento y muchas veces nos referiremos a ellos simplemente como *tratamientos*.

Por ejemplo, un factor tratamiento podría ser una droga, un aditivo químico, la temperatura, el sexo, un método de enseñanza, ... Los niveles respectivos podrían ser: distintas cantidades de la droga, diferentes tipos de aditivos químicos, conjuntos de temperaturas seleccionadas en unos rangos de interés, hombre y mujer, un número determinado de métodos de enseñanza cuyos resultados se quieren comparar, ...

Muy pocos experimentos englobarán más de cuatro niveles por factor.

Si los niveles son cuantitativos suelen ser elegidos equiespaciados y, en otro caso, es habitual codificarlos.

Cuando un experimento contempla la existencia de más de un factor tratamiento, cada observación es una medida en las condiciones determinadas por la combinación de niveles de

los distintos factores tratamiento. En el supuesto de que existan observaciones en todas las posibles combinaciones de niveles se habla de *experimentos factoriales* y a ellos nos referiremos más adelante.

**(ii) Unidades experimentales.**

Son el *material* donde evaluar la variable respuesta y al cual se aplicarán los distintos niveles de los factores tratamiento. Ejemplos podrían ser: en agricultura, parcelas de tierra; en medicina, individuos humanos u animales; en industria, lotes de material, trabajadores, máquinas, . . . Cuando un experimento se ejecuta sobre un período de tiempo, de modo que las observaciones se recogen secuencialmente en instantes de tiempo determinados, entonces los propios instantes de tiempo pueden considerarse unidades experimentales.

Las conclusiones de un experimento que usa a estudiantes universitarios como unidades experimentales no son extrapolables a toda la población adulta de un país. Es decir, será muy importante que las unidades experimentales sean representativas de la población sobre la que se han fijado los objetivos del paso 1.

**(iii) Factores “nuisance”: bloques, factores ruido y covariables.**

En cualquier experimento, además de los factores tratamiento, cuyo efecto sobre la respuesta desea ser evaluado, influirán otra gran cantidad de factores, de escaso interés en el estudio, pero cuya influencia sobre la respuesta puede aumentar significativamente la variabilidad no planificada. En aras de eliminar esa influencia podrían incluirse en el diseño nuevos factores que, atendiendo a su naturaleza, pueden ser de diversos tipos.

En ocasiones el particular factor nuisance puede ser fijado en distintos niveles, de modo que es posible controlar su efecto a esos niveles: se mantiene constante su nivel para un grupo de unidades experimentales, se cambia a otro nivel para otro grupo y así sucesivamente. Estos factores se denominan *factores de bloqueo* y las unidades experimentales evaluadas en un mismo nivel del bloqueo se dice que pertenecen al mismo *bloque*. Incluso cuando el factor nuisance no es medible es a veces posible agrupar las unidades experimentales en bloques de unidades similares: parcelas de tierra contiguas o períodos de tiempo próximos probablemente conduzcan a unidades experimentales más parecidas que parcelas o períodos distantes. A los factores bloque nos referiremos con detalle en la sección 1.4 y en el capítulo 5.

Otras veces el factor nuisance es una propiedad cuantitativa de las unidades experimentales que puede ser medida antes de realizar el experimento (la presión sanguínea de un paciente en un experimento médico o la acidez de una parcela de tierra en un experimento agrícola). El factor se denomina entonces *covariable* y juega además un papel muy importante en el análisis estadístico.

Por último, si el experimentador está interesado en la variabilidad de la respuesta cuando se modifican las condiciones experimentales, entonces los factores nuisance son incluidos deliberadamente en el experimento y no se aísla su efecto vía el bloqueo. Se habla entonces de factores *ruido*.

Terminado el paso 2 conoceremos todas las posibles fuentes de variación del experimento:

| Fuente   | Tipo                               |
|--|------------------------------------|
| Debida a las condiciones de interés<br>(Factores tratamiento)                        | Planificada y sistemática          |
| Debida al resto de condiciones controladas<br>(Factores “nuisance”)                  | Planificada y sistemática          |
| Debida a condiciones no controladas<br>(error de medida, material experimental, ...) | No planificada, pero ¿sistemática? |

**3.- Elegir una regla de asignación de las unidades experimentales a las condiciones de estudio (“tratamientos”).**

La regla de asignación, o *diseño experimental*, especifica que unidades experimentales se observarán bajo cada tratamiento. Diseño factorial o no, anidamiento, asignación al azar en determinados niveles de observación, el orden de asignación, etc. Existen una serie de diseños estándar que a menudo se usan en la práctica.

**4.- Especificar las medidas que se realizarán (la “respuesta”), el procedimiento experimental y anticiparse a las posibles dificultades.**

Los datos que se recogerán en un experimento son medidas de una *variable respuesta*. Es por tanto muy importante precisar de antemano cuál será esta variable y en qué unidades se medirá. Naturalmente, la respuesta a ambas cuestiones estará condicionada por los objetivos del experimento. Por ejemplo, si se desea detectar una diferencia de 0’05 gramos en la respuesta de dos tratamientos no será apropiado toma medidas con una precisión próxima al gramo.

A menudo aparecen dificultades imprevistas en la toma de datos. Es buena idea anticiparse a estos imprevistos pensando detenidamente en los problemas que se pueden presentar o ejecutando un pequeño experimento piloto. Enumerar estos problemas permite en ocasiones descubrir nuevas fuentes de variación o simplificar el procedimiento experimental antes de comenzar.

También es provechoso determinar con claridad y anticipación cómo se harán las mediciones: instrumentos de medida, disponibilidad de éstos, momento en que se tomarán, etc.

**5.- Ejecutar un experimento piloto.**

Un experimento piloto es un mini-experimento involucrando sólo un número pequeño de observaciones. El objetivo de su ejecución no es necesariamente obtener conclusiones sino ayudar a completar y chequear la lista de acciones. Por ejemplo:

- ★ proporciona una oportunidad para practicar la técnica experimental e identificar problemas insospechados en el proceso de recogida de los datos,
- ★ si es suficientemente grande puede ayudar a seleccionar un modelo adecuado al experimento principal,

★ los errores experimentales observados en el experimento piloto pueden ayudar a calcular el número de observaciones que se precisan en el experimento principal, etc.

### 6.- Especificar el modelo.

El modelo matemático especificado debe indicar explícitamente la relación que creemos existe entre la variable respuesta y las principales fuentes de variación que hemos identificado en el paso 2. Dado que las técnicas analíticas a emplear más tarde dependerán de la forma del modelo, es importante que éste represente la realidad con la mayor precisión posible.

El tipo de modelo más habitual es el lineal:  $Y = \sum_{i=1}^k \alpha_i + \varepsilon$ , donde la respuesta se explica por una combinación lineal de términos representando las principales fuentes de variación planificada más un término que aglutina todas las fuentes de variación no planificada. El experimento piloto podría ayudar a ver si el modelo se ajusta razonablemente bien a la realidad.

Un modelo se dice *modelo de efectos fijos* si los niveles de todos los factores han sido seleccionados por el experimentador. Naturalmente es apropiado cuando el interés se centra en comparar el efecto (que se denomina *efecto fijo*) sobre la respuesta de esos niveles específicos.

Ocasionalmente sin embargo un factor tiene un número excesivamente grande de niveles, de modo que se opta por incluir en el experimento tan sólo una muestra aleatoria simple de ellos. El efecto de ese factor se dice entonces *efecto aleatorio*. En este caso no interesa comparar los niveles controlados en el experimento sino examinar la variabilidad de la respuesta debida a la población entera de niveles del factor. Un modelo con la totalidad de factores de efectos aleatorios se dice de *modelo de efectos aleatorios*. Llamamos *modelos mixtos* a aquellos que combinan factores de efectos fijos con factores de efectos aleatorios.

### 7.- Esquematizar los pasos del análisis estadístico.

El análisis estadístico a realizar dependerá de: los objetivos cifrados en el paso 1, el diseño seleccionado en el paso 3 y su modelo asociado que se habrá especificado en el paso 6.

Ahora es el momento de esquematizar los pasos del análisis, incluyendo las estimaciones, contrastes e intervalos de confianza que se calcularán. El análisis debe incluir un completo ejercicio de diagnosis y crítica del grado de ajuste del modelo a la realidad.

### 8.- Determinar el tamaño muestral.

Cálculo previo del número de observaciones que se deben tomar para alcanzar los objetivos del experimento.

Existen, dependiendo del modelo, algunas fórmulas para determinar este tamaño. Todas ellas sin embargo requieren el conocimiento del tamaño de la variabilidad no planificada (no sistemática y sistemática si es el caso) y, estimarlo a priori no es fácil, siendo aconsejable sobrestimarla. Habitualmente suele estimarse a partir del experimento piloto y en base a experiencias previas en trabajos con diseños experimentales semejantes.

**9.- Revisar las decisiones anteriores. Modificar si es necesario.**

Sin duda, el proceso de recogida de datos será, casi siempre, la tarea que mayor tiempo consume. Resulta sin embargo una obviedad afirmar que dedicar una gran cantidad de tiempo a pensar concienzudamente en la planificación previa, detallando los items anteriores, asegurará que los datos sean utilizados de la forma más eficiente posible.

*Ningún método de análisis estadístico, por sofisticado que sea, permitirá extraer conclusiones correctas en un diseño de experimentos mal planificado.*

Recíprocamente, debe quedar claro que el análisis estadístico es un paso más, completamente integrado en el proceso de planificación:

*El análisis estadístico no es un segundo paso independiente de la tarea de planificación. Es necesario comprender la totalidad de objetivos que se buscan antes de comenzar con el análisis. Si esto no se hace así, tratar que el experimento conteste a otras cuestiones a posteriori puede ser (lo será casi siempre) imposible.*

Pero no sólo los objetivos estarán presentes al empezar con el análisis sino también la técnica experimental empleada. Una regla de oro en la experimentación es

*No invertir nunca todo el presupuesto en un primer conjunto de experimentos y utilizar en su diseño toda la información previa disponible.*

Por último,

*Toda persona implicada en la ejecución del experimento y en la recolección de los datos debe ser informada con precisión de la estrategia experimental diseñada.*

Términos estándar en la metodología del diseño experimental son:

**Tratamientos:** Las condiciones experimentales que se desean comparar en el experimento (no confundir con factor tratamiento). En otras palabras, los distintos niveles de un factor tratamiento en un diseño con un sólo factor o las distintas combinaciones de niveles de factores tratamiento en un diseño factorial.

**Diseño Equilibrado o Balanceado:** Todos los tratamientos son asignados a un número igual de unidades experimentales.

**Tamaño del Experimento:** Número total de observaciones recogidas en el diseño.

## 1.4 Tres Principios Básicos

Al planificar un experimento, independientemente de todo lo comentado previamente, existen tres principios básicos a tener siempre en cuenta:

- **El principio de aleatorización.**
- **El bloqueo.**
- **La factorización del diseño.**

Los dos primeros son estrategias eficientes para asignar los tratamientos o condiciones experimentales a las unidades experimentales sin preocuparse de qué tratamientos considerar.

Por el contrario, el tercero define una estrategia eficiente para elegir los tratamientos sin considerar en absoluto como asignarlos después a las unidades experimentales.

**Aleatorizar:** *Todos aquellos factores no controlados por el experimentador en su diseño experimental y que pudiesen influir en los resultados serán asignados al azar (procedimiento objetivo que garantice la aleatoriedad) a las unidades experimentales.*

### ¿Por qué aleatorizar?

- Convierte la variabilidad sistemática no planificada en variabilidad no planificada o ruido aleatorio. En otros términos: previene contra la introducción de sesgos en el experimento.
- Evita la dependencia entre observaciones.
- Valida muchos de los procedimientos estadísticos más comunes.

**Bloquear:** *Dividir o particionar (pudiera ser también al azar) las unidades experimentales en grupos llamados bloques (o niveles de factores bloque) de modo que las observaciones realizadas en cada bloque (o nivel del factor bloque correspondiente) se realicen bajo condiciones experimentales lo más parecidas posibles. A diferencia de lo que ocurre con los factores tratamiento, el experimentador no está interesado en investigar las posibles diferencias de la respuesta entre los niveles de los factores bloque.*

### ¿Por qué bloquear?

- Convierte la variabilidad sistemática no planificada en variabilidad sistemática planificada.

### ¿Cómo elegir los bloques?

Bloquear es una muy buena estrategia siempre y cuando sea posible dividir las unidades experimentales en grupos de unidades similares, entendiendo por similares que probablemente conduzcan a valores semejantes de la respuesta. En ocasiones los principios de aleatorización y de bloqueo son incompatibles (ver diseños en bloques no aleatorizados).

**Ejemplo 1.1** Investigar posibles diferencias en la concentración de una determinada enzima en dos órganos (cerebro y corazón) de ratas. Aleatorizar: podría consistir en seleccionar al azar dos grupos de dos ratas cada uno y, al azar también, evaluar la respuesta en el cerebro de las ratas de un grupo y en el corazón de las ratas del otro grupo. Bloquear: introducir el factor bloque rata con cuatro niveles de observación (las cuatro ratas) y examinar a cada una el cerebro y el corazón. Obviamente el primer planteamiento es menos eficiente que el segundo (cuatro unidades experimentales frente a ocho), pero en el segundo los bloques no pueden ser aleatorizados.

**Diseño Factorial:** *Estrategia experimental que consiste en cruzar los niveles de todos los factores tratamiento a todas las combinaciones posibles.*

#### Ventajas de los diseños factoriales:

- Permiten detectar la existencia de efectos interacción entre niveles de factores tratamiento distintos.
- Es una estrategia más eficiente que la estrategia clásica de examinar la influencia de un factor manteniendo constantes los demás.

## 1.5 Un Ejemplo Real

A continuación se describe un experimento real descrito por Robert Peake en el Journal of Applied Statistics en noviembre de 1953. Es relativo al proceso de obtención del algodón.

**CONTEXTO:** En un paso intermedio del proceso de obtención de algodón se obtiene un hilo más grueso que el producto final llamado *roving*. Con unas guías rotatorias denominadas *flyers*, el roving se retuerce justo antes de enrollarlo a unas bobinas que tienen las máquinas. Cuanto más se retuerce, más fuerte resulta ser el algodón, pero más se ralentiza y encarece el proceso.

**1.- Objetivos.** Fundamentalmente dos:

- (1) Investigar el modo en que los diferentes grados de tensión sobre el roving (medidos en vueltas por pulgadas) afecta a la tasa de roturas del roving.
- (2) Comparar el flyer tradicional con uno de reciente introducción en el mercado.

## 2.- Identificar fuentes de variación.

### Factores Tratamiento y Niveles:

Factor 1: TIPO DE FLYER, con dos niveles codificados como 1 y 2.

Factor 2: GRADO DE TRABAJO SOBRE EL ROVING. Para elegir los niveles de este factor se pensó en un rango admisible y se ejecutó un experimento piloto para determinar los niveles más adecuados, optando por: 1'63, 1'69, 1'78 y 1'90 vueltas por pulgadas, que fueron codificados mediante 1, 2, 3 y 4 respectivamente.

| Flyer    | Grado |      |      |      |
|----------|-------|------|------|------|
|          | 1'63  | 1'69 | 1'78 | 1'90 |
| Estándar | (11)  | 12   | 13   | 14   |
| Nuevo    | 21    | 22   | 23   | (24) |

Los tratamientos 11 y 24 se omitieron porque el experimento piloto permitió comprobar que no daban lugar a resultados apropiados.

Unidades Experimentales: Una unidad experimental consistió en el hilo de algodón de un conjunto de bobinas de una máquina en un determinado día.

En este punto fue importante advertir tres hechos destacados. Primero: no era posible asignar bobinas distintas a tratamientos distintos en una misma máquina. Segundo: las bobinas deberían enrollarse completamente dado que la tensión, y consiguientemente la tasa de roturas, cambiaría a medida que se llenaba la bobina. Y, tercero, llevaba cerca de un día acabar de enrollar todas las bobinas de una máquina.

Factores Nuisance: Además de los factores tratamiento, se identificaron como otras posibles fuentes de variación: las máquinas, los operarios, el material experimental (el algodón) y las condiciones atmosféricas.

Hubo algún debate entre los experimentadores para determinar los factores de bloqueo. Algunos pensaban que, aunque el material sería lo más similar posible y las condiciones de humedad de la fábrica estaban muy controladas, podían originar importantes cambios en la respuesta a lo largo del tiempo, de modo que sería aconsejable considerar el día del experimento como un factor bloque. Sin embargo, finalmente se optó por ignorar la variabilidad del día-a-día y controlar únicamente a los operarios y las máquinas. Así se consideró un factor bloque cuyos niveles serían cada máquina con un único operario.

**3.- Elegir una regla de asignación de las unidades experimentales a los tratamientos.**

Buscando mantener lo más homogéneas posibles las condiciones experimentales dentro de cada bloque, se limitó a seis el número de unidades experimentales por bloque.

En particular se eligió un diseño conocido como **diseño en bloques completamente aleatorizados** (ver sección siguiente). Las seis unidades experimentales de cada bloque se asignaron al azar a los seis tratamientos.

| Bloque | Orden |    |    |     |     |     |
|--------|-------|----|----|-----|-----|-----|
|        | 1     | 2  | 3  | 4   | 5   | 6   |
| I      | 22    | 12 | 14 | 21  | 13  | 23  |
| II     | 21    | 14 | 12 | 13  | 22  | 23  |
| III    | 23    | 21 | 14 | 12  | 13  | 22  |
| IV     | 23    | 21 | 12 | ... | ... | ... |
| ⋮      | ⋮     | ⋮  | ⋮  | ⋮   | ⋮   | ⋮   |

**4.- Especificar las medidas que se realizarán, el procedimiento experimental y anticiparse a las posibles dificultades.** La respuesta consistiría en medir el número de roturas por cada cien libras de material.

Esta respuesta era muy apropiada porque el trabajo del operario incluía reparar las roturas, de modo que sería muy sencillo llevar un registro.

El experimento se realizaría durante la jornada rutinaria y las mayores dificultades serían: el tiempo perdido en tomar cada observación, la pérdida de producción causada por el cambio de flyers y el hecho de no saber por anticipado el número de máquinas que se precisan.

**5.- Ejecutar un experimento piloto.** Permitted identificar los niveles del factor 2. En este caso no se utilizó para estimar el tamaño de la variabilidad no planificada.

**6.- Especificar el modelo.** El modelo era:

$$\text{Tasa de Rotura} = \text{Constante} + \text{Efecto del Tratamiento} + \text{Efecto del Bloque} + \text{Error}$$

**7.- Esquematizar los pasos del análisis.** El análisis se orientó a evaluar las diferencias entre la tasa de roturas debidas a cada combinación de flyer/tensión. También se orientó a examinar la tendencia de la tasa de roturas cuando se incrementaba la tensión en cada flyer por separado.

**8.- Determinar el tamaño muestral.** En base a experiencias previas semejantes, aunque de otra naturaleza, se estimó la variabilidad experimental. Tal estimación implicaba que sería

necesario examinar 56 bloques (¡336 observaciones!) para detectar una diferencia real de al menos 2 roturas por 100 pulgadas con alta probabilidad.

**9.- Revisar las decisiones anteriores. Modificar si es necesario.** Dado que cada bloque suponía una semana de observación, se optó por analizar los datos una vez que se hubiesen examinado trece bloques. Por tanto la pretensión de detectar 2 roturas por 100 pulgadas con alta probabilidad fue desestimada.

## 1.6 Algunos Diseños Experimentales Estándar

Un diseño experimental es una regla que determina la asignación de las unidades experimentales a los tratamientos. Aunque los experimentos difieren unos de otros en muchos aspectos, existen diseños estándar que se usan con mucha frecuencia.

### 1.6.1 Diseños Completamente Aleatorizados

El experimentador asigna las unidades experimentales a los tratamientos completamente al azar, con la única restricción del número de observaciones que se tomarán en cada tratamiento.

De hecho, los valores de  $n_1, n_2, \dots, n_I$ , donde  $n_i$  denota el número de observaciones en el  $i$ -ésimo tratamiento,  $i = 1, \dots, I$ , determinan por completo las propiedades estadísticas del diseño.

Naturalmente, este tipo de diseño se utiliza en experimentos que no incluyen factores bloque.

El modelo matemático es de la forma:

$$\text{Respuesta} = \text{Constante} + \text{Efecto tratamiento} + \text{Error}$$

### 1.6.2 Diseños en Bloques o con Un Factor Bloque

El experimentador reagrupa las unidades experimentales en bloques, a continuación determina la distribución de los tratamientos en cada bloque y, por último, asigna al azar las unidades experimentales a los tratamientos dentro de cada bloque.

En el análisis estadístico de un diseño en bloques, éstos se tratan como los niveles de un único factor de bloqueo, aunque en realidad puedan venir definidos por la combinación de niveles de más de un factor nuisance.

El modelo matemático es:

$$\text{Respuesta} = \text{Constante} + \text{Efecto bloque} + \text{Efecto tratamiento} + \text{Error}$$

El diseño en bloques más simple es el llamado **Diseño en Bloques Completos** en el que cada tratamiento se observa el mismo número de veces en cada bloque.

Un diseño en bloques completos con una única observación por cada tratamiento se denomina **Diseño en Bloques Completamente Aleatorizados** o, simplemente, **Diseño en Bloques Aleatorizados**.

Cuando el tamaño del bloque es inferior al número de tratamientos no es posible observar la totalidad de tratamientos en cada bloque y se habla entonces de **Diseño en Bloques Incompletos**.

### 1.6.3 Diseños con Dos o Más Factores Bloque

En ocasiones se advierte la existencia de dos fuentes de variación lo suficientemente importantes como para ser designadas factores de bloque. En tal caso, ambos factores bloque pueden ser **cruzados** o **anidados**.

Se dicen cruzados cuando existen unidades experimentales en todas las combinaciones posibles de los niveles de los factores bloques. Si, en cambio, cada nivel particular de uno de los factores bloque ocurre en un único nivel del otro factor bloque, los factores bloque se dicen anidados. La Tabla 1.1 ilustra gráficamente la diferencia entre ambos tipos de bloqueo.

| Bloques Cruzados |   |          |   | Bloques Anidados |          |   |          |   |  |
|------------------|---|----------|---|------------------|----------|---|----------|---|--|
|                  |   | Bloque 1 |   |                  |          |   | Bloque 1 |   |  |
|                  |   | 1        | 2 | 3                |          | 1 | 2        | 3 |  |
|                  | 1 | *        | * | *                | 1        | * |          |   |  |
| Bloque 2         | 2 | *        | * | *                | 2        | * |          |   |  |
|                  | 3 | *        | * | *                | 3        | * |          |   |  |
|                  |   |          |   |                  | Bloque 2 | 4 | *        |   |  |
|                  |   |          |   |                  |          | 5 | *        |   |  |
|                  |   |          |   |                  |          | 6 | *        |   |  |
|                  |   |          |   |                  |          | 7 |          | * |  |
|                  |   |          |   |                  |          | 8 |          | * |  |
|                  |   |          |   |                  |          | 9 |          | * |  |

Tabla 1.1: Plan esquemático de experimentos con dos factores bloque

**Diseño con Factores Bloque Cruzados.** También se denomina **diseño fila-columna** y, como se ha dicho, se caracteriza porque existen unidades experimentales en todas las *celdas* (intersecciones de fila y columna).

El modelo matemático es:

$$\text{Respuesta} = \text{Constante} + \text{Efecto bloque fila} + \text{Efecto bloque columna} + \text{Efecto tratamiento} + \text{Error}$$

**Ejemplo 1.2** Si en el experimento del proceso de obtención de algodón, se garantiza que una combinación arbitraria máquina-operario se podrá utilizar cualquier día, entonces la variación debida al día-a-día se podría controlar mediante un nuevo factor bloque: *día de la prueba*, con seis niveles:  $1, \dots, 6$ . En tal caso un diseño fila-columna básico (previo a la aleatorización) como el de la Tabla 1.2 sería adecuado:

| Máq/Oper | Día |    |    |    |    |    |
|----------|-----|----|----|----|----|----|
|          | 1   | 2  | 3  | 4  | 5  | 6  |
| 1        | 12  | 13 | 14 | 21 | 22 | 23 |
| 2        | 13  | 14 | 21 | 22 | 23 | 12 |
| 3        | 14  | 21 | 22 | 23 | 12 | 13 |
| 4        | 22  | 23 | 12 | 13 | 14 | 21 |
| 5        | 23  | 12 | 13 | 14 | 21 | 22 |
| 6        | 21  | 22 | 23 | 12 | 13 | 14 |

Tabla 1.2: Un diseño en cuadrado latino para el experimento del algodón.

Obsérvese que si la cabeceras de las columnas se ignoran resulta un diseño en bloques completamente aleatorizados. Lo mismo ocurre si se ignoran las cabeceras de las filas. Este diseño es muy popular y se suele denominar **Diseño en Cuadrado Latino**.

**Diseño con Factores Bloque Anidados o Jerarquizados.** Dos factores bloque se dicen anidados cuando observaciones pertenecientes a dos niveles distintos de un factor bloque están automáticamente en dos niveles distintos del segundo factor bloque.

**Ejemplo 1.3** Se diseña un experimento para comparar el efecto de un número de dietas (tratamientos) sobre la ganancia en peso (respuesta) de pequeños cerdos (unidades experimentales). El metabolismo de los cerdos es, naturalmente, extremadamente variable. Parte de esta variabilidad puede ser controlada si advertimos que las respuestas serán más semejantes en cerdos procedentes de una misma camada. Más aún, camadas procedentes de una misma cerda presentarán mayor similitud que camadas de cerdas distintas. Así pues, las cerdas y las camadas podrían ser factores bloque. Obviamente, cada cochinito procede de una única camada que, a su vez, lo es de un única cerda. O sea, los cochinitos se encuentran anidados en las camadas y éstas en las cerdas. La asignación aleatoria de las dietas a las unidades experimentales se haría separadamente, camada por camada, de forma análoga a un diseño en bloques.

## 1.7 Más Ejemplos Reales

En esta sección se exponen más ejemplos de planificaciones experimentales para situaciones reales muy concretas. Se siguen los pasos de la lista expuesta en la sección 1.3 hasta el relativo al experimento piloto. Se propone completar el resto de pasos en ejercicios de capítulos posteriores.

### Experimento con Pastillas de Jabón. (Suyapa Silvia, 1985)

**1.- Definir los objetivos del experimento.** Se trata de comparar la disolución en agua de tres tipos de jabón. Los objetivos son fundamentalmente dos: (i) ¿existen diferencias en las pérdidas de peso de los distintos tipos de jabón luego de permanecer en agua un mismo intervalo de tiempo? y (ii) ¿cómo son esas diferencias?

Es importante caer en la cuenta que, de existir, las diferencias observadas no son generalizables a otras marcas de jabón del mismo tipo, ya que siempre existe alguna diferencia en las composiciones y mezclas de componentes. Las limitaciones del laboratorio donde se va a realizar el experimento no permiten simular el tratamiento habitual que se da al jabón (fricción, . . .) y, consiguientemente, las conclusiones se limitan a las condiciones impuestas en la experimentación.

**2.- Identificar las posibles fuentes de variación.** (i) *Factores tratamiento y niveles.* Existe un factor tratamiento, el tipo de jabón, con tres niveles de efectos fijos: una marca de jabón habitual, una marca de jabón desodorante y una marca de jabón hidratante. Los niveles son de efectos fijos porque las marcas concretas eran de particular interés para el experimentador.

Las pastillas de jabón se compraron en hipermercados locales y se cortaron en pequeños cubos de tamaño y peso similares. Los cubos se extrajeron de cada pastilla con una pequeña sierra metálica, muy fina y afilada, de modo que las caras de los cubos eran muy suaves. En realidad se cortaron los cubos y después se rebajaron todos hasta tener tamaño y peso similares al cubo más pequeño (con una precisión de diez miligramos). Antes del experimento se registraron los pesos de cada cubo.

Nótese que el experimentador no tiene ni idea de la antigüedad de las pastillas de jabón. Él ha supuesto que son una muestra típica de las que habitualmente se encuentran en los hiper. De entrada conviene resaltar que cada cubo se cortó en pastillas procedentes de una muestra aleatoria de hipermercados.

(ii) *Unidades experimentales.* Se calentaba el agua a  $100^{\circ}F$  (temperatura más o menos usual en un baño) y rápidamente se cubría con ella la cuarta parte de unas tazas metálicas idénticas (un estudio piloto mostró que con la cuarta parte llegaba para cubrir los cubos de jabón).

Las unidades experimentales eran las tazas con el agua y se asignarían a los tratamientos como se indica más adelante.

(iii) *Factores bloque, factores ruido y covariables.* Los tamaños iniciales de los cubos no eran idénticos y las secciones de las tazas con agua no estarían todas necesariamente expuestas al mismo grado de ventilación.

Con respecto a la primera fuente de variación podría considerarse el peso inicial como una covariable. Sin embargo el experimentador optó por controlar su efecto restando al peso final su peso inicial.

La segunda fuente de variación podría haber sido bloqueada, sin embargo se pensó que la variabilidad causada por esas diferencias de temperatura sería menor y no demasiado preocupante. Otras fuentes de variación, también consideradas como menores, serían las imprecisiones en las medidas de los pesos y de la temperatura del agua.

**3.- Elegir una regla de asignación de las u.e. a los tratamientos.** Se planificó un diseño balanceado: se necesitarían  $r$  cubos de cada tipo de jabón que se asignarían al azar a las tazas usando una tabla de números aleatorios.

**4.- Especificar las medidas a realizar, el procedimiento experimental y posibles dificultades.** Los cubos serían introducidos con mucho cuidado e inmediatamente se sellarían las tazas con papel aluminico para evitar una excesiva pérdida de humedad. Las tazas se colocarían lo más cerca posible unas de otras para que recibieran el mismo grado de ventilación (en todo caso, si eso influía significativamente en la temperatura del agua, las diferencias se habrían distribuido al azar entre los tratamientos debido a la regla de asignación).

Al cabo de 24 horas se vaciarían las tazas dejando filtrar el agua y poniendo los cubos a secar durante cuatro días. A continuación se pesarían de nuevo los cubos.

Dificultades esperadas: (i) A lo peor se requiere más tiempo para que se disuelvan los cubos. De ser así no se detectarían diferencias entre los pesos pre y post-experimento.

(ii) El tiempo de secado puede ser mayor que el supuesto y distinto para los tres tipos de jabón. ¿Estarán completamente secos?

Dificultades encontradas: (i) Cuando los cubos se depositaron en el agua caliente se hizo evidente que algunos jabones absorbían agua mucho más deprisa que otros. De hecho, algunos cubos no estuvieron totalmente cubiertos de agua las 24 horas. Para no alterar el diseño se optó por no añadir más agua.

(ii) El tiempo de secado fue distinto. El jabón normal estaba todavía húmedo (incluso parecía más grande que al principio) cuando los otros dos tipos empezaban a astillarse y separarse. Esto sí que generó un serio problema: la pérdida de peso no se podría evaluar hasta que los cubos estuviesen totalmente secos. Se optó por esperar otros dos días de modo que el jabón normal perdió parte del agua retenida.

### **Experimento con Pilas.** (Dean y Voss, 1999)

**1.- Definir los objetivos del experimento.** Debido al elevado gasto de una familia en pilas para linternas, el padre de familia desea investigar que tipo de pila no recargable es más económica. Le interesa comparar la duración por unidad de coste de dos marcas: una particular que usa con frecuencia y otra marca estándar del hiper donde acostumbra a realizar la compra

diaria. Además, quiere saber si merece la pena pagar el sobreprecio que las pilas alcalinas tienen con respecto a las pilas de larga duración.

Un objetivo posterior sería comparar las duraciones de las pilas independientemente del coste.

**2.- Identificar las posibles fuentes de variación.** (i) *Factores tratamiento y niveles.* Existen en este experimento una serie de fuentes de variación muy fáciles de determinar. Además de los factores tipo de pila y marca, que definirán los tratamientos, se podrían incluir: las fechas de fabricación de las pilas, como controlar la duración, la temperatura del entorno, la edad y la variabilidad de las lámparas de las linternas, ...

Las pilas que se van a usar en el experimento se compraron en fechas y lugares distintos al objeto de obtener una amplia variedad de fechas de fabricación. Si se hubiesen apuntado las fechas de los envases, la variabilidad de esta fuente podría haberse planificado a través de una covariable; al no ser así, su variabilidad engrosa la variabilidad no planificada.

Se fijaron las condiciones de ejecución del experimento: se encenderían las linternas y se evaluaría su duración. Sin duda este sistema no reproduce las condiciones de uso habituales de las linternas pero permite controlar muy fácilmente las condiciones de experimentación.

El experimento se ejecutará en una habitación donde la temperatura será prácticamente estable durante las horas de funcionamiento de las pilas. Además, no se espera que pequeñas fluctuaciones de la temperatura ambiente tengan una fuerte incidencia en la respuesta.

La variabilidad debida a la edad de las bombillas sí es más preocupante y, por ende, más difícil de manipular. Hubo que decidir entre usar una nueva bombilla para cada observación (arriesgándose a confundir los efectos de bombilla y pila) o usar una misma bombilla durante todo el experimento (con el riesgo de sesgar los resultados por la edad de la bombilla). Una tercera posibilidad sería agrupar las observaciones en bloques y usar una única bombilla para cada bloque pero bombillas distintas para bloques distintos. Al final se optó por la primera de las opciones debido a que el tiempo de vida de una bombilla resultaba considerablemente mayor que el de una pila.

Resumiendo:

(i) *Factores tratamiento y niveles.* Dos factores tratamiento, cada uno de ellos con dos niveles de efectos fijos. Se codificaron de dos formas distintas:

| Codificación 1 | Codificación 2 | Tratamientos                     |
|----------------|----------------|----------------------------------|
| 1              | 11             | alcalina, marca particular       |
| 2              | 12             | alcalina, marca del hiper        |
| 3              | 21             | larga duración, marca particular |
| 4              | 22             | larga duración, marca del hiper  |

(ii) *Unidades experimentales.* Una unidad experimental es el orden temporal en que se obtiene cada observación. Se asignaron al azar a los cuatro tipos de pilas (tratamientos).

(iii) *Factores bloque, factores ruido y covariables.* No hubo.

**3.- Elegir una regla de asignación de las u.e. a los tratamientos.** Dado que no se planificaron otras fuentes de variación se optó por un diseño completamente aleatorizado.

**4.- Especificar las medidas a realizar, el procedimiento experimental y posibles dificultades.** La primera dificultad radica en cómo medir la duración de una pila ya que cada linterna usa dos. Se optó por manipular cada linterna mediante un circuito que conectaba a la bombilla con una única pila. Una vez más no se reproducen las condiciones de uso habituales pero, dado que las condiciones de ejecución se iban a mantener constantes, se creyó que la tasa relativa duración de la pila por unidad de coste se preservaría.

Otra dificultad estriba en determinar cuando se acaba la pila. Primero porque tardaría unas cuantas horas y no todas las observaciones podrían ser monitorizadas y, segundo, porque la luz de la bombilla se va atenuando poco a poco a medida que la pila se está terminando. Para evitar estos dos problemas simultáneamente se optó por conectar al circuito un pequeño reloj que se pararía justo antes de apagarse por completo la bombilla. El tiempo marcado en el reloj se computaría como duración de la bombilla.

Finalmente, el coste de la pila equivaldría a la mitad del valor de un pack doble de pilas y la respuesta: *duración por coste unitario*, vendría dada en minutos por dólar.

**5.- Ejecutar un experimento piloto.** Se ejecutó un experimento piloto con pocas observaciones, lo que, de entrada, permitió garantizar que el circuito ideado trabajaría adecuadamente. En particular se descubrió que bombilla y reloj tenían que estar conectados en paralelo y no en serie (como se proyectó en un principio). El experimento piloto también dio una idea de la duración aproximada de cada observación (al menos cuatro horas) y proporcionó una primera estimación de la variabilidad del error que fue útil para determinar (en el paso 8 de la lista) que serían necesarias cuatro observaciones para cada tratamiento.

*Dificultades encontradas:* Sólo una: al ejecutar la cuarta observación se descubrió que el reloj continuaba funcionando cuando la bombilla ya no alumbraba. Se había desconectado el sistema. Se reparó y se continuó el experimento sin dificultades adicionales.

### **Experimento de la Cocción de un Pastel.** (Compañía Spillers Ltd., 1979)

**1.- Definir los objetivos del experimento.** Interesa saber cómo afecta a la calidad final de un pastel la cantidad de glicerina y ácido tartárico que se añade a su masa antes de la cocción.

**2.- Identificar las posibles fuentes de variación.**

(i) *Factores tratamiento y niveles.* Dos factores tratamiento: la cantidad de glicerina y la cantidad de ácido tartárico. Los experimentadores estaban familiarizados con la cocción y la determinación de calidad de los pasteles. Su conocimiento previo les permitía saber qué niveles de cada factor tratamiento eran interesantes para la experimentación. Seleccionaron cuatro

grados de glicerina (codificados del 1 al 4) y tres de ácido tartárico (codificados del 1 al 3); de modo que, en ambos casos, se trataba de cantidades equiespaciadas. La codificación para las doce combinaciones de tratamientos resultantes fue 11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 33, y 34.

(ii) *Unidades experimentales.* Antes de determinar las unidades experimentales era preciso tener claro el procedimiento experimental. Cada volumen de masa de pastel se dividiría en doce porciones y cada una de estas porciones recibiría una combinación distinta de las cantidades de glicerina y ácido tartárico que se investigan. Se amasaría meticulosamente cada mezcla y luego se introduciría en un envase (un molde) para horno. Los moldes se colocarían en una bandeja del horno a una temperatura dada y durante un período de tiempo establecido. En cada estante del horno sólo era posible acoplar una bandeja y, para obtener una medida correcta de la calidad del pastel, era preciso llenar cada bandeja. Una unidad experimental sería por tanto: *un estante del horno con una bandeja de moldes llenos de masa del pastel.* Dichas unidades experimentales serían asignadas al azar a los tratamientos.

(iii) *Factores bloque, factores ruido y covariables.* Se determinaron dos factores bloque cruzados: el momento del día en que se cocía (con dos niveles: mañana y tarde) y el horno (con tres niveles, uno por cada horno utilizado). Dado que cada horno tenía seis estantes, cada celda (intersección horno y momento del día) contendría seis unidades experimentales ¿Cómo? Cada conjunto de seis unidades experimentales se asignó al azar a seis de los doce tratamientos pero, previamente, se estipuló qué seis tratamientos concretos serían sorteados entre cada conjunto de seis unidades experimentales (ver paso 3).

Posiblemente habría alguna variabilidad en los resultados debido a las diferentes bandejas de los hornos pero, con total seguridad, esta fuente de variabilidad sería menor. Así, no se optó por un tercer factor de bloqueo definido por las bandejas.

Resultaba muy sencillo controlar la cantidad de masa a colocar en cada molde, pese a lo cual no se midieron los pesos pre-experimentales de cada porción (obviando así introducir una covariable en el diseño). En realidad se sabía que pequeñas diferencias entre estos pesos iniciales no afectarían a la calidad final del producto.

**3.- Elegir una regla de asignación de las u.e. a los tratamientos.** Se determinó un diseño fila-columna con seis unidades experimentales por celda. Obviamente tal diseño no permitiría observar a los doce tratamientos en cada celda, sin embargo, la planificación se realizó de modo que hubiese observaciones de todos los tratamientos en cada horno, bien por la mañana bien por la tarde. Esto parecía recomendable por pura precaución: si algún horno no funcionaba el día del experimento aún se disponía de dos observaciones de cada tratamiento. El diseño básico (previo a la aleatorización) se desglosa en la Tabla 1.3.

| Código<br>horno | Código del momento del día |    |    |    |    |    |    |    |    |    |    |    |
|-----------------|----------------------------|----|----|----|----|----|----|----|----|----|----|----|
|                 | 1                          |    |    |    |    |    | 2  |    |    |    |    |    |
|                 | 1                          | 11 | 13 | 22 | 24 | 32 | 34 | 12 | 14 | 21 | 23 | 31 |
| 2               | 12                         | 14 | 21 | 23 | 32 | 34 | 11 | 13 | 22 | 24 | 31 | 33 |
| 3               | 12                         | 14 | 22 | 24 | 31 | 33 | 11 | 13 | 21 | 23 | 32 | 34 |

Tabla 1.3: Diseño básico para el experimento de la cocción.



## Capítulo 2

# Diseños con Una Fuente de Variación

### 2.1 Introducción

Un diseño es una regla para determinar como asignar las unidades experimentales a los distintos tratamientos. El diseño más simple posible es el **Diseño Completamente Aleatorizado**, donde las unidades experimentales se asignan a los tratamientos completamente al azar, con la única restricción del número de observaciones a tomar para cada tratamiento. Este tipo de diseños no incluye factores bloque.

### 2.2 Aleatorización

Veamos a continuación un procedimiento muy sencillo para aleatorizar un diseño completamente aleatorizado. En este capítulo se empleará la siguiente notación:

$n_i$  representa el número de unidades experimentales que reciben el  $i$ -ésimo tratamiento,

$I$  denota el número total de tratamientos,

$n = \sum_{i=1}^I n_i$  denota el número total de observaciones.

De modo que podemos codificar los tratamientos de 1 a  $I$  y etiquetar las unidades experimentales de 1 a  $n$ . Entonces se seguirán los siguientes pasos:

**Paso 1:** Crear una columna con  $n_1$  unos,  $n_2$  doses,  $\dots$ ,  $n_I$   $I$ 'es; representando a los tratamientos.

**Paso 2:** Crear una segunda columna con  $n$  números aleatorios (uniformes en  $[0, 1]$ ) con dígitos suficientes para evitar empates.

**Paso 3:** Reordenar simultáneamente ambas columnas de modo que los números aleatorios se reubiquen en orden ascendente. Este paso consigue acomodar las etiquetas de los tratamientos en un orden aleatorio.

**Paso 4:** Asignar la unidad experimental  $t$  al tratamiento cuya etiqueta se ubica en la fila  $t$ .

### 2.3 Modelo Matemático para un Diseño Completamente Aleatorizado

En este caso el modelo matemático describirá la dependencia de la variable respuesta con respecto a los tratamientos (combinaciones de distintos niveles de los factores tratamientos involucrados en el diseño).

Denotemos por:

- $Y_{it}$  a la variable aleatoria que representa el valor de la respuesta en la  $t$ -ésima observación del  $i$ -ésimo tratamiento. En adelante usaremos la notación  $Y_{it}$  para referirnos a la variable y la notación  $y_{it}$  para referirnos a una observación concreta.
- $\mu_i$  a la respuesta real en el  $i$ -ésimo tratamiento. Es decir, a la respuesta que se obtendría siempre con el  $i$ -ésimo tratamiento si se ejecutase el experimento en, *exactamente*, las mismas condiciones.
- $\varepsilon_{it}$  a la variable aleatoria  $Y_{it} - \mu_i$ . O sea, la diferencia entre la  $t$ -ésima observación del  $i$ -ésimo tratamiento y su valor real. Por tanto  $\varepsilon_{it}$  agrupa la contribución de las fuentes de variación menores y no planificadas. En adelante la llamaremos variable **error** o **error experimental**, si bien el término error no tiene aquí un carácter peyorativo.

Para cada  $t = 1, \dots, n_i$ ,  $i = 1, \dots, I$ , el modelo matemático es:

$$\underbrace{Y_{it}}_{\text{aleatorio}} = \overbrace{\mu_i}^{\text{determinista}} + \underbrace{\varepsilon_{it}}_{\text{aleatorio}} \quad (2.1)$$

Un modelo alternativo se obtiene al reparametrizar la respuesta real por tratamiento en la forma:  $\mu_i = \mu + \tau_i$ ,  $i = 1, \dots, I$ , resultando:

$$\underbrace{Y_{it}}_{\text{aleatorio}} = \overbrace{\mu + \tau_i}^{\text{determinista}} + \underbrace{\varepsilon_{it}}_{\text{aleatorio}} \quad (2.2)$$

de modo que  $\mu$  es una constante y  $\tau_i$  representa la desviación, positiva o negativa, de esta constante cuando se observa el tratamiento  $i$ . Así, examinar las diferencias entre tratamientos equivale a examinar las diferencias entre los parámetros  $\mu_i$  en el modelo (2.1) o entre los parámetros  $\tau_i$  en el modelo (2.2). Los parámetros  $\tau_i$  suelen llamarse **efectos**.

En la sección siguiente se muestra que en el segundo modelo no se pueden estimar de forma única los parámetros.

Los modelos (2.1) y (2.2) son *modelos lineales* que, si bien raramente se adecúan a una situación real, permiten obtener aproximaciones razonables a modelos más complejos.

Sobre las variables error, que engloban a todas las fuentes de variación menores del experimento, incluyendo los errores de medida, es necesario establecer una serie de hipótesis iniciales acerca de su comportamiento en distribución en aras de poder realizar el análisis estadístico. Para este modelo las hipótesis sobre el error son:

$$\begin{aligned} \varepsilon_{it} \text{ son mutuamente independientes e idénticamente} \\ \text{distribuidas (i.i.d.) según una } N(0, \sigma) \quad \forall i, t \\ \Downarrow \\ \forall i : Y_{it} \text{ son i.i.d. según una } N(\mu + \tau_i, \sigma), \forall t \end{aligned}$$

## 2.4 Estimación de los Parámetros

### 2.4.1 Funciones Estimables de Parámetros

En el modelo matemático (2.1) hay  $I + 1$  parámetros a estimar:  $\mu_1, \mu_2, \dots, \mu_I$  y  $\sigma$ , en tanto que en el modelo (2.2) hay que estimar  $I + 2$ :  $\mu, \tau_1, \tau_2, \dots, \tau_I$  y  $\sigma$ .

Una función de parámetros se dice *estimable* si y sólo si puede escribirse como el valor esperado de una combinación lineal de variables respuesta.

Técnicamente, la ventaja de las funciones estimables de parámetros es que para ellas existe un único estimador lineal insesgado y, en ese sentido, es deseable que el análisis estadístico de cualquier diseño incluya, preferiblemente, tan sólo funciones estimables de los parámetros.

En un diseño completamente aleatorizado cualquier función estimable de los parámetros es de la forma:

$$E \left[ \sum_{i=1}^I \sum_{t=1}^{n_i} a_{it} Y_{it} \right] = \sum_{i=1}^I \sum_{t=1}^{n_i} a_{it} E[Y_{it}] = \sum_{i=1}^I \sum_{t=1}^{n_i} a_{it} (\mu + \tau_i) = \sum_{i=1}^I b_i (\mu + \tau_i)$$

siendo  $b_i = \sum_{t=1}^{n_i} a_{it}$  y los  $a_{it}$  números reales arbitrarios.

Obsérvese que, por ejemplo:

- $\mu_i = \mu + \tau_i$  es estimable para todo  $i$ . En efecto, basta tomar  $b_i = 1$  y  $b_j = 0$  para todo  $j \neq i$ .

- No existen valores para los  $b_i$  que permitan obtener los parámetros  $\mu, \tau_1, \dots, \tau_I$ . Luego estos parámetros no son individualmente estimables.

- Cualquier función de la forma  $\sum_{i=1}^I b_i \tau_i$  tal que  $\sum_{i=1}^I b_i = 0$  se denomina **contraste**. Pues bien, todos los contrastes son estimables en nuestro modelo.

En particular, haciendo  $b_r = 1, b_s = -1$  y  $b_i = 0 \forall i \neq r, s$ , el contraste resulta ser  $\tau_r - \tau_s$ . Por tanto las diferencias entre dos efectos cualesquiera son estimables.

### 2.4.2 Estimadores Puntuales Mínimo-Cuadráticos

Para estimar los parámetros se utiliza el método de *Estimación Mínimo Cuadrática*, que consiste en seleccionar como estimadores los valores de los parámetros que minimizan la suma de los cuadrados de los errores. O sea, se trata de seleccionar valores  $\hat{\mu}, \hat{\tau}_1, \dots, \hat{\tau}_I$  que verifiquen:

$$\sum_{i=1}^I \sum_{t=1}^{n_i} [y_{it} - (\hat{\mu} + \hat{\tau}_i)]^2 = \min_{\mu, \tau_1, \dots, \tau_I} \sum_{i=1}^I \sum_{t=1}^{n_i} \overbrace{[y_{it} - (\mu + \tau_i)]^2}^{e_{it}^2}$$

El problema de minimización anterior conduce a un sistema de  $I + 1$  ecuaciones (se llaman *ecuaciones normales*) con rango  $I$  que aporta soluciones únicas para cada  $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i$ :

$$\boxed{\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \frac{1}{n_i} \sum_{t=1}^{n_i} y_{it} = \bar{y}_i.} \tag{2.3}$$

No existe solución mínimo cuadrática única para los parámetros  $\mu, \tau_1, \dots, \tau_I$  debido a que el sistema de ecuaciones tiene rango  $I$ . Aunque a efectos prácticos cualquiera de las infinitas soluciones es válida, es habitual añadir alguna restricción más al objeto de obtener una solución única. Veamos algunos de los ejemplos más usados:

- Imponer  $\hat{\mu} = 0$ , de donde:  $\begin{cases} \hat{\mu} = 0 \\ \hat{\tau}_i = \bar{y}_i, \forall i = 1, \dots, I \end{cases}$
- Imponer  $\sum_{i=1}^I n_i \hat{\tau}_i = 0$ , de donde:  $\begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^I \sum_{t=1}^{n_i} y_{it} = \bar{y}_{..} \\ \hat{\tau}_i = \bar{y}_i - \bar{y}_{..}, \forall i = 1, \dots, I \end{cases}$
- Imponer  $\hat{\tau}_I = 0$ , de donde:  $\begin{cases} \hat{\mu} = \bar{y}_I. \\ \hat{\tau}_i = \bar{y}_i - \bar{y}_I, \forall i \neq I \end{cases}$

La última de las restricciones es la empleada por el paquete estadístico SPSS.

Una propiedad fundamental de los estimadores mínimo-cuadráticos la establece un resultado clave en los problemas de modelización lineal estadística, el **Teorema de Gauss-Markov**, según el cual:

Para todo modelo lineal cuyas variables error sean normales, independientes y con varianza común  $\sigma^2$ , los estimadores mínimo-cuadráticos de cualquier función estimable son únicos, insesgados y de varianza mínima.

Se tiene además que, como los estimadores mínimo-cuadráticos de  $\mu + \tau_i$ ,  $i = 1, \dots, I$ , tienen distribución normal:

$$\boxed{\bar{Y}_i \sim N\left(\mu + \tau_i, \frac{\sigma}{\sqrt{n_i}}\right)} \quad (2.4)$$

y son todos ellos independientes (están basados en observaciones independientes), entonces el estimador mínimo-cuadrático de cualquier contraste  $\sum_{i=1}^I c_i \tau_i$ , con  $\sum_{i=1}^I c_i = 0$ , es también normal:

$$\boxed{\sum_{i=1}^I c_i \bar{Y}_i \sim N\left(\sum_{i=1}^I c_i \tau_i, \sigma \sqrt{\sum_{i=1}^I \frac{c_i^2}{n_i}}\right)} \quad (2.5)$$

### 2.4.3 Estimación Puntual de $\sigma^2$

Para cualquier modelo estadístico, se denominan **residuos** a las diferencias entre cada valor observado y el valor previsto por el modelo una vez que han sido estimados sus parámetros. Es decir:

$$\boxed{\underbrace{e}_{\text{Residuo}} = \underbrace{y}_{\text{Valor observado}} - \underbrace{\hat{y}}_{\text{Previsto modelo}}} \quad (2.6)$$

En el modelo actual, para todo  $t = 1, \dots, n_i$  e  $i = 1, \dots, I$  se tiene:

$$\boxed{e_{it} = y_{it} - \hat{y}_{it} = y_{it} - \hat{\mu}_i} \quad (2.7)$$

siendo  $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i$  los estimadores mínimo-cuadráticos de  $\mu_i = \mu + \tau_i$ . Sustituyendo los estimadores mínimo-cuadráticos de  $\mu_i$  por su valor,  $\bar{y}_i$ , se calcula **la suma de residuos al cuadrado** o **suma de cuadrados residual**:

$$\boxed{scR = \sum_{i=1}^I \sum_{t=1}^{n_i} e_{it}^2 = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_i)^2} \quad (2.8)$$

El valor concreto  $scR$  dado en (2.8) es una realización particular de la variable aleatoria  $SCR$  (que también llamaremos suma de cuadrados residual). Es decir, la variable aleatoria  $SCR$  viene dada por:

$$SCR = \sum_{i=1}^I \sum_{t=1}^{n_i} (Y_{it} - \bar{Y}_i)^2$$

Se demuestra que:

- $SCR = \sum_{i=1}^I \sum_{t=1}^{n_i} (Y_{it} - \bar{Y}_i)^2 = \sum_{i=1}^I (n_i - 1) \hat{S}_i^2$ , siendo  $\hat{S}_i^2$  la varianza muestral corregida del  $i$ -ésimo tratamiento:  $\hat{S}_i^2 = \frac{1}{n_i - 1} \sum_{t=1}^{n_i} (Y_{it} - \bar{Y}_i)^2$ .

- Como  $\hat{S}_i^2$  es un estimador insesgado de la varianza del error  $\sigma^2$ , el valor esperado de  $SCR$  es:

$$E(SCR) = \sum_{i=1}^I (n_i - 1) E(\hat{S}_i^2) = (n - I)\sigma^2$$

de modo que un estimador insesgado de  $\sigma^2$  es:

$$\boxed{\hat{\sigma}^2 = SCMR = \frac{SCR}{n - I}} \tag{2.9}$$

que recibe indistintamente los nombres de **error cuadrático medio**, **varianza residual** y **varianza dentro de los tratamientos**.

- $(n - I) \frac{SCMR}{\sigma^2} = \frac{SCR}{\sigma^2} \sim \chi_{n-I}^2$

Así, un intervalo de confianza, al  $100(1 - \alpha)\%$ , para la varianza  $\sigma^2$  debida a factores no incluidos en la planificación del experimento viene dado por:

$$\left( 0, \frac{scR}{\chi_{n-I, 1-\alpha}^2} \right)$$

En ocasiones se desea estimar el error cuadrático medio a partir de la realización de un experimento piloto. Como se explicará en la sección 2.6, una primera estimación de  $\sigma^2$  permitirá calcular el número aproximado de observaciones necesarias para garantizar una potencia predeterminada de los contrastes. Pues bien, para no infraestimar dicho número, una práctica aconsejable es seleccionar como estimación de  $\sigma^2$  el valor del extremo superior del anterior intervalo de confianza.

## 2.5 Análisis de la Varianza de Una Vía

### 2.5.1 Idea General

Interesa contrastar:

$$\begin{cases} H_0 : \tau_1 = \tau_2 = \dots = \tau_I \\ H_1 : \text{Existen } i, j, \text{ con } i \neq j, \text{ tales que } \tau_i \neq \tau_j \end{cases}$$

En principio la hipótesis nula se refiere a parámetros no estimables, sin embargo puede reescribirse en términos de  $I - 1$  contrastes estimables como sigue:

$$H_0 : \tau_1 - \tau_I = 0 \text{ y } \tau_2 - \tau_I = 0 \text{ y } \dots \text{ y } \tau_{I-1} - \tau_I = 0$$

Existen más formas de reescribir  $H_0$  en términos de funciones estimables de los parámetros pero cualquiera de ellas dependerá de  $I - 1$  contrastes distintos. El número  $I - 1$  se denomina **grados de libertad de los tratamientos**.

La idea básica del test **análisis de la varianza** es comparar la suma de cuadrados residual bajo el modelo matemático que se propone, (se denomina **modelo completo**), con la suma de cuadrados residual del modelo que resulta cuando la hipótesis nula es cierta (llamado **modelo reducido**).

Es decir:

$$\left\{ \begin{array}{l} H_0 : \left\{ \begin{array}{l} Y_{it} = \mu + \tau + \varepsilon_{it}^0, \text{ con } \varepsilon_{it}^0 \text{ i.i.d. según una } N(0, \sigma) \forall i, t \\ \text{(Modelo Reducido)} \end{array} \right. \\ \\ H_1 : \left\{ \begin{array}{l} Y_{it} = \mu + \tau_i + \varepsilon_{it}, \text{ con } \varepsilon_{it} \text{ i.i.d. según una } N(0, \sigma) \forall i, t \\ \text{(Modelo Completo)} \end{array} \right. \end{array} \right.$$

Se demuestra que el único estimador mínimo-cuadrático de  $\mu + \tau$  es la media muestral de todas las observaciones,  $\bar{y}_{..}$ , de modo que la suma de cuadrados residual del modelo reducido es:

$$scR_0 = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{..})^2$$

Si  $H_0$  es falsa (al menos dos efectos tratamiento difieren) la suma de cuadrados residual  $scR$  bajo el modelo completo es considerablemente más pequeña que la suma de cuadrados residual del modelo reducido  $scR_0$ . Por el contrario, si  $H_0$  es cierta ambas serán muy similares.

Ambas sumas de cuadrados se pueden relacionar descomponiendo las fuentes de variabilidad como se indica en el siguiente apartado. Se prueba que:

$$\boxed{scR_0 - scR = scT} \quad (2.10)$$

donde  $scT$  se denomina indistintamente **suma de cuadrados entre tratamientos** o **suma de cuadrados explicada por diferencias entre tratamientos**. Si la cantidad obtenida en (2.10) es grande en relación con  $scR$  se rechazará  $H_0$ .

### 2.5.2 Descomposición de la Variabilidad

$$\begin{array}{ccc} \begin{array}{c} \text{Suma de} \\ \text{Cuadrados} \\ \text{Global (scG)} \end{array} & \begin{array}{c} \text{Suma de} \\ \text{Cuadrados} \\ \text{Explicada (scT)} \end{array} & \begin{array}{c} \text{Suma de} \\ \text{Cuadrados} \\ \text{Residual (scR)} \end{array} \\ \hline \underbrace{\sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{..})^2}_{g.l. = n-1} & = \underbrace{\sum_{i=1}^I n_i (\bar{y}_{i.} - \bar{y}_{..})^2}_{g.l. = I-1} + \underbrace{\sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{i.})^2}_{g.l. = n-I} \end{array}$$

Dividiendo por los correspondientes grados de libertad se obtienen tres estimaciones distintas de  $\sigma^2$ :

$$\begin{aligned} \hat{\sigma}_1^2 &= \frac{1}{n-1} \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{..})^2 = scmG \\ \hat{\sigma}_2^2 &= \frac{1}{I-1} \sum_{i=1}^I n_i (\bar{y}_{i.} - \bar{y}_{..})^2 = scmT \\ \hat{\sigma}_3^2 &= \frac{1}{n-I} \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{i.})^2 = scmR \end{aligned}$$

Se verifica:

- Bajo  $H_0$ :  $\frac{SCT}{\sigma^2} \sim \chi_{I-1}^2$  y  $\frac{SCMT}{SCMR} \sim F_{I-1, n-I}$
- En general:  $E(SCMT) = \sigma^2 + Q(\tau_i)$ , siendo

$$Q(\tau_i) = \frac{1}{I-1} \sum_{i=1}^I n_i \left( \tau_i - \frac{1}{n} \sum_{j=1}^I n_j \tau_j \right)^2 \quad (2.11)$$

- Por tanto:  $\boxed{\text{Rechazar } H_0 \text{ al nivel } \alpha \text{ si } \hat{F} = \frac{scmT}{scmR} > F_{I-1, n-I, \alpha}}$

| <b>CUADRO DEL ANÁLISIS DE LA VARIANZA</b>  |   |             |                            |                     |                        |
|--|---|-------------|----------------------------|---------------------|------------------------|
| <b>— UNA VÍA —</b>   |   |             |                            |                     |                        |
| <i>Fuente de Variación</i>   | <i>Suma de Cuadrados</i>  | <i>g.l.</i> | <i>scm</i>                 | $\hat{F}$           | $E(SCM)$               |
| Tratamientos   | $scT = \sum_{i=1}^I \sum_{t=1}^{n_i} (\bar{y}_{i.} - \bar{y}_{..})^2$ | $I - 1$     | $scmT = \frac{scT}{I - 1}$ | $\frac{scmT}{scmR}$ | $\sigma^2 + Q(\tau_i)$ |
| Residual   | $scR = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{i.})^2$       | $n - I$     | $scmR = \frac{scR}{n - I}$ |                     |                        |
| Global   | $scG = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$       | $n - 1$     | $scmG = \frac{scG}{n - 1}$ |                     |                        |
| $Q(\tau_i) = \frac{1}{I - 1} \sum_{i=1}^I n_i \left( \tau_i - \frac{1}{n} \sum_{j=1}^I n_j \tau_j \right)^2$ |   |             |                            |                     |                        |
| Rechazar $H_0 : \tau_i = \tau_j \forall i, j$ en base al p-valor $p = P(\hat{F} > F_{I-1, n-I})$             |   |             |                            |                     |                        |
| Coeficiente de Determinación: $R^2 = \frac{scT}{scG}$  |   |             |                            |                     |                        |

Tabla 2.1: Cuadro del análisis de la varianza para un diseño completamente aleatorizado de efectos fijos.

## 2.6 Determinación del Tamaño Muestral en base al Contraste Anova

Generalmente los efectos reales de los tratamientos no serán nunca *exactamente* iguales. Es claro que el objetivo del experimento es detectar diferencias significativas y no es preocupante dejar de descubrir la existencia de diferencias irrelevantes entre tratamientos. Por ejemplo, muy posiblemente un consumidor no valore en exceso que las vidas medias reales de dos marcas de lavadoras difieran en una semana; sin embargo, sí podría estar interesado en conocer que esa diferencia sea de un año.

En muchos experimentos hay un valor mínimo, llámese  $\Delta$ , tal que el experimentador desea que el contraste ANOVA conduzca, con alta probabilidad, a rechazar la hipótesis nula de que todos los tratamientos son iguales cuando existan al menos dos de ellos cuyos efectos difieran en

una cantidad mayor o igual que  $\Delta$ .

El valor de  $\Delta$  lo determina el experimentador y muchos textos se refieren a él como *mínima diferencia significativa*. Se denomina **potencia** del contraste en  $\Delta$ , y se denota por  $\Pi(\Delta)$ , a la probabilidad de rechazar  $H_0$  cuando al menos dos de los tratamientos difieren en  $\Delta$ .

La potencia del contraste es función de los tamaños muestrales ( $n_i$ ), del número de tratamientos ( $I$ ), del nivel de significación ( $\alpha$ ) y de la varianza del error ( $\sigma^2$ ). Así, conocidos  $\Delta$ ,  $\Pi(\Delta)$ ,  $I$ ,  $\alpha$  y  $\sigma^2$ , se podrían determinar los tamaños muestrales  $n_i$ .

Ahora bien, mientras que a  $\Delta$ ,  $\Pi(\Delta)$ ,  $I$  y  $\alpha$  los elige el experimentador,  $\sigma^2$  debe ser previamente estimado a través de un experimento piloto o en base a informaciones previas procedentes de experimentos similares. En general conviene sobreestimar algo la varianza del error, ya que, en caso de infraestimarla podríamos no obtener la potencia deseada para el contraste Anova.

Veamos a continuación como deducir un algoritmo que permita calcular los tamaños muestrales  $n_i$  para obtener una potencia predeterminada  $\Pi(\Delta)$ , dados  $\Delta$ ,  $I$ ,  $\alpha$  y  $\sigma^2$ . Para facilitar los cálculos se supondrá que los tamaños muestrales son iguales para todos los tratamientos:  $r = n_1 = n_2 = \dots = n_I$ .

Como  $\Pi_\alpha(\Delta)$  es igual a la probabilidad de rechazar  $H_0$  cuando  $H_1$  es cierto (al menos dos tratamientos difieren en  $\Delta$ ), es preciso conocer la distribución de probabilidad del estadístico de contraste  $F = SCMT / SCMR$  bajo  $H_1$ . Se tiene:

$$\left\{ \begin{array}{l} \hat{F} | H_0 \sim F_{I-1, n-I} \quad (\text{F de Fisher-Snedecor}) \\ \hat{F} | H_1 \sim F_{I-1, n-I, \delta^2}^* \quad (\text{F de Fisher-Snedecor no centrada}) \end{array} \right.$$

siendo  $\delta^2$  el parámetro de *no centralidad* que para el ANOVA de una vía es:

$$\delta^2 = \frac{I-1}{\sigma^2} Q(\tau_i) \quad \underbrace{\text{si } r=n_i \forall i}_{= \quad} \quad \frac{r}{\sigma^2} \sum_{i=1}^I (\tau_i - \bar{\tau})^2$$

Entonces

$$\Pi_\alpha(\Delta) = P\left(\hat{F} > F_{I-1, n-I, \alpha} \mid \hat{F} \sim F_{I-1, n-I, \delta^2}^*\right) \quad \text{con} \quad \delta^2 = \frac{r}{\sigma^2} \sum_{i=1}^I (\tau_i - \bar{\tau})^2$$

En las tablas del Apéndice se obtiene  $\Pi_\alpha(\phi)$  como función de

$$\phi = \frac{\delta}{\sqrt{I}}, \quad \nu_1 = I - 1, \quad \nu_2 = n - I \quad y \quad \alpha$$

Por otro lado, la situación más desfavorable para detectar una diferencia real  $\Delta$  entre dos tratamientos cualesquiera es:

$$\mu + \tau_i = k + \frac{\Delta}{2}, \quad \mu + \tau_j = k - \frac{\Delta}{2} \quad y \quad \mu + \tau_s = k, \quad \forall s \neq i, j$$

En este caso:

$$\delta^2 = \frac{r}{\sigma^2} \sum_{i=1}^I (\tau_i - \bar{\tau})^2 = \frac{r\Delta^2}{2\sigma^2} \Rightarrow \phi^2 = \frac{\delta^2}{I} = \frac{r\Delta^2}{2I\sigma^2} \Rightarrow r = \frac{2I\sigma^2\phi^2}{\Delta^2}$$

Algoritmo

1. Encontrar sección tabla para  $\alpha$  y  $\nu_1 = I - 1$ .
2. En la primera iteración usar  $\nu_2 = 1000$ . En las siguientes considerar  $\nu_2 = n - I = I(r - 1)$ .
3. Según la potencia especificada para  $\Delta$ , deducir de la tabla el valor de  $\phi$  por interpolación.
4. Calcular  $r = \frac{2I\sigma^2\phi^2}{\Delta^2}$ , redondeando al entero más próximo.
5. Repetir los pasos 2 a 4 hasta que el valor de  $r$  no varíe o se mueva entre dos valores. En este último caso seleccionar el mayor.

**Ejemplo 2.1** Sean  $I = 3$ ,  $\Delta = 0'25$ ,  $\hat{S}_R^2 = 0'007$ ,  $\Pi(\Delta) = 0'90$  y  $\alpha = 0'05$ .

En la primera iteración:  $r = \frac{2I\sigma^2\phi^2}{\Delta^2} = 0'672\phi^2$ ,  $\nu_1 = 2$  y  $\nu_2 = 1000$ .

| $r$ | $\nu_2 = 3(r - 1)$ | $\phi$ | $r = 0'672\phi^2$ | Acción              |
|-----|--------------------|--------|-------------------|---------------------|
|     | 1000               | 2'33   | 3'65              | Redondear a $r = 4$ |
| 4   | 9                  | 2'48   | 4'13              | Redondear a $r = 5$ |
| 5   | 12                 | 2'33   | 3'65              | Redondear a $r = 4$ |

Conclusión: Tomar  $r = 5$  en cada tratamiento.

## 2.7 Inferencia de Contrastes y Medias de Tratamientos

### 2.7.1 Contrastes

En la sección 2.4.1 se denominó **contraste** a cualquier combinación lineal de los efectos de los tratamientos:

$$\sum_{i=1}^I b_i \tau_i \quad \text{tal que} \quad \sum_{i=1}^I b_i = 0 \quad (2.12)$$

En un diseño completamente aleatorizado todo contraste es estimable y su estimador mínimo-cuadrático es:

$$\sum_{i=1}^I b_i \bar{Y}_i.$$

que, debido a la normalidad e independencia de las observaciones, se distribuye según:

$$\boxed{\sum_{i=1}^I b_i \bar{Y}_i \sim N \left( \sum_{i=1}^I b_i \tau_i, \sigma \sqrt{\sum_{i=1}^I \frac{b_i^2}{n_i}} \right)} \quad (2.13)$$

Muchas veces es útil representar un contraste por la lista de sus coeficientes. Es decir, el contraste puede representarse en cualquiera de las dos formas equivalentes siguientes:

$$\sum_{i=1}^I b_i \tau_i \quad \iff \quad [b_1, b_2, \dots, b_I]$$

Un contraste se dice **normalizado** si sus coeficientes en la combinación lineal se reescalan para que la varianza del estimador coincida con la varianza del error  $\sigma^2$ .

De (2.13) se deduce que para normalizar un contraste basta considerar los nuevos coeficientes  $h_i = \left( \sum_{i=1}^I \frac{b_i^2}{n_i} \right)^{-1/2} b_i$ . Entonces:

$$\boxed{\sum_{i=1}^I h_i \bar{Y}_i \sim N \left( \sum_{i=1}^I h_i \tau_i, \sigma \right) \quad \text{donde} \quad h_i = \frac{b_i}{\sqrt{\sum_{i=1}^I \frac{b_i^2}{n_i}}} } \quad (2.14)$$

Suele ser interesante realizar inferencia sobre contrastes como:

**Comparar tratamientos a pares (“pairwise”)** Es decir, estimar contrastes del tipo:  $\tau_i - \tau_j$ , donde la lista de coeficientes tiene un 1 en el  $i$ -ésimo lugar, un  $-1$  en el  $j$ -ésimo lugar y un 0 en el resto. Por ejemplo,  $[0, 0, 1, 0, \dots, 0, -1, 0]$  sería el contraste  $\tau_3 - \tau_{I-1}$ . Existen  $m = \binom{I}{2}$  contrastes de comparaciones por pares.

Obsérvese que  $\hat{\tau}_i - \hat{\tau}_j = \bar{Y}_i - \bar{Y}_j$ . De (2.13) se deduce que el estimador mínimo-cuadrático de este tipo de contrastes se distribuye:

$$\boxed{\bar{Y}_i - \bar{Y}_j \sim N\left(\tau_i - \tau_j, \sigma \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right)} \quad (2.15)$$

**Tratamientos frente a control** Un subconjunto de contrastes del grupo anterior muy particular es el formado por los  $I-1$  contrastes  $\tau_1 - \tau_I$  ( $[1, 0, \dots, 0, -1]$ ),  $\tau_2 - \tau_I$  ( $[0, 1, \dots, 0, -1]$ ),  $\dots$ ,  $\tau_{I-1} - \tau_I$  ( $[0, 0, \dots, 1, -1]$ ). El objetivo es comparar el efecto de cada uno de los tratamientos con un tratamiento concreto, que se suele denominar **control**, y que, en la notación anterior, se ha identificado con el tratamiento  $I$ .

**Diferencia de medias** Cuando los niveles de los factores tratamiento se dividen de un modo natural en dos o más grupos, puede ser de interés comparar el efecto medio de un grupo con los efectos medios de otros grupos.

Por ejemplo, supóngase que se desea comparar el efecto del color del papel de examen en los resultados finales de éste. Se ha probado con dos tonos claros: blanco y amarillo (niveles 1 y 2 del factor) y con tres tonos más fuertes: azul, verde y rosa (niveles 3, 4 y 5). El contraste:

$$\frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{3}(\tau_3 + \tau_4 + \tau_5) \Leftrightarrow \left[\frac{1}{2}, \frac{1}{2}, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right]$$

permite observar diferencias entre la influencia del papel claro con respecto a la del papel oscuro.

**Tendencias** Cuando los niveles del factor tratamiento son cuantitativos y tienen un orden natural, el experimentador podría estar interesado en saber si la respuesta crece o decrece con un incremento del nivel o, más aún, si esa tendencia se mantiene o no constante. Se habla entonces de **contrastos de tendencia**.

Los coeficientes de contrastes de tendencias para  $I$  niveles **igualmente espaciados** de un factor tratamiento con **tamaños muestrales iguales** se dan en la Tabla 2.2 para un número de tratamientos entre 3 y 7.

El método general para obtener coeficientes como los de la Tabla 2.2 (incluyendo tamaños muestrales distintos o niveles no igualmente espaciados) consiste en ajustar un modelo de regresión polinómica a los niveles no codificados del factor tratamiento. Para el caso lineal, es sencillo comprobar que los coeficientes se calculan a través de la siguiente expresión:

$$b_i = n_i \left( n_i x_i - \sum n_i x_i \right) \quad (2.16)$$

siendo  $x_i$  el valor del nivel no codificado.

| Tendencia  | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ | $b_6$ | $b_7$ |
|------------|-------|-------|-------|-------|-------|-------|-------|
| $I = 3$    |       |       |       |       |       |       |       |
| Lineal     | -1    | 0     | 1     |       |       |       |       |
| Cuadrática | 1     | -2    | 1     |       |       |       |       |
| $I = 4$    |       |       |       |       |       |       |       |
| Lineal     | -3    | -1    | 1     | 3     |       |       |       |
| Cuadrática | 1     | -1    | -1    | 1     |       |       |       |
| Cúbica     | -1    | 3     | -3    | 1     |       |       |       |
| $I = 5$    |       |       |       |       |       |       |       |
| Lineal     | -2    | -1    | 0     | 1     | 2     |       |       |
| Cuadrática | 2     | -1    | -2    | -1    | 2     |       |       |
| Cúbica     | -1    | 2     | 0     | -2    | 1     |       |       |
| Grado 4    | 1     | -4    | 6     | -4    | 1     |       |       |
| $I = 6$    |       |       |       |       |       |       |       |
| Lineal     | -5    | -3    | -1    | 1     | 3     | 5     |       |
| Cuadrática | 5     | -1    | -4    | -4    | -1    | 5     |       |
| Cúbica     | -5    | 7     | 4     | -4    | -7    | 5     |       |
| Grado 4    | 1     | -3    | 2     | 2     | -3    | 1     |       |
| Grado 5    | -1    | 5     | -10   | 10    | -5    | 1     |       |
| $I = 7$    |       |       |       |       |       |       |       |
| Lineal     | -3    | -2    | -1    | 0     | 1     | 2     | 3     |
| Cuadrática | 5     | 0     | -3    | -4    | -3    | 0     | 5     |
| Cúbica     | -1    | 1     | 1     | 0     | -1    | -1    | 1     |
| Grado 4    | 3     | -7    | 1     | 6     | 1     | -7    | 3     |
| Grado 5    | -1    | 4     | -5    | 0     | 5     | -4    | 1     |
| Grado 6    | 1     | -6    | 15    | -20   | 15    | -6    | 1     |

Tabla 2.2: Coeficientes de contrastes ortogonales de tendencias para  $I$  niveles igualmente espaciados de un factor tratamiento con tamaños muestrales iguales.

### 2.7.2 Contrastes Individuales y Medias de Tratamientos

#### Intervalo de Confianza para un Contraste Individual

En general, si  $\sum_{i=1}^I b_i \hat{\tau}_i$  es el estimador mínimo cuadrático de un contraste individual  $\sum_{i=1}^I b_i \tau_i$ , con  $\sum_{i=1}^I b_i = 0$ , entonces, bajo las hipótesis de normalidad, independencia y homocedasticidad del modelo, se prueba que un intervalo de confianza, al nivel  $1 - \alpha$ ,  $0 < \alpha < 1$ , para dicho contraste tiene la forma:

$$\boxed{\sum_{i=1}^I b_i \tau_i \in \left( \sum_{i=1}^I b_i \hat{\tau}_i \mp t_{g.l., \alpha/2} \sqrt{\widehat{Var} \left( \sum_{i=1}^I b_i \hat{\tau}_i \right)} \right)} \quad (2.17)$$

donde *g.l.* representa los grados de libertad con que se ha estimado la varianza del error.

En el caso particular del modelo de un diseño completamente aleatorizado, la expresión (2.17) resulta:

$$\boxed{\sum_{i=1}^I b_i \tau_i \in \left( \sum_{i=1}^I b_i \bar{Y}_i \mp t_{n-I, \alpha/2} \sqrt{scmR \sum_{i=1}^I \frac{b_i^2}{n_i}} \right)} \quad (2.18)$$

donde  $\sigma^2$  se ha estimado por medio de la varianza residual, *scmR*, con  $n - I$  grados de libertad.

### Intervalo de Confianza para un Media Individual

Aunque es muy poco probable que exista interés en construir un intervalo de confianza para la respuesta media en un único tratamiento, su cálculo es elemental.

En efecto, dado que  $\mu_i = \mu + \tau_i$  es estimable y su estimador mínimo-cuadrático es  $\bar{Y}_i$ , el intervalo de confianza es:

$$\boxed{\mu_i = \mu + \tau_i \in \left( \bar{Y}_i \mp t_{n-I, \alpha/2} \sqrt{\frac{scmR}{n_i}} \right)} \quad (2.19)$$

### Test de Hipótesis para un Contraste Individual o la Media de un Tratamiento

Se desea resolver el contraste de hipótesis:

$$\begin{cases} H_0 : \sum_{i=1}^I b_i \tau_i = 0 \\ H_1 : \sum_{i=1}^I b_i \tau_i \neq 0 \end{cases}$$

La regla de decisión a un nivel de significación  $\alpha$  es:

$$\text{Rechazar } H_0 \text{ si } \left| \frac{\sum_{i=1}^I b_i \bar{Y}_i}{\sqrt{scmR \sum_{i=1}^I \frac{b_i^2}{n_i}}} \right| > t_{n-I, \alpha/2} \quad (2.20)$$

O, equivalentemente, elevando al cuadrado:

$$\text{Rechazar } H_0 \text{ si } \frac{\left(\sum_{i=1}^I b_i \bar{Y}_i\right)^2}{scmR \sum_{i=1}^I \frac{b_i^2}{n_i}} > t_{n-I, \alpha/2}^2 = F_{1, n-I, \alpha} \quad (2.21)$$

El valor

$$scC = \frac{\left(\sum_{i=1}^I b_i \bar{Y}_i\right)^2}{\sum_{i=1}^I \frac{b_i^2}{n_i}} \quad (2.22)$$

se denomina **suma de cuadrados del contraste** y las reglas de decisión equivalentes (2.20) y (2.21) pueden simplificarse escribiendo:

$$\text{Rechazar } H_0 \text{ si } \frac{scC}{scmR} > F_{1, n-I, \alpha} \quad (2.23)$$

Para tratar con modelos futuros, la forma general de  $scC/scmR$  es:

$$\frac{scC}{scmR} = \frac{\left(\sum_{i=1}^I b_i \hat{\tau}_i\right)^2}{\widehat{Var}\left(\sum_{i=1}^I b_i \hat{\tau}_i\right)} \quad (2.24)$$

### Análisis Basado en Contrastes Ortogonales

Dos contrastes  $\sum_{i=1}^I c_i \tau_i$  y  $\sum_{j=1}^I d_j \tau_j$  se dicen ortogonales si y solo si sus estimadores mínimo cuadráticos son incorrelados.

$$Cov\left(\sum_{i=1}^I c_i \bar{Y}_i, \sum_{j=1}^I d_j \bar{Y}_j\right) = \sum_{i=1}^I c_i d_i Var(\bar{Y}_i) + \sum_{i=1}^I \sum_{j \neq i}^I c_i d_j Cov(\bar{Y}_i, \bar{Y}_j) = \sigma^2 \sum_{i=1}^I \frac{c_i d_i}{n_i}$$

$$\sum_{i=1}^I c_i \mu_i \text{ y } \sum_{j=1}^I d_j \mu_j \text{ son ortogonales } \Leftrightarrow \sum_{i=1}^I \frac{c_i d_i}{n_i} = 0 \quad (2.25)$$

Para los modelos aquí considerados, todos los estimadores de los contrastes están normalmente distribuidos, por tanto la ortogonalidad de dos contrastes implica que sus estimadores mínimo cuadráticos son independientes.

Se prueba que, dados  $I$  tratamientos, existe un máximo de  $I - 1$  contrastes mutuamente ortogonales (que se llama **conjunto completo de contrastes ortogonales**). Entonces, por la independencia de los respectivos estimadores, la suma de cuadrados de un contraste puede descomponerse en la suma de las sumas de cuadrados de cada uno de los contrastes de un conjunto completo.

## 2.8 Métodos de Comparaciones Múltiples

En ocasiones se desea analizar un número de comparaciones muy grande, de modo que la probabilidad de que *alguna* comparación individual resulte significativa erróneamente es alarmantemente grande.

Supóngase que interesa resolver todas las pruebas de hipótesis siguientes:

$$\text{Prueba } ij : \quad \begin{cases} H_0^{ij} : \tau_i = \tau_j \\ H_1^{ij} : \tau_i \neq \tau_j \end{cases} \quad \forall i, j = 1, \dots, I, \quad i \neq j$$

Existen  $m = \binom{I}{2}$  pruebas (si, por ejemplo,  $I = 6$  entonces  $m = 15$ ). Si se resuelven una por una a un nivel  $\alpha$  y se denomina  $A_{ij}$  al suceso:

$$A_{ij} = \left\{ \text{aceptar } H_0^{ij} \text{ siendo } \tau_i = \tau_j \right\}$$

Entonces:

$$P(A_{ij}) = P\left( |\bar{Y}_i - \bar{Y}_j| \leq t_{n-I, \alpha/2} \sqrt{scmR} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right) = 1 - \alpha$$

Sin embargo, lo que nos debe de preocupar, y mucho, es la probabilidad del suceso  $A$ , siendo:

$$A = \left\{ \text{rechazar erróneamente alguna } H_0^{ij} \right\} = \cup_{ij}^m \bar{A}_{ij},$$

¿Cuál es la probabilidad de  $A$ ? En el supuesto de que los sucesos  $A_{ij}$  fuesen independientes (obviamente no lo son):

$$P(A) = P\left(\cup_{ij}^m \bar{A}_{ij}\right) = P\left(\overline{\cap_{ij}^m A_{ij}}\right) = 1 - (1 - \alpha)^m$$

Si  $\alpha = 0'05$  y  $m = 15$ , entonces  $P(A) = 1 - 0'95^{15} = 1 - 0'46 = 0'54$ . O sea, más de la mitad de las veces se concluiría erróneamente que algún par de tratamientos son significativamente distintos.

Hay distintos métodos para abordar el problema de la resolución de pruebas de hipótesis *simultáneas* (es decir, garantizando para todas ellas un nivel de significación predeterminado). Unos han sido desarrollados con carácter general y otros orientados a problemas concretos como puede ser la comparación de distintos tratamientos con un tratamiento control.

El método adecuado dependerá, en general, de qué pruebas de hipótesis particulares interesa resolver y del número total de ellas.

Se acostumbra a distinguir entre:

*Pruebas preplanificadas:* Antes de realizar el experimento se establece como objetivo primordial comparar un grupo determinado de medias y de contrastes.

*Pruebas no preplanificadas:* Una vez tomados los datos los resultados no concuerdan con lo esperado a priori y pruebas de hipótesis que inicialmente no eran de excesivo interés son ahora relevantes para las conclusiones.

A continuación se exponen varios procedimientos distintos pero todos ellos con un denominador común:

Dado un conjunto de  $m$  contrastes  $\left\{ \sum_{i=1}^I b_i^j \tau_i \right\}_{j=1}^m$ , se rechaza la hipótesis  $H_0^j : \sum_{i=1}^I b_i^j \tau_i = 0$ , a un nivel de significación como máximo  $\alpha$ , siempre que:

$$\left| \sum_{i=1}^I b_i^j \hat{\tau}_i \right| \geq \omega \times \sqrt{\widehat{Var} \left( \sum_{i=1}^I b_i^j \hat{\tau}_i \right)} \quad (2.26)$$

El valor de  $\omega$  es el valor crítico de la regla de decisión y dependerá de:

- cada método,
- del número de pruebas de hipótesis simultáneas,  $m$ ,
- y del número de grados de libertad del estimador de la varianza del contraste.

Para cada contraste, al valor:  $\omega \times \sqrt{\widehat{Var} \left( \sum_{i=1}^I b_i^j \hat{\tau}_i \right)}$ , que es la frontera entre el rechazo y la aceptación de la prueba de hipótesis relativa a ese contraste a un nivel de significación global  $\alpha$ , se le denomina a menudo **mínima diferencia significativa (mds) al nivel de significación global  $\alpha$** .

### 2.8.1 Método de Bonferroni

Consiste en determinar previamente a qué nivel de significación  $\alpha^*$  es necesario realizar cada una de las  $m$  pruebas de hipótesis para garantizar un nivel de significación global prefijado  $\alpha$  para el conjunto de las  $m$  pruebas.

Supóngase que, como antes, se tienen  $I$  niveles y  $m$  pruebas de hipótesis individuales. Sean los sucesos:

$A_k$ : aceptar la hipótesis nula del contraste  $k$ -ésimo cuando ésta es cierta.

$A$ : rechazar erróneamente la hipótesis nula de uno o más contrastes.

¿A qué  $\alpha^*$  habría que realizar cada prueba de hipótesis individual para garantizar que  $P(A)$  no excede  $\alpha$ ?

$$P(A) = P\left(\cup_{k=1}^m \bar{A}_k\right) \leq \sum_{k=1}^m P(\bar{A}_k) = m\alpha^* = \alpha \implies \boxed{\alpha^* = \frac{\alpha}{m}} \quad (2.27)$$

Por tanto, para el modelo matemático de un diseño completamente aleatorizado, el método de Bonferroni consiste en resolver cada prueba de hipótesis individual conforme al siguiente criterio:

Dado un conjunto de  $m$  contrastes  $\left\{\sum_{i=1}^I b_i^j \tau_i\right\}_{j=1}^m$ , se rechaza la hipótesis  $H_0^j : \sum_{i=1}^I b_i^j \tau_i = 0$ , a un nivel de significación global no superior a  $\alpha$ , siempre que

$$\left|\sum_{i=1}^I b_i^j \bar{y}_i\right| \geq \omega_B \times \sqrt{scmR \sum_{i=1}^I \frac{(b_i^j)^2}{n_i}} \quad (2.28)$$

donde el valor crítico es:

$$\boxed{\omega_B = t_{n-I, \alpha/(2m)}}$$

Entre las ventajas y desventajas de este procedimiento cabe enumerar las siguientes:

- Si  $m$  es muy grande,  $\alpha/(2m)$  será tan pequeño que  $t_{n-I, \alpha/(2m)}$  no vendrá en las tablas. Aproximar por:

$$t_{n-I, \alpha/(2m)} = z_{\alpha/(2m)} \left(1 - \frac{z_{\alpha/(2m)} + 1}{4(n-I)}\right)^{-1}$$

- Usar sólo para pruebas de hipótesis preplanificadas.

- Es excesivamente conservador y sólo resulta más potente que otros procedimientos cuando  $m$  es muy pequeño.
- Es válido para cualquier tipo de diseño.

### 2.8.2 Método de Scheffé

Del apartado anterior se deduce que los dos serios inconvenientes del método de comparaciones de Bonferroni son: primero, que las  $m$  pruebas de hipótesis a resolver debieran haber sido preplanificadas y, segundo, que si  $m$  es muy grande la mínima diferencia significativa al nivel global  $\alpha$  para cada prueba es excesivamente amplia.

Por el contrario, el método de Scheffé proporciona una mínima diferencia significativa que no depende del número de pruebas  $m$  a realizar.

La llave teórica del procedimiento radica en que cualquier contraste  $\sum_{i=1}^I b_i \tau_i$  puede escribirse como una combinación lineal del conjunto de  $I-1$  contrastes tratamiento frente a control:  $\tau_1 - \tau_I$ ,  $\tau_2 - \tau_I$ ,  $\dots$ ,  $\tau_{I-1} - \tau_I$ . Se puede probar entonces (los detalles matemáticos no son elementales) que las mínimas diferencias significativas, al nivel global  $\alpha$ , para las pruebas de hipótesis asociadas a los  $I-1$  contrastes anteriores, determinan también las m.d.s., al mismo nivel de significación global, para cualquier contraste  $\sum_{i=1}^I b_i \tau_i$  y para cualquier número de pruebas  $m$ .

El valor crítico de Scheffé es:

$$\omega_S = \sqrt{(I-1) F_{I-1, n-I, \alpha}}$$

de modo que, para cualquier contraste individual  $\sum_{i=1}^I b_i \tau_i$  se rechaza la hipótesis

$H_0 : \sum_{i=1}^I b_i \tau_i = 0$ , a un nivel de significación global no superior a  $\alpha$ , siempre que

$$\left| \sum_{i=1}^I b_i \bar{y}_i \right| \geq \omega_S \times \sqrt{scmR \sum_{i=1}^I \frac{b_i^2}{n_i}} \quad (2.29)$$

Sobre el método de Scheffé conviene conocer:

- Sólo depende de  $I$  y de  $n$ , pero no de  $m$ .
- Es especialmente adecuado cuando se precisen comparar otros contrastes además de las comparaciones a pares.

- Puede usarse con contrastes no preplanificados.
- Si  $m$  es muy grande, resulta más potente (y por ello más recomendable) que el método de Bonferroni.
- Es válido para cualquier tipo de diseño.
- El  $F$ -test del Anova resulta significativo al nivel  $\alpha$  si al menos una de las infinitas pruebas de hipótesis simultáneas de Scheffé lo es.

### 2.8.3 Método de Tukey para Todas las Comparaciones a Pares

Este método ha sido propuesto por Tukey en 1953 y es especialmente adecuado cuando se desean analizar tan sólo los contrastes de comparaciones a pares o *pairwise*.

Supóngase en primer lugar que  $n_i = r \forall i$ . Si se ordenan las medias muestrales en sentido creciente se demuestra que, bajo las hipótesis de normalidad e independencia, el estadístico:

$$Q = \frac{\max_{1 \leq i \leq I} \bar{Y}_i - \min_{1 \leq i \leq I} \bar{Y}_i}{\sqrt{\frac{SCMR}{r}}}$$

tiene distribución conocida, que se denomina, *distribución del rango estudentizado* y que depende de los grados de libertad del estimador de la varianza y de  $I$ .

Sea  $q_{I,n-I,\alpha}$  el correspondiente percentil  $100(1 - \alpha)$ . Entonces:

$$P \left( \left| \max_{1 \leq i \leq I} \bar{Y}_i - \min_{1 \leq i \leq I} \bar{Y}_i \right| \leq q_{I,n-I,\alpha} \sqrt{\frac{SCMR}{r}} \right) = 1 - \alpha \Rightarrow$$

$$P \left( \left| \bar{Y}_i - \bar{Y}_j \right| \leq q_{I,n-I,\alpha} \sqrt{\frac{SCMR}{r}}; \forall i, j \right) = 1 - \alpha$$

Por tanto, reemplazando  $\bar{Y}_i$  y  $SCMR$  por sus respectivos valores observados  $\bar{y}_i$  y  $scmR$ , se concluye:

Rechazar cada hipótesis nula del tipo:  $H_0 : \tau_i - \tau_j = 0$ , de comparación de un par de medias a un nivel de significación exactamente  $\alpha$ , siempre que

$$\left| \bar{y}_i - \bar{y}_j \right| \geq \omega_T \times \sqrt{scmR \left( \frac{1}{r} + \frac{1}{r} \right)} \quad (2.30)$$

donde, el valor crítico de Tukey es:

$$\boxed{\omega_T = q_{I,n-I,\alpha} / \sqrt{2}}$$

Sobre el método de Tukey conviene saber que:

- El nivel de significación global es exactamente  $\alpha$  cuando  $n_i = r, \forall i = 1, \dots, I$ .
- En otro caso es sólo aproximado y el nivel de significación es *como máximo*  $\alpha$ .
- Está especialmente recomendado para contrastes de comparaciones a pares balanceados, donde es más potente que los métodos de Bonferroni y Scheffé.
- Si  $I = 5, n = 35$  y  $\alpha = 0'05, \omega_B = t_{30,0'025/10} = 3'02, \omega_S = \sqrt{4F_{4,30,0'05}} = 3'28$  y  $\omega_T = q_{5,30,0'05}/\sqrt{2} = 2'91$ .

### 2.8.4 Método de Dunnett para Comparaciones entre Tratamientos y Control

Para analizar las comparaciones preplanificadas entre cada tratamiento y un tratamiento control (por ejemplo  $\tau_I$ ):  $\tau_i - \tau_I, i = 1, \dots, I-1$ , Dunnett desarrolló un método de comparaciones múltiples que proporciona regiones de rechazo simultáneas a un nivel global exacto  $\alpha$  más extensas que las proporcionadas por los métodos de Bonferroni, Scheffé y Tukey. Ahora bien, este método no es apropiado para analizar cualquier otro tipo de contraste.

El método se basa en que la distribución conjunta de los estimadores  $\bar{Y}_i - \bar{Y}_I, i = 1, \dots, I-1$  es un caso especial de la distribución  $t$  multivariante y depende de la correlación entre  $\bar{Y}_i - \bar{Y}_I$  y  $\bar{Y}_j - \bar{Y}_I, i \neq j$ .

En un diseño completamente aleatorizado con  $n_1 = \dots = n_{I-1} = r$  y  $n_I = c$ , el coeficiente de correlación es

$$\rho = \frac{r}{r+c}$$

A menudo se toma  $c = r$  (diseño balanceado), en cuyo caso  $\rho = 0'5$ . Sin embargo, para optimizar la potencia de las pruebas de hipótesis simultáneas con el procedimiento de Dunnett, se debe tomar  $c/r$  lo más próximo posible a  $\sqrt{I-1}$ . Aquí se proporciona la tabla de la distribución  $t$  multivariante con  $\rho = 0'5$  y es por ello que nos centramos en el caso  $c = r$ . Si  $n_i = r \forall i$ :

Rechazar cada hipótesis nula del tipo:  $H_0 : \tau_i - \tau_I = 0$ , de comparación de un tratamiento con un control (tratamiento  $I$ ), a un nivel de significación global exactamente  $\alpha$ , siempre que

1. Si la alternativa es que los tratamientos son superiores en la respuesta media al control (contrastos de hipótesis unilaterales a la derecha):

$$\bar{y}_i - \bar{y}_I \geq -\omega_{D1} \times \sqrt{\frac{scmR}{r}} \quad (2.31)$$

2. Si la alternativa es que los tratamientos son inferiores en la respuesta media al control (contrastes de hipótesis unilaterales a la izquierda):

$$\bar{y}_i - \bar{y}_I \leq \omega_{D1} \times \sqrt{\frac{scmR}{r}} \quad (2.32)$$

3. Y, finalmente, si la alternativa es que los tratamientos son distintos en la respuesta media al control (contrastes de hipótesis bilaterales):

$$|\bar{y}_i - \bar{y}_I| \geq \omega_{D2} \times \sqrt{\frac{scmR}{r}} \quad (2.33)$$

donde los valores críticos de Dunnett son:

$$\omega_{D1} = t_{I-1, n-I, \alpha}^{0'5} \quad \text{y} \quad \omega_{D2} = |t_{I-1, n-I, \alpha}^{0'5}|$$

con  $t_{I-1, n-I, \alpha}^{0'5}$  el  $\alpha$ -cuantil del máximo de la distribución  $t$  multivariante y  $|t_{I-1, n-I, \alpha}^{0'5}|$  el  $\alpha$ -cuantil del máximo del valor absoluto de la distribución  $t$  multivariante (ver tablas 3 y 4 en Apéndice).

### 2.8.5 Otros Métodos

Los cuatro métodos de comparaciones múltiples expuestos en las secciones anteriores permiten al experimentador controlar el nivel de significación global. Existen otros procedimientos, algunos incluso más potentes (es decir, más sensibles para detectar un contraste no nulo), pero que no mantienen una tasa de error global. Entre otros se pueden citar:

LSD (Least Significant Difference) de Fisher

Test de Rango Múltiple de Duncan

Test de Newman y Keuls

Test de Hsu (éste sí mantiene el nivel de significación global y es muy útil para comparar cada tratamiento con el *mejor*).

## 2.9 Tamaños Muestrales

En la sección 2.6 se expuso un método para determinar el tamaño muestral apropiado que permita obtener una potencia preespecificada al testar  $H_0 : \tau_1 = \tau_2 = \dots = \tau_I$  mediante el cociente  $F$  del Anova.

Los contrastes de comparaciones múltiples proporcionan un camino alternativo. Como en la sección 2.6, en base a una estimación previa de la varianza del error, se aproximan los  $n_i$  mediante secuencias de *ensayo y error*. Veamos un ejemplo.

**Ejemplo 2.2** Se desea comparar el efecto de 5 formas distintas de regadío sobre el crecimiento de semillas de alubias. La respuesta es la longitud de un nuevo brote 48 horas después de regar. Un experimento piloto ha mostrado que la varianza del error no debiera superar los  $10 \text{ mm}^2$  y, por otro lado, interesa realizar las comparaciones pareadas de tratamientos con un nivel de significación global del 5%, de modo que diferencias superiores o iguales a  $3 \text{ mm}$  se detecten como significativas.

Utilizando Tukey:  $\omega_T \times \sqrt{10 \left( \frac{1}{r} + \frac{1}{r} \right)} \leq 3$ , siendo  $\omega_T = q_{5,5r-5,0'05}/\sqrt{2}$ , o, equivalentemente,

$$q_{5,5r-5,0'05}^2 \leq 0'9r$$

Empezando en  $r = 10$ , vamos corrigiendo la decisión como sigue:

| $r$ | $5r - 5$ | $q_{5,5r-5,0'05}^2$ | $0'9r$ | Acción        |
|-----|----------|---------------------|--------|---------------|
| 10  | 45       | $4'03^2 = 16'24$    | 9'00   | Aumentar $r$  |
| 20  | 95       | $3'95^2 = 15'60$    | 18'00  | Disminuir $r$ |
| 15  | 70       | $3'97^2 = 15'76$    | 13'50  | Aumentar $r$  |
| 18  | 85       | $3'96^2 = 15'68$    | 16'20  | Disminuir $r$ |
| 17  | 80       | $3'97^2 = 15'68$    | 15'30  |               |

A menos que se desee más información de unos tratamientos que de otros, o que existan varianzas muy desiguales dentro de cada tratamiento, es recomendable elegir tamaños muestrales iguales. Tal criterio presenta, entre otras, las dos ventajas siguientes:

1. Se obtienen intervalos de confianza de igual longitud para las comparaciones por pares de tratamientos, lo que facilita su comparación y análisis.
2. Las comparaciones múltiples y el test  $F$  del Anova son menos sensibles a una suposición incorrecta de normalidad de los errores.

## 2.10 Comprobación de la Idoneidad del Modelo

Las hipótesis establecidas a priori sobre los errores del modelo:  $\varepsilon_{it} = Y_{it} - (\mu + \tau_i)$  pueden y deben chequearse (tómese como norma rutinaria) a partir de los **residuos**.

Según la expresión (2.7) el  $it$ -ésimo residuo,  $e_{it}$ , asociado al modelo de un diseño completamente aleatorizado viene dado por

$$e_{it} = y_{it} - \bar{y}_i.$$

Para facilitar la identificación de observaciones atípicas (*outliers*), suele ser preferible trabajar con los **residuos estandarizados**:

$$e_{it}^* = \frac{e_{it}}{\sqrt{scmR}} \quad (2.34)$$

Si las suposiciones sobre el modelo son correctas:  $\frac{\varepsilon_{it}}{\sigma}$  es una muestra aleatoria simple de una distribución  $N(0, 1)$  y, por tanto, cabe esperar un comportamiento similar para  $e_{it}^*$ .

Tomando como base los residuos, la estrategia a seguir consiste en chequear las siguientes hipótesis:

1. **Chequear la bondad del ajuste del modelo.** ¿Se describen adecuadamente las respuestas medias de los tratamientos por medio de  $E(Y_{it}) = \mu + \tau_i$ ,  $i = 1, \dots, I$ ?
2. **Chequear la existencia de observaciones atípicas.**
3. **Chequear la independencia de las observaciones.**
4. **Chequear la homocedasticidad del error.** ¿Tiene la variable error aproximadamente la misma varianza en cada tratamiento?
5. **Chequear la normalidad.**

### 2.10.1 Bondad del Ajuste del Modelo

En general se debe observar la nube de residuos estandarizados en un diagrama de dispersión con respecto a los niveles de cada factor incluido en el modelo (factores tratamiento, factores bloque, covariables, ...).

En el modelo actual basta observar la nube de puntos  $\{i, e_{it}^*\}$ ,  $t = 1, \dots, n_i$ ,  $i = 1, \dots, I$ .

Por ejemplo, la Figura 2.1 muestra la nube de puntos correspondiente a los residuos estandarizados que se obtienen al estimar un modelo como el estudiado. En particular, esta nube se corresponde con un experimento sobre los niveles de hemoglobina en sangre de una especie de truchas según el suministro de distintas cantidades de sulfamerazina en su dieta alimenticia (ver Problema 5 en la sección 2.12 de Problemas Propuestos de este capítulo para una información más completa acerca de este experimento).

Cuando una nube de este tipo muestre un patrón no aleatorio alrededor del 0 (residuos excesivamente positivos para algunos niveles y excesivamente negativos para otros) delatará **falta de ajuste**. El modelo supuesto no será válido y deberá modificarse.

En ocasiones, la observación de los residuos frente a los niveles de factores que se omitieron en el experimento permite modificar de forma adecuada el modelo.

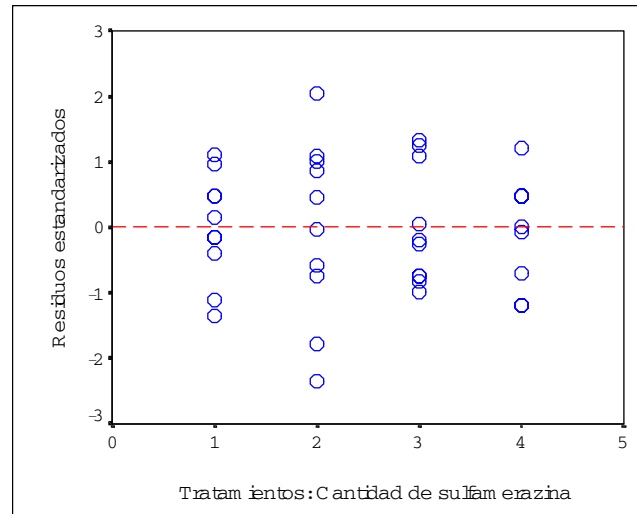


Figura 2.1: Diagrama de dispersión para el experimento de las truchas.

### 2.10.2 Existencia de Observaciones Atípicas

Una observación atípica es un registro mayor o menor de lo esperado que detectaremos por su correspondiente residuo que, lógicamente, resultará ser un valor “inusual”. En efecto, los residuos estandarizados son una muestra aleatoria de una distribución normal estándar y, por tanto, aproximadamente un 68% de ellos debe ubicarse entre  $-1$  y  $1$ , alrededor del 95% entre  $-2$  y  $2$  y prácticamente todos entre  $-3$  y  $3$ . Un residuo estandarizado que diste más de 3 o 4 unidades del 0 es potencialmente una observación atípica.

De acuerdo con el razonamiento anterior, un diagrama de dispersión de los puntos  $\{i, e_{it}^*\}_{i,t}$ , como el utilizado para chequear la bondad del ajuste del modelo, es una posibilidad gráfica válida y habitual para chequear asimismo la existencia de observaciones atípicas. Otra alternativa es observar conjuntamente los diagramas de caja de los residuos por niveles del factor tratamiento como se observa en la Figura 2.2.

En ocasiones las observaciones atípicas son fruto de errores concretos en el proceso de registro de los datos. Otras veces en cambio delatan ausencia de normalidad, heterocedasticidad o una incorrecta especificación del modelo.

Una vez se ha identificado la presencia de una observación atípica, se debe investigar pormenorizadamente su procedencia (si ello fuese posible) y eliminarla tan sólo si se concluye que se ha generado por errores en el muestreo. En otro caso, es conveniente repetir el análisis estadístico sin la observación atípica y examinar las nuevas conclusiones. Si dichas conclusiones son, en esencia, semejantes a las obtenidas con ella, la observación puede mantenerse. Si, por el contrario, las conclusiones se modifican drásticamente, la observación atípica se denomina

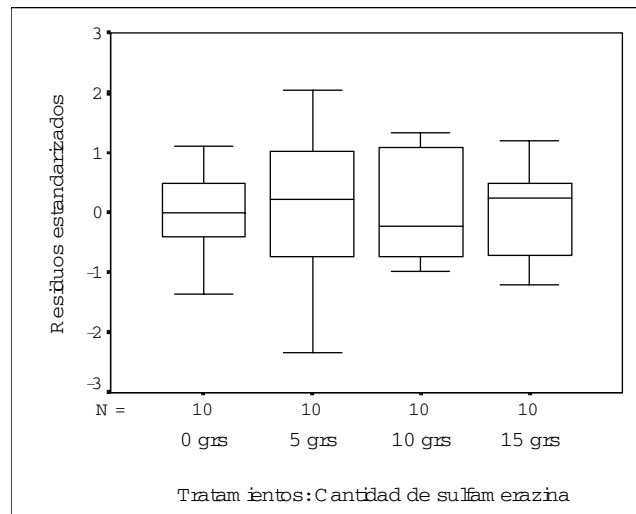


Figura 2.2: Diagramas de caja de los residuos para el experimento de las truchas.

**influyente.** En este punto, el experimentador debe enjuiciar si es posible su aparición por un error experimental (eliminarlo) o si tal observación podría volver a aparecer (buscar modelos más complejos).

**Ejemplo 2.3** La Figura 2.3 muestra la nube de residuos estandarizados que se obtuvo inicialmente con el experimento de las pilas del final del capítulo 1.

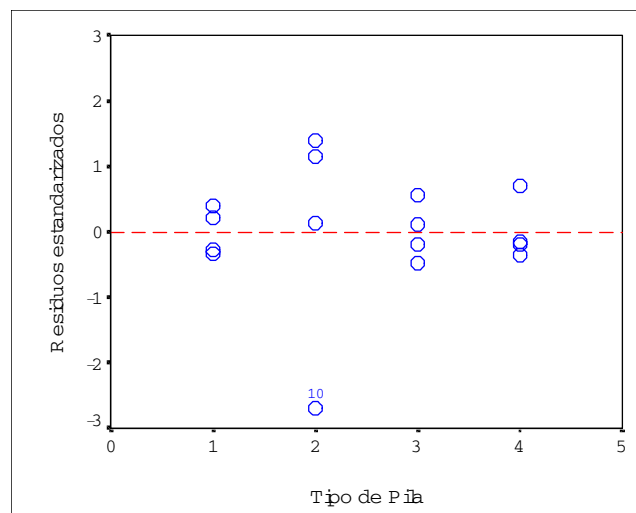


Figura 2.3: Diagrama de dispersión para el experimento de las pilas.

En la Figura 2.3 se perciben dos relativas anomalías. Por un lado, la décima observación (correspondiente a una de las pilas del tipo 2) genera un residuo de valor  $-2'77$ . Además, la totalidad de residuos procedentes de los otros tres tipos de pilas se encuentran entre  $-1$  y  $1$ , lo que constituye bastante más que el 68% esperado.

Un análisis del origen de la décima observación reveló un error en el registro. Se había introducido en el ordenador una duración de 473 minutos cuando en la hoja de anotaciones figuraba una duración de 773 minutos. Subsano el error se repitió el ajuste del modelo y los nuevos residuos estandarizados se muestran en la Figura 2.4. Ambas anomalías desaparecen ahora. Ya no se detectan observaciones atípicas y, más aún, la varianza residual es más pequeña, lo que conduce a residuos algo mayores.

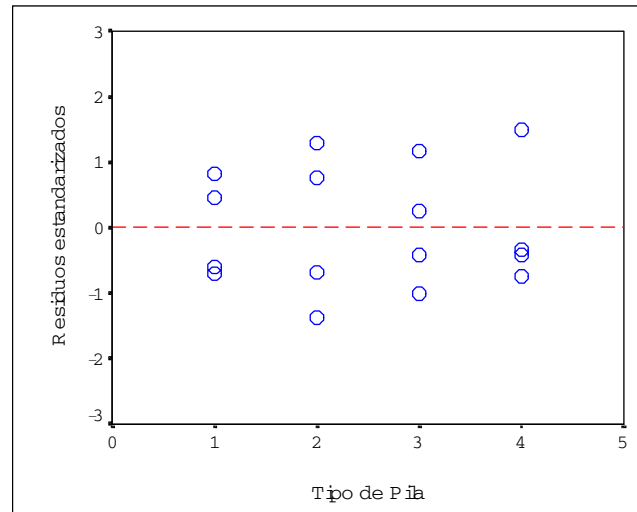


Figura 2.4: Diagrama de dispersión corregido para el experimento de las pilas.

### 2.10.3 Independencia de los Errores

La causa más común de dependencia entre los errores es la similitud de unidades experimentales “próximas” en tiempo o espacio.

Gráficos especialmente útiles para detectar este problema son:

- Un diagrama de dispersión de los residuos frente al orden (temporal o espacial) del registro de las observaciones. En el caso de independencia la nube de puntos responde a puntos marcados al azar en torno al cero, sin discernir patrón o tendencia determinista alguna. Así ocurre en la Figura 2.5 para el ejemplo del experimento de las pilas.

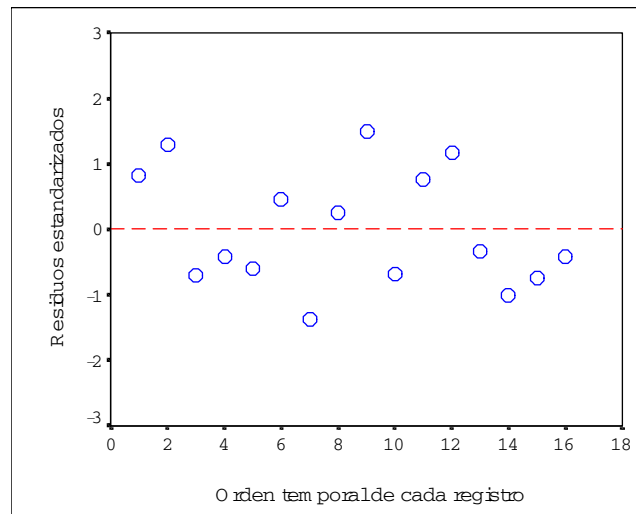


Figura 2.5: Diagrama de dispersión de los residuos frente al tiempo para el experimento de las pilas.

- Un diagrama de dispersión de los puntos ( $i$ -ésimo residuo,  $(i + 1)$ -ésimo residuo),  $i = 1, \dots, n - 1$ . Si los datos son dependientes, esta nube tendería a ubicarse en torno al eje  $x = y$ . Si son independientes nos volveremos a encontrar con una nube totalmente aleatoria. Véase la Figura 2.6 para el experimento de las pilas.

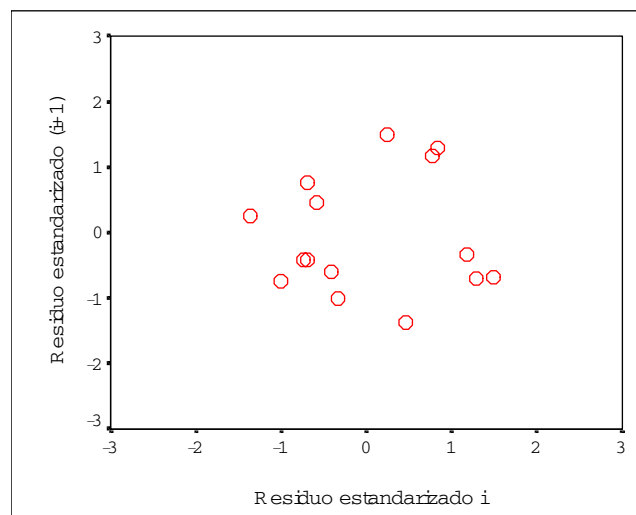


Figura 2.6: Nube (residuo  $i$ , residuo  $i + 1$ ) para el experimento de las pilas.

- Un diagrama de autocorrelaciones o *autocorrelograma* de los residuos (ver Figura 2.7).

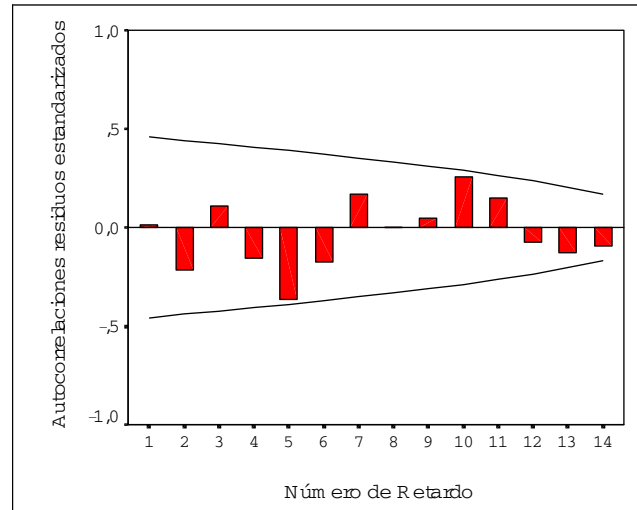


Figura 2.7: Autocorrelograma de los residuos para el experimento de las pilas.

Desde un punto de vista analítico se dispone de métodos de distribución libre basados en rachas o pruebas de hipótesis basadas en la hipótesis de normalidad como Ljung y Box o Durbin y Watson.

### Prueba de aleatoriedad basada en el número total de rachas

Para introducir el concepto de “racha” considérese una muestra de una variable con dos posibles resultados. Por ejemplo NSSSNNSNNSNS, con N: “no verifica una determinada característica” y S: “sí la verifica”.

Se denomina **racha** a una sucesión de valores consecutivos repetidos que esté flanqueada por valores adyacentes distintos. En el ejemplo anterior hay 8 rachas: N, SSS, NN, S, NN, S, N, S.

Independientemente de lo probable que sea observar los valores concretos de la variable, es obvio que el número total de rachas constituye una medida de lo aleatoriamente que están repartidos los valores en cuestión a lo largo de la muestra observada:

- Demasiadas rachas implican excesiva alternancia de valores (dependencia negativa).
- Pocas rachas indican largas sucesiones de valores contiguos repetidos (dependencia positiva).

La prueba de aleatoriedad se basa en el siguiente razonamiento:

Sea una muestra de tamaño  $n$  correspondiente a una variable con dos posibles resultados, de modo que  $n_1$  son de un tipo y  $n_2$  del otro ( $n_1 + n_2 = n$ ).

Sea  $R$  el número total de rachas observadas en la muestra. La distribución exacta de  $R$  condicionada a los valores  $n_1$  y  $n_2$  es conocida (se encuentra tabulada) y, en particular:

$$E(R) = 1 + \frac{2n_1n_2}{n} \quad \text{y} \quad \text{Var}(R) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}$$

Por ejemplo, si  $n = 3$ ,  $n_1 = 2$  y  $n_2 = 1$  (o sea dos S y una N), sea cuál sea la probabilidad de S, la distribución de  $R$  condicionada a la presencia de dos S y una N será:

$$\begin{cases} SSN (\Rightarrow R = 2) \text{ ocurre con } p = 1/3 \\ SNS (\Rightarrow R = 3) \text{ ocurre con } p = 1/3 \\ NSS (\Rightarrow R = 2) \text{ ocurre con } p = 1/3 \end{cases} \Rightarrow R = \begin{cases} 2 \text{ con probabilidad } 2/3 \\ 3 \text{ con probabilidad } 1/3 \end{cases}$$

Cuando  $n$  tiende a infinito ( $n > 40$ ) la distribución de  $R$  tiende a la de una normal con la media y varianza especificadas. En caso de que los valores  $n_1$  y  $n_2$  sean pequeños debe acudir a unas tablas. Se rechazará la hipótesis de independencia para un número de rachas significativamente grande o significativamente pequeño.

Esta prueba puede adaptarse a distribuciones continuas sin más que codificar las observaciones con los valores  $+$  o  $-$  según que el dato en cuestión quede por arriba o por debajo de la mediana muestral.

La ausencia de aleatoriedad entre las observaciones es muy difícil de corregir y es especialmente grave en el sentido de invalidar por completo las conclusiones del análisis estadístico (pueden incluso conducir a conclusiones erróneas). En particular, todas las expresiones utilizadas para las varianzas son incorrectas y, por tanto, los intervalos de confianza y las pruebas de hipótesis deducidos a partir de ellas, tendrán una confianza o una potencia distinta a la supuesta.

Para ejemplificar este efecto supóngase una variable  $X$  tal que:

$$\text{Cov}(X_i, X_t) = \begin{cases} \rho\sigma^2 & \text{si } t = i + 1 \\ 0 & \text{en otro caso} \end{cases}$$

Entonces se mantiene que  $E(\bar{X}) = \mu$ , pero

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \sum E[(X_i - \mu)]^2 + \frac{2}{n^2} \sum E[(X_i - \mu)(X_t - \mu)] = \frac{\sigma^2}{n} + \frac{2}{n^2}(n-1)\rho\sigma^2$$

Suponiendo  $n$  grande y, por tanto,  $n + 1/n \approx 1$ , se observa que para un valor  $\rho = 0.45$ :

$$\text{Var}(\bar{X}) \approx 1.9 \frac{\sigma^2}{n}$$

Si se observase una tendencia muy clara en los gráficos, por ejemplo linealidad de los residuos frente al tiempo, se pueden deducir otros modelos como:

$$Y_{it} = \mu + \tau_i + \gamma X_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \text{ i.i.d } N(0, \sigma)$$

con  $X_{it}$  la covariable *tiempo* (Modelo de Análisis de la Covarianza).

**Ejemplo 2.4** (Meily Lyin, 1985) La experimentadora venía observando en distintas fiestas de cumpleaños que unos globos de una determinada marca resultaban más difíciles de inflar que otros. Ella achacaba las diferencias al color del globo y diseñó un experimento para determinar si los tiempos necesarios de inflado eran similares con globos del mismo color y diferentes para distintos colores. Eligió globos de cuatro colores (rosa, amarillo, naranja y azul) y solicitó a un asistente que los inflara hasta alcanzar un diámetro de 7 pulgadas. Medía los tiempos con un reloj y una precisión cercana a 0'1 segundos. La Figura 2.8 muestra el diagrama de dispersión de los residuos generados por el ajuste del modelo matemático asociado a un diseño completamente aleatorizado.

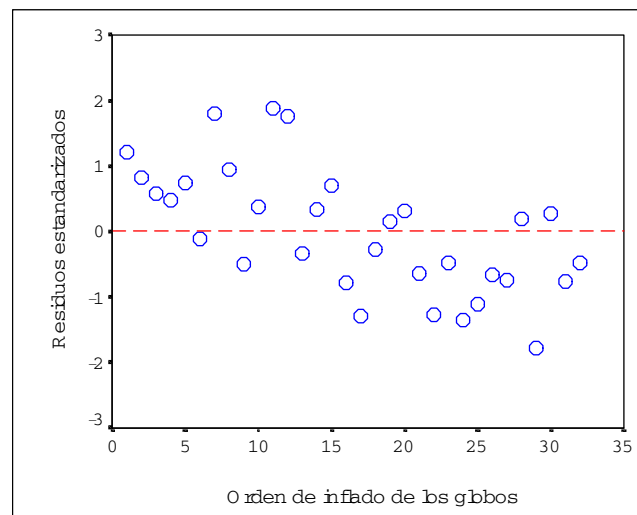


Figura 2.8: Diagrama de dispersión de los residuos para el experimento de los globos.

De la Figura 2.8 se deduce una flagrante violación de la hipótesis de independencia de las observaciones: es claro un acusado descenso del valor de los residuos a medida que progresamos en el tiempo.

### 2.10.4 Homocedasticidad de los Errores

Se ha demostrado que si el diseño es balanceado ( $n_i = r \forall i = 1, \dots, I$ ) los niveles de significación de los pruebas de hipótesis y los niveles de confianza de los intervalos apenas se ven afectados por la existencia de heterocedasticidad, a no ser que la varianza de la respuesta para algún tratamiento particular sea *considerablemente* mayor que para otros.

Estudios de simulación han mostrado que, en este supuesto de tamaños muestrales iguales, la heterocedasticidad no afecta al  $F$ -test ni a los distintos métodos de comparaciones múltiples siempre que:

$$\frac{\hat{S}_{Max}^2}{\hat{S}_{Min}^2} < 3$$

Si los tamaños muestrales son muy distintos, de modo que:

(a) tratamientos con mayor variabilidad en la respuesta han sido muestreados con los tamaños muestrales más pequeños, entonces se habrá aumentado artificialmente la probabilidad de cometer un error de tipo I en las pruebas de hipótesis y los niveles de confianza de los intervalos serán inferiores a lo que se cree;

(b) los mayores  $n_i$  coinciden con los tratamientos  $i$  de mayor varianza, entonces se tendrá el efecto contrario y las pruebas serán conservadoras.

Existen procedimientos analíticos para detectar heterocedasticidad (Hurtley, Bartlett, ...), pero todos ellos son muy conservadores y muy sensibles a la ausencia de normalidad. Una técnica muy empleada es el procedimiento de Levene que, en esencia, consiste en efectuar un análisis de la varianza sobre las diferencias en valor absoluto entre las observaciones y la mediana (u otra medida de tendencia central) manteniendo el diseño original.

Muchas veces la heterocedasticidad responde al modelo:

$$\sigma_i = k\mu_i^q \quad \text{con } k \text{ y } q \text{ constantes desconocidas,} \quad (2.35)$$

motivo por el que suele ser útil obtener un diagrama de dispersión de los residuos frente a los valores previstos o pronosticados por el modelo estimado:  $\{\bar{y}_i, e_{it}^*\}_{i,t}$  (véase la Figura 2.9).

Si se da este tipo de tendencia entonces es posible transformar los datos para obtener homocedasticidad (en otro caso puede resultar imposible encontrar transformaciones adecuadas). En efecto, se puede probar que si (2.35) es cierto, entonces cualquier transformación del tipo  $Y^* = Y^\lambda$  es tal que los datos transformados  $Y^*$  tienen varianza de la forma  $\mu^{\lambda+q-1}$ . Por tanto, si se puede determinar el valor aproximado de  $q$ , los datos transformados elevando los originales a  $\lambda = 1 - q$  tendrán varianza constante (cuando  $q = 0$  tómesese la transformación logarítmica).

Una determinación empírica de  $q$  puede realizarse en base al siguiente razonamiento. De (2.35) se deduce que  $\log(\sigma_i) = \log(k) + q \log(\mu_i)$ . Entonces una gráfica de  $\log(\sigma_i)$  contra

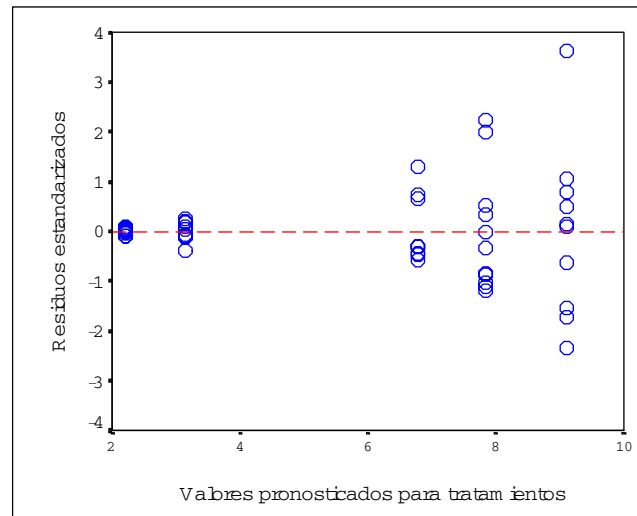


Figura 2.9: Diagrama de dispersión de residuos frente valores pronosticados.

$\log(\mu_i)$  debe resultar en una recta con pendiente igual a  $q$ . Como  $\sigma_i$  y  $\mu_i$  son desconocidas se reemplazan por estimaciones apropiadas y la pendiente de la recta sugiere el valor aproximado de  $q$  (véase el Problema 7 en la sección 2.12 de Problemas Propuestos). Otra posibilidad es utilizar la familia de transformaciones de Box y Cox.

### 2.10.5 Normalidad

La premisa de que los errores siguen una distribución normal puede chequearse también tanto gráfica como analíticamente.

Gráficos especialmente útiles son: los histogramas móviles, la estimación no paramétrica de la densidad, el Q-Q plot (cuantiles empíricos frente cuantiles teóricos de una distribución normal - ver Figura 2.10 -), el diagrama de cajas, ... Por otro lado, pruebas no paramétricas de bondad del ajuste de los errores a una distribución normal son: Kolmogorov-Smirnov-Lilliefors, Shapiro-Wilks o D'Agostino. A continuación se expone muy brevemente en que consisten algunos de estos procedimientos.

#### La prueba de normalidad de Kolmogorov-Smirnov-Lilliefors

Se denomina **función de distribución empírica**,  $F_n$ , asociada a una muestra  $x_1, x_2, \dots, x_n$ , a la función que asigna a cada número real,  $x$ , la frecuencia relativa de observaciones menores o iguales que  $x$ . Es decir:

$$F_n(x) = \frac{\text{“número de } x_i \leq x\text{”}}{n}$$

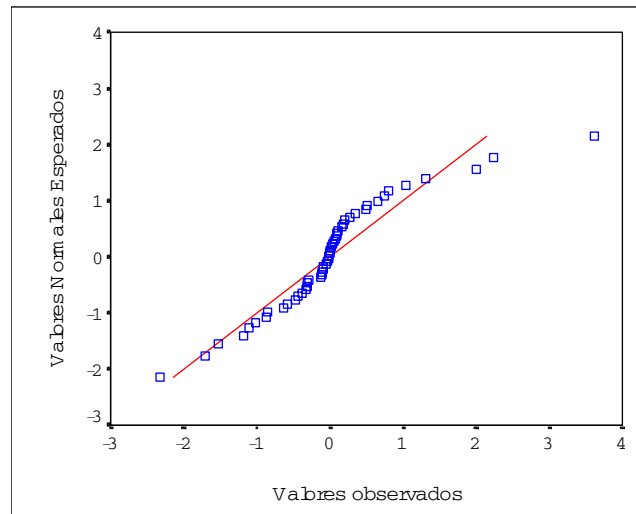


Figura 2.10: Q-Q plot (Diagrama de puntos cuantil-cuantil).

El test de Kolmogorov-Smirnov-Lilliefors se basa en comparar la distribución empírica de la muestra con la distribución normal supuesta para la población de procedencia. Para ello se evalúa el estadístico

$$D_n = \sup_{x \in R} \left| F_n(x) - \Phi \left( \frac{x - \bar{X}}{\hat{S}} \right) \right|$$

o sea, la máxima distancia, en vertical, entre la función de distribución empírica y la función de distribución de la normal ajustada (es decir la de la normal con media y varianza estimadas mediante la media y varianza muestrales respectivamente).

La distribución exacta del estadístico  $D_n$  cuando los datos proceden de la citada normal está tabulada.

### La prueba de normalidad de Shapiro y Wilks

Basado en la distancias entre los estadísticos ordenados de lugar simétrico. Así se tienen en cuenta la distancia entre la mayor y la menor observación muestral, la distancia entre la segunda y la penúltima (dispuestos los datos en orden creciente), etc.

El estadístico de Shapiro-Wilks tiene la forma:

$$\omega = \frac{A^2}{nS^2}$$

siendo

$$A = \sum_{i=1}^{[n/2]} a_{i,n} (x_{(n-i+1)} - x_{(i)})$$

donde las constantes  $a_{i,n}$  vienen dadas a partir de la inversa de la distribución de la normal estándar.

Shapiro y Wilks han tabulado los valores de estas constantes y de  $\omega$ .

Dado que el estadístico  $\omega$  puede interpretarse como el cuadrado del coeficiente de correlación lineal de los puntos muestrales dibujados sobre papel probabilístico normal, la hipótesis de normalidad se rechazará para valores pequeños del estadístico.

¿Qué consecuencias acarrea la ausencia de normalidad de los errores? Este problema afecta directa y especialmente a la estimación de la varianza del modelo pero no al  $F$ -test ni a las estimaciones puntuales de las medias de los tratamientos.

## 2.11 Alternativas al Análisis de la Varianza

### 2.11.1 Introducción

No siempre la técnica del análisis de la varianza es apropiada para comparar tratamientos.

Un modelo Anova univariante es apropiado siempre que:

- la variable respuesta sea univariante,
- la variable respuesta sea medible en una escala de razón o intervalo,
- las variables explicativas o de clasificación sean categóricas,
- sea razonable asumir que los datos procedan de distribuciones normales.

Para fijar conceptos puede resultar muy interesante establecer una clasificación, ni mucho menos estricta, sobre los métodos univariantes como la que presenta George W. Coob en su monografía "Introduction to design and analysis of experiments" y que se reproduce en la Tabla 2.3.

|                                 |                     | La variable respuesta es:          |                                  |  |
|---------------------------------|---------------------|------------------------------------|----------------------------------|--|
| Las variables explicativas son: | Tasa/Intervalo      | Ordinal                            | Nominal                          |  |
| Categóricas                     | ANOVA               | ANOVA<br>Mét. no paramétricos      | Análisis de datos<br>categóricos |  |
| Numéricas                       | Regresión           | Regresión Logística<br>y variantes |                                  |  |
| Algo de todo                    | Regresión<br>ANCOVA |                                    |                                  |  |

Tabla 2.3: Posible clasificación de los métodos univariantes atendiendo al tipo de variables explicativas y respuesta.

Con la excepción del capítulo 4, dedicado al análisis de la covarianza (Ancova), el presente manual se centra exclusivamente en el caso en que las variables de clasificación son categóricas, siendo las categorías los distintos tratamientos, bloques, etc. En otras palabras, el manual aborda la problemática planteada en la segunda fila de la Tabla 2.3. Atendiendo a dicha tabla únicamente resta considerar las alternativas al análisis de la varianza cuando la variable respuesta es categórica, bien nominal bien ordinal.

Para el caso particular de respuesta ordinal todavía la técnica Anova resulta válida si la hipótesis de normalidad es asumible. Naturalmente esto ocurrirá pocas veces ya que si el número de categorías no es suficientemente grande las respuestas tenderán a repetirse muchas veces. No debemos olvidar de todos modos que la ausencia de normalidad no afecta significativamente a, por ejemplo, el contraste  $F$  del cuadro Anova. En cualquier caso, los métodos no paramétricos o de distribución libre proporcionan una alternativa al Anova perfectamente válida cuando la respuesta es ordinal y, en general, cuando la suposición de normalidad no se mantiene. De entre estos métodos se presentan en la memoria algunos **contrastos basados en rangos** como el **contraste de Kruskal-Wallis** (ver sección siguiente), el **contraste de Friedman** (ver sección 5.5.1) o el **contraste de Durbin** (ver sección 5.5.2), válidos para distintos tipos de diseños experimentales. Su potencia, sencillez y validez para distribuciones arbitrarias los convierten en especialmente útiles y aconsejables.

Naturalmente la técnica Anova es del todo inapropiada para el caso de respuesta cualitativa o nominal. Aquí todavía los métodos no paramétricos podrían tener validez pero, en todo caso, la solución más correcta pasa por el análisis de datos categóricos mediante el empleo de tablas de frecuencias y de herramientas como la distribución  $\chi^2$ . Podríamos introducir en este contexto un gran abanico de procedimientos estadísticos (contrastos de bondad de ajuste, contrastos de independencia, análisis de correlación, etc), sin embargo, en esta memoria únicamente se presenta el **contraste de  $\chi^2$  de homogeneidad** (ver sección 2.11.3) que permite comparar distintos tratamientos en base a una respuesta cualitativa.

### 2.11.2 Alternativa No Paramétrica: Prueba de Kruskal-Wallis

Cuando la suposición de normalidad no es asumible y definitivamente se está dispuesto a utilizar un procedimiento alternativo a la prueba  $F$  del análisis de la varianza que no dependa de esta hipótesis, se debe utilizar la prueba de Kruskal-Wallis basada en los rangos de las observaciones.

La prueba de Kruskal-Wallis contrasta la hipótesis nula de que las  $I$  muestras independientes proceden de la misma población y, en particular, que tengan la misma media. En realidad esta prueba es una extensión a  $I$  muestras, con  $I > 2$ , de la prueba de la *suma de rangos de Wilcoxon*, que contrasta la homogeneidad de dos poblaciones a partir de dos muestras aleatorias simples e independientes de ambas sin necesidad de asumir la hipótesis de normalidad.

Para la obtención del estadístico de contraste se siguen los siguientes pasos.

Se mezclan las  $I$  muestras y se ordenan todas las observaciones en sentido creciente para obtener la muestra conjunta ordenada. A continuación se reemplazan las observaciones por sus rangos  $R_{it}$ ,  $i = 1, \dots, I$ ,  $t = 1, \dots, n_i$ , en la muestra conjunta. Se suman entonces los rangos de las observaciones procedentes del  $i$ -ésimo tratamiento,  $R_{i.} = \sum_{t=1}^{n_i} R_{it}$ , para  $i = 1, \dots, I$ . La estadística de prueba es:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^I n_i \left( \frac{R_{i.}}{n_i} - \frac{n+1}{2} \right)^2 \quad (2.36)$$

Obsérvese que  $\bar{R}_{i.} = R_{i.}/n_i$  es el rango medio de las observaciones del tratamiento  $i$ -ésimo. Si la respuesta en todos los tratamientos tuviese la misma distribución entonces  $\bar{R}_{i.} = (n+1)/2$ . O sea, que esencialmente el estadístico de Kruskal-Wallis compara los rangos promedio observados dentro de cada tratamiento con los esperados  $((n+1)/2$  para todos ellos) si la hipótesis nula es cierta.

La conclusión del comentario anterior es clara: rechazar la igualdad de medias para valores grandes de  $H$ . Para ello se tendrá en cuenta que la distribución de  $H$  puede aproximarse bajo la nula por una  $\chi^2$  con  $I - 1$  grados de libertad, si bien conviene precisar que dicha aproximación será aceptable siempre que  $I \geq 4$  y con todos los  $n_i$  mayores o iguales que 6.

Por tanto y resumiendo:

Con el contraste de Kruskal-Wallis, rechazar la igualdad de medias a un nivel de significación  $\alpha$  siempre que  $H > \chi_{I-1, \alpha}^2$ .

La expresión para  $H$  dada en (2.36) admite una forma equivalente pero más cómoda de manejar computacionalmente que viene dada por:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^I \frac{R_{i.}^2}{n_i} - 3(n+1) \quad (2.37)$$

En el supuesto que existan “empates”, esto es observaciones repetidas, el estadístico  $H$  dado por (2.37) se corrige por:

$$H' = C^{-1} H \quad \text{siendo} \quad C = 1 - \frac{\sum_{i=1}^s T_i}{(n-1)n(n+1)}$$

donde  $s$  denota el número de grupos de empates y  $T_i = (t_i - 1)t_i(t_i + 1)$ , con  $t_i$  el número de empates en el  $i$ -ésimo grupo de empates.

Obsérvese que, por definición,  $C$  es siempre menor o igual que 1 y, de ahí,  $H \leq H'$ . Luego es recomendable evaluar primero  $H$  y si el contraste ya resulta significativo no será preciso calcular el valor de  $C$  puesto que con  $H'$  seguirá siendo significativo.

**Ejemplo 2.5** Se consideran las cuatro muestras siguientes:

Muestra número 1: 64, 72, 68, 77, 56 y 95

Muestra número 2: 78, 91, 97, 82, 85 y 77

Muestra número 3: 75, 93, 78, 71, 63 y 76

Muestra número 4: 55, 66, 49, 64, 70 y 68

Por tanto:  $I = 4$ ,  $n_i = 6$ , para  $i = 1, 2, 3, 4$  y  $n = 24$ . El siguiente cuadro esquematiza el proceso de ordenación de las muestras y cómputo de los rangos.

| Datos ordenados |    |    |    | Rangos        |              |               |             | Empates         |       |
|-----------------|----|----|----|---------------|--------------|---------------|-------------|-----------------|-------|
| M1              | M2 | M3 | M4 | M1            | M2           | M3            | M4          | $t_i$           | $T_i$ |
|                 |    | 49 |    |               |              |               | 1           |                 |       |
|                 |    | 55 |    |               |              |               | 2           |                 |       |
| 56              |    |    |    | 3             |              |               |             |                 |       |
|                 |    | 63 |    |               |              | 4             |             |                 |       |
| 64              |    |    |    | 5'5           |              |               |             |                 |       |
|                 |    | 64 |    |               |              |               | 5'5         | 2               | 6     |
|                 |    | 66 |    |               |              |               | 7           |                 |       |
| 68              |    |    |    | 8'5           |              |               |             |                 |       |
|                 |    | 68 |    |               |              | 8'5           |             | 2               | 6     |
|                 |    | 70 |    |               |              |               | 10          |                 |       |
| 72              |    | 71 |    |               |              | 11            |             |                 |       |
|                 |    | 75 |    | 12            |              |               |             |                 |       |
|                 |    | 76 |    |               |              | 13            |             |                 |       |
|                 |    |    |    |               |              | 14            |             |                 |       |
| 77              |    |    |    | 15'5          |              |               |             |                 |       |
|                 | 77 |    |    |               | 15'5         |               |             | 2               | 6     |
|                 | 78 |    |    |               | 17'5         |               |             |                 |       |
|                 |    | 78 |    |               |              | 17'5          |             | 2               | 6     |
|                 | 82 |    |    |               | 19           |               |             |                 |       |
|                 | 85 |    |    |               | 20           |               |             |                 |       |
|                 | 91 |    |    |               | 21           |               |             |                 |       |
|                 |    | 93 |    |               |              | 22            |             |                 |       |
| 95              |    |    |    |               | 23           |               |             |                 |       |
|                 | 97 |    |    |               |              | 24            |             |                 |       |
|                 |    |    |    | $R_1. = 67'5$ | $R_2. = 117$ | $R_3. = 81'5$ | $R_4. = 34$ | $\sum T_i = 24$ |       |

Se obtiene el valor de  $C = 1 - \frac{24}{23 \times 24 \times 25} = 0'9983$  y:

$$H = \left\{ \frac{12}{24 \times 25} \left( \frac{67'5^2}{6} + \frac{117^2}{6} + \frac{81'5^2}{6} + \frac{34^2}{6} \right) - 3 \times 25 \right\} / 0'9983 = 11'83$$

Como  $P(\chi_3^2 > 11'83) < 0'01$  se concluye que los grupos no son homogéneos.

### Contrastes Múltiples

Si y solo si se rechazase la hipótesis nula a un nivel de significación  $\alpha$  sería factible emplear el siguiente criterio de comparaciones múltiples para los contrastes pairwise.

Dos tratamientos  $i$  y  $p$  son considerados distintos al nivel de significación global  $\alpha$  siempre que:

$$\left| \frac{R_{i.}}{n_i} - \frac{R_{p.}}{n_p} \right| > t_{\alpha/2} \left( S^2 \frac{n-1-H}{n-I} \right)^{1/2} \left( \frac{1}{n_i} - \frac{1}{n_p} \right)^{1/2}$$

siendo  $S^2 = \frac{1}{n-1} \left( \sum_{i=1}^I \sum_{t=1}^{n_i} R_{it}^2 - n \frac{(n+1)^2}{4} \right)$ .

### 2.11.3 Alternativa para Variables Respuesta Nominales: Contraste $\chi^2$ de Homogeneidad

Cuando la respuesta es nominal, ni el Anova ni los contrastes basados en rangos parecen la solución apropiada.

**Ejemplo 2.6** Los psiquiatras y profesionales en el estudio de la salud mental tienen todavía un largo camino que recorrer para reconocer con claridad el origen y las causas de determinadas enfermedades mentales. Se sabe que las cuestiones hereditarias tienen algo que ver, así como el entorno familiar actual. Se realizó un pequeño experimento en relación con estas cuestiones que consistió en lo siguiente.

Se seleccionó un grupo de veinte madres con hijos esquizofrénicos y otro grupo de veinte madres control con hijos normales. Todos los hijos tenían el mismo sexo y aproximadamente la misma edad.

Cada madre realizó un test denominado Thematic Apperception Test (TAT). El TAT consiste en mostrar al sujeto una imagen ambigua (por ejemplo un joven mirando a un violín que se encuentra en una mesa situada enfrente de él) y pedirle entonces que elabore una historia imaginaria en base a lo que le sugiere la imagen. Los evaluadores del test puntúan entonces esta historia espontánea de acuerdo a unas reglas y criterios bien definidos. En el experimento que nos ocupa a cada madre se le mostraron 10 imágenes TAT que en general sugerirían historias involucrando a sus padres, hijos y entorno familiar. Las 10 historias se clasificaban en una de cinco posibles categorías (A,B,C,D y E) referentes a su personalidad. Los resultados se recogieron en el formato siguiente:

| Hijo           | Madre | Tipo de historia |   |   |   |   |
|----------------|-------|------------------|---|---|---|---|
|                |       | A                | B | C | D | E |
| Esquizofrénico | 1     | 2                | 2 | 4 | 1 | 1 |
|                | 2     | 1                | 0 | 2 | 1 | 6 |
|                | ⋮     | ⋮                | ⋮ | ⋮ | ⋮ | ⋮ |
|                | 20    | 1                | 1 | 4 | 2 | 2 |
| Normal         | 21    | 8                | 0 | 1 | 0 | 1 |
|                | 22    | 4                | 0 | 0 | 1 | 5 |
|                | ⋮     | ⋮                | ⋮ | ⋮ | ⋮ | ⋮ |
|                | 40    | 4                | 0 | 2 | 0 | 4 |

Podría modelizarse el experimento como: dos factores de interés cruzados: “diagnos de los hijos” (con dos niveles: esquizofrénico y normal) y “tipo de historia” (con cinco niveles A, . . . , E) y un factor de bloqueo, “madre”, anidado en el factor “diagnos de los hijos” y cruzado con el factor “tipo de historia”. En realidad se trataría de un diseño denominado Split-Plot (no abordado en la memoria). A continuación emplearíamos la técnica Anova para establecer las comparaciones entre los factores incluidos en el modelo.

El planteamiento anterior sería a todas luces erróneo. Obsérvese que para evaluar diferencias entre madres (bloques) habría que calcular el valor medio de la respuesta para cada madre y comparar estos promedios. Pero obviamente el número medio de historias por categoría para cada madre es siempre igual:  $10/5 = 2$ . Análogamente la respuesta media para cada nivel de diagnos es también dos historias por madre y por categoría.

La estructura del diseño convierte en inútil el empleo de las medias para comparar a las madres o a los niveles de diagnos, el principal factor de interés.

Es muy importante tener siempre presente que el Anova emplea como herramienta fundamental medias y desviaciones típicas para comparar dos o más grupos de unidades. Por tanto, antes de decidir si un modelo Anova es adecuado o no cabe preguntarse si en verdad los valores medios son interpretables y significativos y si esa interpretabilidad es válida para resolver las comparaciones que deseamos hacer.

En el ejemplo 2.6 los números “en crudo” de la tabla no constituyen la respuesta. En realidad son frecuencias, cuentas, de modo que cada número nos dice cuántas historias pertenecían a un grupo particular. Luego la verdadera respuesta es nominal, con categorías A, B, C, D y E.

### Presentación de los datos

Se desean comparar  $I$  tratamientos de un factor de interés en base a una respuesta nominal  $Y$  con  $K$  posibles categorías. Para ello se dispone de una muestra aleatoria de cada tratamiento de tamaño  $n_i$ :  $y_{i1}, y_{i2}, \dots, y_{in_i}$ , para cada  $i = 1, \dots, I$ ; de modo que las  $I$  muestras son independientes.

Dado que la respuesta es nominal, cada dato  $y_{it}$  equivale a clasificar la observación  $t$ -ésima del  $i$ -ésimo tratamiento en alguna de las  $K$  características nominales. Por ello es habitual presentar los resultados en una tabla de frecuencias con  $I$  filas y  $K$  columnas donde las filas representan a los distintos tratamientos, las columnas a las distintas categorías de la respuesta nominal y en la celda  $(i, k)$  figura la frecuencia observada de la categoría  $k$  con el tratamiento  $i$ .

**Ejemplo 2.7** Se investiga que tratamiento de cuatro posibles (codificados de 1 a 4) es más adecuado para curar una cierta enfermedad. Se seleccionan al azar cuatro grupos de enfermos de tamaños  $n_1 = 150$ ,  $n_2 = 120$ ,  $n_3 = 130$  y  $n_4 = 160$ . A cada grupo se le aplica un tratamiento distinto y, pasado un tiempo razonable, se observa si el paciente HA MEJORADO (categoría **Mejor**), HA EMPEORADO (categoría **Peor**) o ESTÁ IGUAL (categoría **Igual**).

La información del experimento puede resumirse en una tabla como la que sigue:

| Tratamientos  | Categorías    |               |                | Total             |
|---------------|---------------|---------------|----------------|-------------------|
|               | Peor          | Igual         | Mejor          |                   |
| Tratamiento 1 | $O_{11} = 7$  | $O_{12} = 28$ | $O_{13} = 115$ | $F_1 = n_1 = 150$ |
| Tratamiento 2 | $O_{21} = 15$ | $O_{22} = 20$ | $O_{23} = 85$  | $F_2 = n_2 = 120$ |
| Tratamiento 3 | $O_{31} = 10$ | $O_{32} = 30$ | $O_{33} = 90$  | $F_3 = n_3 = 130$ |
| Tratamiento 4 | $O_{41} = 5$  | $O_{42} = 40$ | $O_{43} = 115$ | $F_4 = n_4 = 160$ |
| Total         | $C_1 = 37$    | $C_2 = 118$   | $C_3 = 405$    | $T = 560$         |

En lo que sigue y tal y como se detalla en la tabla del ejemplo 2.7 se adoptará la siguiente notación:

$O_{ik}$  = frecuencia de observaciones con el tratamiento  $i$  clasificadas en la categoría  $k$ ,

$F_i = \sum_{k=1}^K O_{ik}$  = total de observaciones realizadas con el tratamiento  $i$ ,

$C_k = \sum_{i=1}^I O_{ik}$  = total de observaciones clasificadas en la categoría  $k$ ,

$T = \sum_{i=1}^I \sum_{k=1}^K O_{ik} = \sum_{i=1}^I F_i = \sum_{k=1}^K C_k$  = total de observaciones.

**Contraste de interés**

Como en capítulos precedentes el contraste de hipótesis de interés es:

$$\begin{cases} H_0 : \text{no existen diferencias entre los tratamientos en términos de la respuesta considerada} \\ H_1 : \text{al menos dos tratamientos difieren en términos de la respuesta considerada} \end{cases} \quad (2.38)$$

Ahora bien, en este caso al tratarse de una respuesta nominal carece de sentido plantear la comparación en términos de medias. Los tratamientos serán “iguales” si la probabilidad de clasificar a un individuo en una determinada categoría de la respuesta es igual para todos los tratamientos. En otros términos, si la distribución de probabilidad para la respuesta nominal es la misma bajo todos los tratamientos.

Se trata por tanto de un contraste de homogeneidad. De hecho, denotando por  $p_{ik}$  a la probabilidad de la categoría  $k$  bajo el tratamiento  $i$ , el contraste de hipótesis (2.38) puede escribirse en la forma:

$$\begin{cases} H_0 : p_{1k} = p_{2k} = \dots = p_{Ik} = p_{.k} \text{ para todo } k = 1, \dots, K \\ H_1 : \text{existen } i \text{ y } j \text{ tales que } p_{ik} \neq p_{jk} \text{ para algún } k = 1, \dots, K \end{cases} \quad (2.39)$$

**Ejemplo 2.8** El escenario real (poblacional) en el problema planteado en el ejemplo 2.7 es:

|               | Categorías |          |          |       |
|---------------|------------|----------|----------|-------|
| Tratamientos  | Peor       | Igual    | Mejor    | Total |
| Tratamiento 1 | $p_{11}$   | $p_{12}$ | $p_{13}$ | 1     |
| Tratamiento 2 | $p_{21}$   | $p_{22}$ | $p_{23}$ | 1     |
| Tratamiento 3 | $p_{31}$   | $p_{32}$ | $p_{33}$ | 1     |
| Tratamiento 4 | $p_{41}$   | $p_{42}$ | $p_{43}$ | 1     |

donde, por ejemplo,  $p_{11}$  representa la probabilidad de mejoría con el tratamiento 1 y  $p_{43}$  representa la probabilidad de empeoramiento con el tratamiento 3. Entenderemos que no existen diferencias entre los tratamientos siempre que las cuatro filas de la tabla anterior sean iguales pues en ese caso las proporciones reales de mejoría, empeoramiento o de nulo efecto no dependerán del tratamiento.

El contraste de interés:

$$\begin{cases} H_0 : p_{1k} = p_{2k} = p_{3k} = p_{4k} = p_{.k} \text{ para todo } k = 1, 2, 3, 4 \\ H_1 : \text{existen } i \text{ y } j \text{ tales que } p_{ik} \neq p_{jk} \text{ para algún } k = 1, 2, 3, 4 \end{cases}$$

### Resolución del contraste

La idea del contraste  $\chi^2$  de homogeneidad de poblaciones se basa en comparar las frecuencias observadas ( $O_{ik}$ ) en cada celda de la tabla de frecuencias con las frecuencias “esperadas” para dichas celdas si la hipótesis nula del contraste (2.39) fuese cierta.

Veamos como obtener estimaciones de dichas frecuencias esperadas bajo  $H_0$  a partir de los datos muestrales:

1. Sea  $\mu_{ik}$  el número esperado de observaciones con el tratamiento  $i$  en la categoría  $k$ , para  $i = 1, \dots, I$  y  $k = 1, \dots, K$ .

2. Es claro que

$$\mu_{ik} = F_i \times p_{ik} \quad \text{para } i = 1, \dots, I \text{ y } k = 1, \dots, K$$

3. Si  $H_0$  es cierta entonces  $p_{ik}$  es independiente de  $i$ , de modo que  $p_{ik} = p_{.k}$  para todo  $i$ , donde  $p_{.k}$  denota la proporción real de observaciones en la categoría  $k$  para cualquier tratamiento.

Por tanto, bajo  $H_0$ :

$$\mu_{ik} = F_i \times p_{.k} \quad \text{para } i = 1, \dots, I \text{ y } k = 1, \dots, K$$

4. Un estimador natural de  $p_{.k}$  será

$$\hat{p}_{.k} = \frac{C_k}{T} \quad \text{para } k = 1, \dots, K$$

de modo que un estimador para  $\mu_{ik}$  se obtiene haciendo:

$$\hat{\mu}_{ik} = E_{ik} = F_i \times \hat{p}_{.k} = \frac{F_i \times C_k}{T} \quad \text{para } i = 1, \dots, I \text{ y } k = 1, \dots, K \quad (2.40)$$

Concluyendo:

El estimador de la frecuencia esperada para una celda  $(i, k)$  cuando  $H_0$  es cierta se obtiene multiplicando el total de la fila  $i$ -ésima por el total de la columna  $k$ -ésima y dividiendo el resultado por el número total de observaciones.

Se prueba entonces que el estadístico

$$\chi^2 = \sum_{i=1}^I \sum_{k=1}^K \frac{(O_{ik} - E_{ik})^2}{E_{ik}} = \sum_{i=1}^I \sum_{k=1}^K \frac{O_{ik}^2}{E_{ik}} - T \quad (2.41)$$

se distribuye aproximadamente según una distribución  $\chi^2$  de Pearson con  $(I-1)(K-1)$  grados de libertad.

Valores grandes del estadístico sugieren una gran discrepancia entre frecuencias observadas y frecuencias esperadas bajo  $H_0$  en, al menos, dos de las celdas. Por tanto sugieren rechazar la hipótesis nula de homogeneidad.

En el contraste (2.39), rechazar la hipótesis nula a un nivel de significación  $\alpha$  siempre que el valor observado del estadístico dado en (2.41) sea mayor que  $\chi^2_{(I-1)(K-1),\alpha}$ .

**Ejemplo 2.9** Retomando el ejemplo 2.7, se puede completar la tabla de frecuencias observadas con las frecuencias esperadas bajo la hipótesis nula de igualdad de los tratamientos sin más que aplicar (2.40).

| Tratamientos  | Categorías                       |                                   |                                     | Total       |
|---------------|----------------------------------|-----------------------------------|-------------------------------------|-------------|
|               | Peor                             | Igual                             | Mejor                               |             |
| Tratamiento 1 | $O_{11} = 7$<br>$E_{11} = 9'91$  | $O_{12} = 28$<br>$E_{12} = 31'61$ | $O_{13} = 115$<br>$E_{13} = 108'48$ | $F_1 = 150$ |
| Tratamiento 2 | $O_{21} = 15$<br>$E_{21} = 7'93$ | $O_{22} = 20$<br>$E_{22} = 25'28$ | $O_{23} = 85$<br>$E_{23} = 86'79$   | $F_2 = 120$ |
| Tratamiento 3 | $O_{31} = 10$<br>$E_{31} = 8'59$ | $O_{32} = 30$<br>$E_{32} = 27'39$ | $O_{33} = 90$<br>$E_{33} = 94'02$   | $F_3 = 130$ |
| Tratamiento 4 | $O_{41} = 5$<br>$E_{41} = 10'57$ | $O_{42} = 40$<br>$E_{42} = 33'72$ | $O_{43} = 115$<br>$E_{43} = 115'71$ | $F_4 = 160$ |
| Total         | $C_1 = 37$                       | $C_2 = 118$                       | $C_3 = 405$                         | $T = 560$   |

Se calcula a continuación el valor del estadístico de contraste empleando (2.41) y resulta:

$$\chi^2 = \sum_{i=1}^I \sum_{k=1}^K \frac{O_{ik}^2}{E_{ik}} - T = 13'87 > \chi_{6,0'05}^2$$

concluyendo que, con un 5% de significación, los tratamientos no conducen a los mismos resultados de la variable respuesta.

## 2.12 Problemas Propuestos

**Problema 1** Se desea planificar un experimento de un sólo factor tratamiento con cuatro niveles predeterminados y sin factores de bloqueo. Se requieren 5 observaciones por nivel. Realizar la aleatorización del experimento.

Repetir el planteamiento anterior pero suponiendo que ahora se precisan 3 unidades para el primer nivel y 5 para el resto.

**Problema 2** Completar los pasos de la lista para la planificación y análisis del experimento de los jabones expuesto en el capítulo 1. Téngase en cuenta que del experimento piloto

resultó una primera estimación de la varianza igual a  $0'007 \text{ gr}^2$ . Los datos están en el archivo JABON.SAV.

Además:

1. Si el experimentador desea comparar la pérdida de peso del jabón normal con la pérdida promedio de los otros dos, ¿qué decisión tomaría?
2. Si inicialmente el objetivo era comparar la pérdida de peso de los jabones desodorante e hidratante con respecto al normal, ¿qué método utilizaríamos? Comparar los resultados con distintos métodos al mismo nivel de significación.
3. Encontrar dos contrastes ortogonales y comprobar que sus sumas de cuadrados suman  $scT$ .

**Problema 3** Completar el análisis estadístico del ejemplo de la relación duración por coste unitario de las pilas expuesto en el capítulo 1. Los datos están en el archivo PILAS.SAV.

**Problema 4** (Richard Davis de la Universidad de Ohio en 1987) Interesa determinar el efecto de cambiar el número de revoluciones por minuto (rpm) de una bomba rotatoria que impulsa un líquido por un circuito. Las rpm se controlan directamente mediante el tacómetro situado en la consola de la bomba. La tasa de corriente del fluido se mide en litros por minuto. Se seleccionan 5 niveles equiespaciados de rpm: 10, 75, 100, 125 y 150 rpm (codificados respectivamente del 1 al 5). Los resultados de las observaciones se recogen en el archivo BOMBA.SAV.

1. Realizar el análisis de la varianza y testar la hipótesis “el número de rpm no afecta a la tasa de corriente del líquido”.
2. ¿Qué opinas del resultado?
3. Calcular un intervalo de confianza para la varianza al 90%

**Problema 5** (Gutsell, Biometrics, 1951) Los datos del archivo TRUCHAS.SAV muestran las cantidades de hemoglobina (en gramos por cien mililitros) en sangre de una especie de truchas marrones. Las truchas se seleccionaron al azar en cuatro dependencias distintas de una piscifactoría. En cada dependencia se añadía a la comida de los peces una cantidad distinta de sulfamerazina por cada cien libras de comida. En concreto: 0, 5, 10 y 15 gramos (codificados del 1 al 4). Las mediciones de hemoglobina se tomaron a diez truchas seleccionadas al azar de cada dependencia después de 35 días.

1. Construir un diagrama de dispersión con los datos y comentar los resultados.
2. Escribir un modelo adecuado para este experimento.
3. Calcular los estimadores mínimo-cuadráticos para la respuesta media de cada uno de los tratamientos. Plasmarlos en algún gráfico y hacer valoraciones.

4. Contrastar la hipótesis nula “la sulfamerazina no influye sobre el contenido de hemoglobina”.
5. Calcular un intervalo de confianza al 95% para  $\sigma^2$ .
6. Supuesto que se repite el experimento con los mismos cuatro niveles del factor tratamiento: ¿qué número de truchas deberíamos examinar en cada dependencia para que el test Anova detectase una diferencia real (i)  $\Delta = 1'5$ , (ii)  $\Delta = 1$ , (iii)  $\Delta = 2$  el 95% de las veces con un nivel de significación del 5%.
7. Realizar las comparaciones pareadas de los cuatro tratamientos usando el método de Tukey con un 2% de significación.
8. Comparar el efecto de no incluir en la comida sulfamerazina con el efecto medio de los otros tres niveles (otra vez al 2% de significación).
9. Contrastar las hipótesis de tendencias lineal y cuadrática en el contenido de hemoglobina debido a la sulfamerazina.

**Problema 6** Completar el análisis estadístico del ejemplo de los tiempos de inflado de globos expuesto en el Ejemplo 2.4 en la sección 2.10.3. Los datos están en el archivo GLOBOS.SAV.

**Problema 7** Un ingeniero civil está interesado en determinar si cuatro métodos diferentes para estimar la frecuencia de inundaciones producen estimaciones equivalentes del flujo máximo (medido en pies cúbicos por segundo) cuando se aplican a una misma cuenca. Cada procedimiento se evaluó seis veces en la misma cuenca y las observaciones figuran en el archivo denominado INUNDA.SAV. Realizar el análisis estadístico y comprobar la heterocedasticidad de los residuos. ¿Qué transformación es más apropiada para conseguir varianza constante?

**Problema 8** Una empresa ha experimentado con veinte nuevos empleados la validez de hasta cuatro programas distintos de entrenamiento y capacitación para un determinado trabajo. Los veinte empleados se dividieron aleatoriamente en grupos de cinco y cada grupo fue asignado (también al azar) a un programa distinto bajo el control de un mismo supervisor. Al final de un cierto período de tiempo preespecificado de antemano, el supervisor estableció un ranking entre los veinte empleados según su capacidad para el trabajo resultando:

| Programa | Ranking |    |    |    |    |
|----------|---------|----|----|----|----|
| <b>1</b> | 4       | 6  | 7  | 2  | 10 |
| <b>2</b> | 1       | 8  | 12 | 3  | 11 |
| <b>3</b> | 20      | 19 | 16 | 14 | 5  |
| <b>4</b> | 18      | 15 | 17 | 13 | 9  |

¿Existen diferencias significativas entre las efectividades de los cuatro programas de capacitación? Si fuese así, ¿cuáles parecen ser diferentes? Explicar y justificar el procedimiento analítico empleado.



## Capítulo 3

# Diseños con Dos Factores Tratamiento Cruzados

### 3.1 Introducción

En este capítulo se estudian diseños completamente aleatorizados con dos factores tratamiento  $A$  y  $B$  cuyos niveles se cruzan. El factor  $A$  tiene  $a$  niveles que se codifican mediante  $1, 2, \dots, a$  y el factor  $B$   $b$  niveles codificados por  $1, 2, \dots, b$ . Se tienen entonces un total de  $ab$  tratamientos (o combinaciones de tratamientos) distintos codificados por  $11, 12, \dots, 1b, 21, 22, \dots, 2b, \dots, ab$ . A lo largo del capítulo,  $y_{ijt}$  denota la  $t$ -ésima observación del tratamiento  $ij$  (una realización particular de la variable aleatoria  $Y_{ijt}$ ) y  $n_{ij}$  denota el número de unidades experimentales asignadas al tratamiento  $ij$ , de modo que el tamaño del experimento es  $n = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$ .

### 3.2 El Significado de la Interacción

Se dice que no existe interacción entre los dos factores tratamiento cuando sus efectos sobre la respuesta son aditivos. En otros términos, la diferencia de las respuestas medias teóricas en dos niveles cualesquiera de un factor es constante en todos los niveles del otro factor y viceversa:

$$\boxed{\text{Para todo } i = 1, \dots, a: \quad \mu_{ij} - \mu_{1j} = \mu_{ik} - \mu_{1k}, \quad j \text{ y } k \text{ arbitrarios en } 1, \dots, b}$$

Un gráfico ilustrativo de la posible existencia de interacción es el denominado **gráfico de interacción**. Para construir un gráfico de interacción se marcan en el eje de abscisas los niveles de uno de los dos factores tratamiento, por ejemplo el  $A$ , y se dibuja la nube de puntos

$$\left\{ \left( i, \bar{y}_{ij} \right), \text{ para } i = 1, \dots, a, j = 1, \dots, b, \text{ siendo } \bar{y}_{ij} = \frac{1}{n_{ij}} \sum_{t=1}^{n_{ij}} y_{ijt} \right\},$$

uniendo a continuación con segmentos las medias muestrales  $\bar{y}_{ij}$  con igual  $j$ .

Aunque los gráficos de interacción son muy intuitivos y útiles, pueden conducir a interpretaciones peligrosas debido a que en ellos no se refleja el tamaño del error experimental. A veces percibiremos la existencia de interacción en el gráfico y, sin embargo, el error experimental será lo suficientemente grande como para que el análisis de la varianza no la detecte como significativa (y viceversa). Por consiguiente, habrá que ser muy prudentes con las conclusiones derivadas de un gráfico de este tipo.

Frecuentemente el interés del diseño radica fundamentalmente en evaluar la contribución individual de cada factor tratamiento sobre la respuesta observada. A los efectos marginales de cada factor se les denomina **efectos principales**. Sin embargo, cuando el efecto interacción entre ambos factores sea importante, puede resultar imposible examinar por separado cada uno de los efectos principales.

El Ejemplo 3.1 ilustra adecuadamente el significado de la interacción y los problemas de interpretar por separado los efectos de los factores tratamiento en su presencia.

**Ejemplo 3.1** Se evalúan, de algún modo (por ejemplo vía calificaciones en pruebas escritas), los resultados obtenidos con dos métodos de enseñanza distintos (método 1 y método 2) y con tres profesores distintos (profesores 1, 2 y 3). Las Figuras 3.1 y 3.2 muestran ocho posibles gráficos de interacción reflejando otros tantos resultados posibles para el experimento.

En las cuatro situaciones descritas en la Figura 3.1 la interacción no existe. Las líneas uniendo las medias obtenidas con cada método son paralelas o coincidentes y, por ello, las diferencias (cuando existen) entre los niveles de cada factor son estables a través de los niveles del otro factor.

Esto no ocurre en los gráficos en la Figura 3.2. En todos ellos el efecto interacción es significativo. Por ejemplo, en el gráfico (a) todos los profesores obtienen las calificaciones más altas con el método de enseñanza 1; ahora bien, las diferencias son muy grandes en el caso del profesor 1 y muy pequeñas con el profesor 3. Es claro que existen diferencias entre los métodos de enseñanza pero, ¿existen entre los profesores? (obsérvese que los promedios muestrales de todos ellos son idénticos). Por el contrario, en el gráfico (c), está claro que existen diferencias entre los profesores. Dos obtienen mejores puntuaciones con el método 1 y uno con el método 2. Sin embargo, si se hubiesen ignorado los métodos, las diferencias no existirían pues las puntuaciones medias de los tres son idénticas. Más aún, un Anova no detectaría diferencias significativas.

El Ejemplo 3.1 permite establecer una primera conclusión a tener siempre en cuenta:

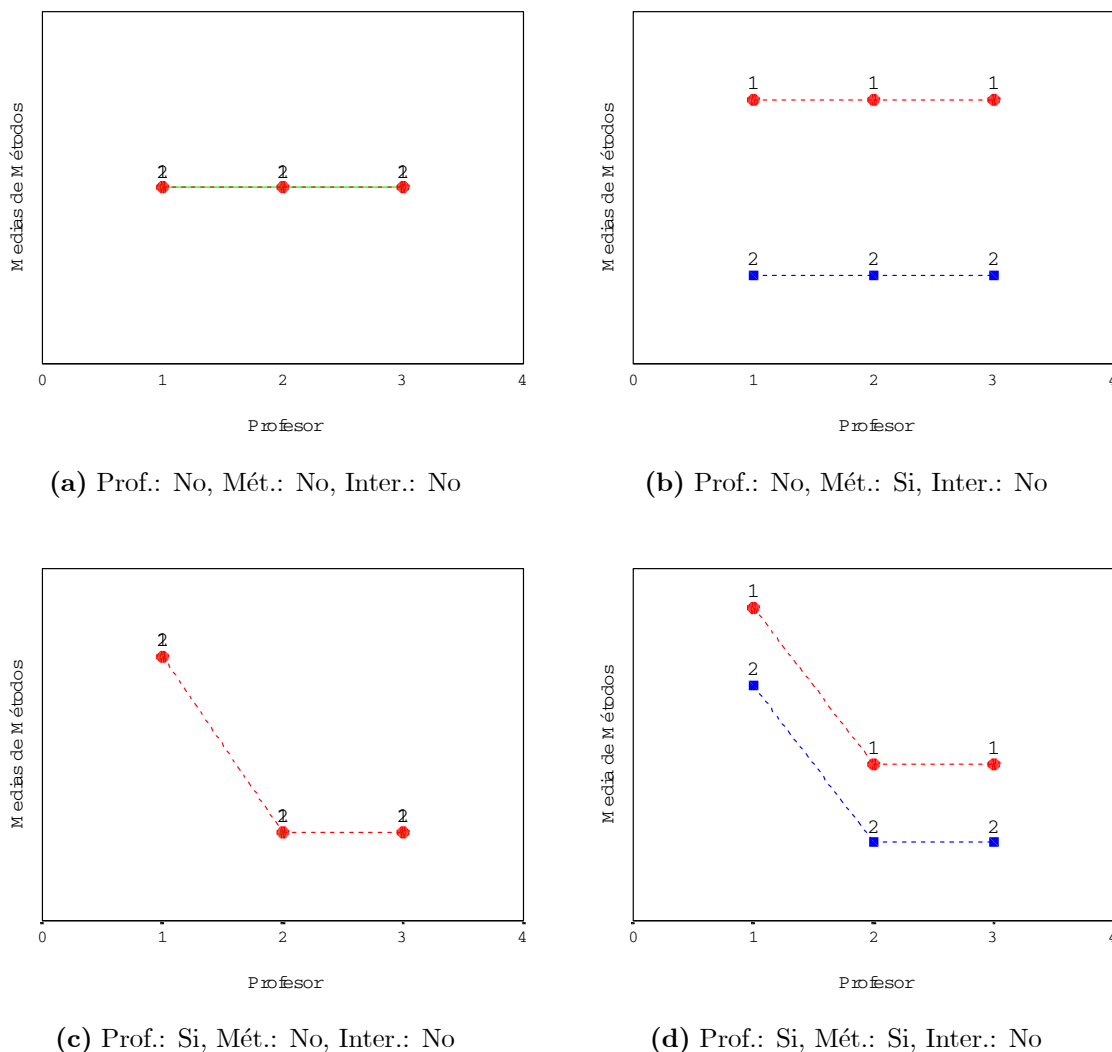


Figura 3.1: Cuatro posibles configuraciones de los efectos sin presencia de interacción.

Si la interacción es significativa, será complejo examinar los efectos de cada factor tratamiento por separado. Por ejemplo, la presencia de interacción significativa podría encubrir diferencias reales entre los niveles de algún factor, de modo que no se detectasen diferencias significativas entre ellos en el análisis estadístico.

### 3.3 Modelos Matemáticos Para Dos Factores Tratamiento

De la misma forma que en el modelo de una vía, si  $n_{ij}$  unidades experimentales son asignadas al tratamiento  $ij$ , entonces, para cada  $t = 1, \dots, n_{ij}$ ,  $i = 1, \dots, a$ ,  $j = 1, \dots, b$ , se tiene:

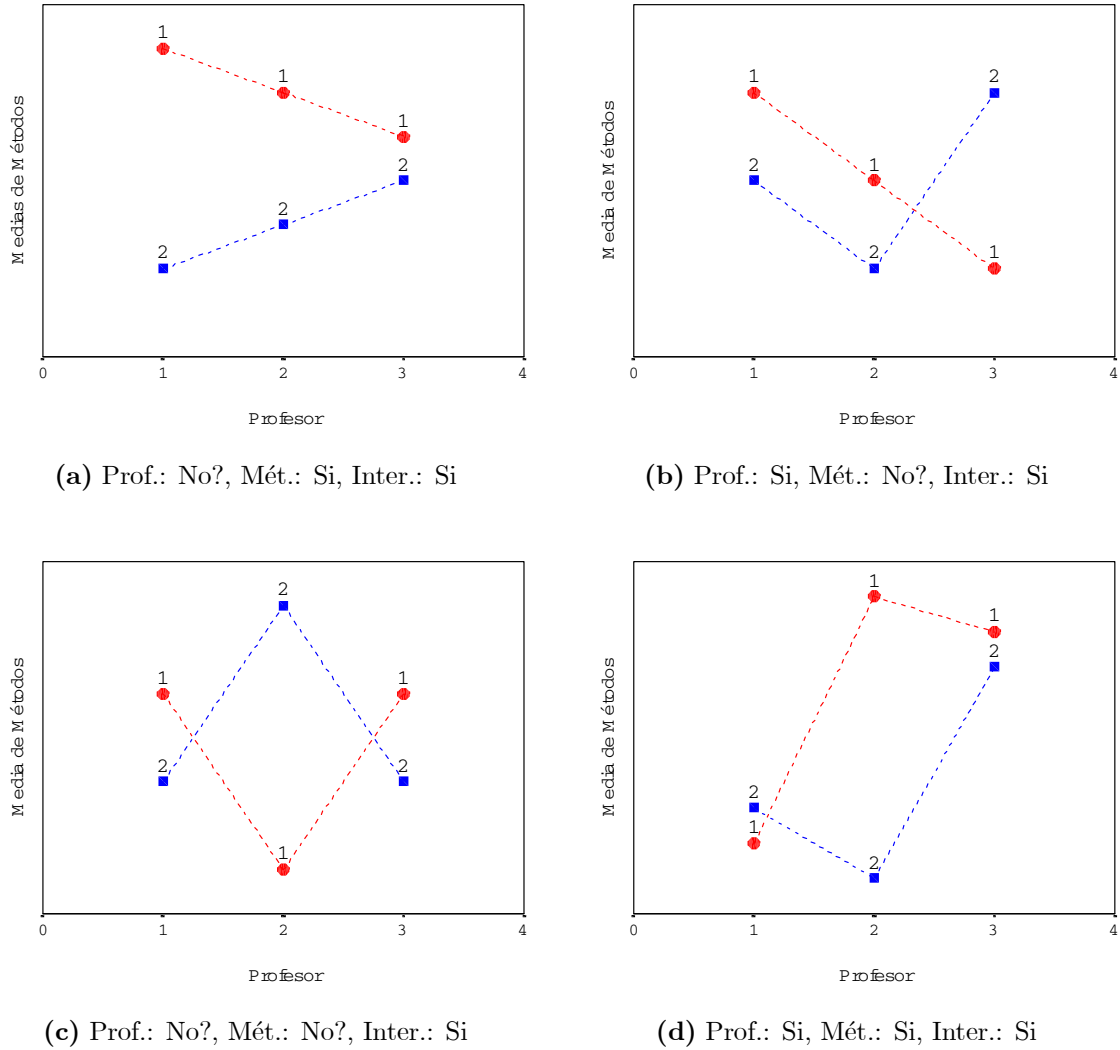


Figura 3.2: Cuatro posibles configuraciones de los efectos con presencia de interacción.

$$\underbrace{Y_{ijt}}_{\text{aleatorio}} = \underbrace{\mu + \tau_{ij}}_{\text{determinista}} + \underbrace{\varepsilon_{ijt}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{ijt} \text{ i.i.d. según una } N(0, \sigma) \quad \forall i, j, t \quad (3.1)$$

de modo que  $\tau_{ij}$  denota el efecto sobre la respuesta media esperada debido al tratamiento  $ij$ . Denotando a dicha respuesta por  $\mu_{ij}$ , se tiene:  $\mu_{ij} = \mu + \tau_{ij}$ . A menudo este modelo se denomina **modelo en medias por celda**.

Una representación equivalente que modeliza la respuesta real media de cada tratamiento como la suma de los efectos principales de los tratamientos y del efecto interacción es:

$$\underbrace{Y_{ijt}}_{\text{aleatorio}} = \overbrace{\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}}^{\text{determinista}} + \underbrace{\varepsilon_{ijt}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{ijt} \text{ i.i.d. según una } N(0, \sigma) \quad \forall i, j, t \quad (3.2)$$

de modo que  $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ , siendo  $\mu$  una constante,  $\alpha_i$  el efecto (positivo o negativo) sobre la respuesta debido a que se observa el nivel  $i$  del factor  $A$ ,  $\beta_j$  el efecto (positivo o negativo) sobre la respuesta debido a que se observa el nivel  $j$  del factor  $B$  y  $(\alpha\beta)_{ij}$  el efecto extra (positivo o negativo) sobre la respuesta debido a que se observan conjuntamente los niveles  $i$  y  $j$  de los factores  $A$  y  $B$  respectivamente.

El modelo (3.2) se conoce como **modelo completo de dos vías** o **modelo de análisis de la varianza de dos vías**.

Si, ocasionalmente, experimentos similares previos o hechos científicos contrastados garantizan con una razonable *seguridad* que ambos factores no interaccionan, el experimento se modeliza a través de:

$$\underbrace{Y_{ijt}}_{\text{aleatorio}} = \mu + \alpha_i + \beta_j + \underbrace{\varepsilon_{ijt}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{ijt} \text{ i.i.d. según una } N(0, \sigma) \quad \forall i, j, t \quad (3.3)$$

El modelo (3.3) es un “submodelo” del modelo completo de dos vías y se denomina **modelo de efectos principales de dos vías** o **modelo aditivo de dos vías** dado que el efecto sobre la respuesta del tratamiento  $ij$  se modeliza como la suma de los efectos individuales de cada factor.

Usar el modelo de efectos principales sólo cuando la certeza de que no exista interacción entre los factores sea importante.

Si no se tiene un conocimiento razonable acerca de la interacción (cuidado con los gráficos de interacción) debe seleccionarse un modelo completo. El motivo es que la inferencia sobre los efectos principales cuando no se ha considerado interacción erróneamente puede ser muy confusa (el error experimental estará inflado artificialmente).

Por tanto la estrategia a seguir es:

1. Con sospechas de interacción empezar contrastando el efecto de ésta en un modelo completo de dos vías.

2. Si no resulta significativa, entonces continuar con el análisis examinando los efectos principales **en el mismo modelo**. En otras términos, no deberíamos cambiar al modelo de efectos principales pues no son modelos equivalentes.
3. Si resulta significativo el efecto interacción, entonces generalmente es preferible pasar a un modelo de una vía donde los niveles son todas las combinaciones de tratamientos y examinar así sus posibles diferencias. Otra posibilidad es examinar las diferencias entre niveles de un factor manteniendo fijos los niveles del otro.

### 3.4 Contrastes

#### 3.4.1 Contrastes de Efectos Principales e Interacciones

Naturalmente, por analogía con el modelo de una vía, los contrastes relativos a efectos de tratamientos:

$$\boxed{\sum_{i=1}^a \sum_{j=1}^b d_{ij} \tau_{ij}, \text{ con } \sum_{i=1}^a \sum_{j=1}^b d_{ij} = 0} \quad (3.4)$$

son todos ellos estimables. Un ejemplo es:  $\tau_{14} - \tau_{23}$ , donde se ha tomado  $d_{14} = 1$ ,  $d_{23} = -1$  y  $d_{ij} = 0$  para todo  $ij$  distinto de 14 y 23.

Pero, además de los contrastes de efectos tratamientos, en el modelo completo de dos vías interesan también *los contrastes de interacciones y los contrastes de efectos principales*.

Los **contrastos de interacción** son aquellos orientados a medir si las líneas de los gráficos de interacción son o no paralelas. Por ejemplo, si en (3.4) se toman los valores  $d_{sh} = d_{(s+1)q} = 1$ ,  $d_{(s+1)h} = d_{sq} = -1$  y 0 en el resto de coeficientes se obtiene:

$$\left( \tau_{sh} - \tau_{(s+1)h} \right) - \left( \tau_{sq} - \tau_{(s+1)q} \right) \iff \left( (\alpha\beta)_{sh} - (\alpha\beta)_{(s+1)h} \right) - \left( (\alpha\beta)_{sq} - (\alpha\beta)_{(s+1)q} \right) \quad (3.5)$$

De hecho los contrastes de interacción son siempre de la forma:

$$\boxed{\sum_{i=1}^a \sum_{j=1}^b d_{ij} \tau_{ij} = \sum_{i=1}^a \sum_{j=1}^b d_{ij} (\alpha\beta)_{ij} \text{ con } \sum_{i=1}^a d_{ij} = 0 \text{ para cada } j \text{ y } \sum_{j=1}^b d_{ij} = 0 \text{ para cada } i} \quad (3.6)$$

Los **contrastos de efectos principales** vienen dados por:

$$\boxed{\sum_{i=1}^a c_i \bar{\tau}_{i.}, \text{ con } \sum_{i=1}^a c_i = 0 \quad \text{y} \quad \sum_{j=1}^b k_j \bar{\tau}_{.j}, \text{ con } \sum_{j=1}^b k_j = 0} \quad (3.7)$$

Obsérvese que, en el modelo completo, los contrastes de efectos principales incluyen los efectos promedio de las interacciones. En efecto, sea para cada  $i$ :

$$\overline{(\alpha\beta)}_{.i} = \frac{1}{b} \sum_{j=1}^b (\alpha\beta)_{ij},$$

entonces:

$$\boxed{\sum_{i=1}^a c_i \bar{\tau}_{.i} = \sum_{i=1}^a c_i \left( \alpha_i + \overline{(\alpha\beta)}_{.i} \right) = \sum_{i=1}^a c_i \alpha_i^*, \quad \text{con } \alpha_i^* = \alpha_i + \overline{(\alpha\beta)}_{.i}.} \quad (3.8)$$

De (3.8) se deduce que si el efecto de interacción es muy pequeño entonces  $\alpha_i^* \approx \alpha_i$  y el contraste compara los efectos marginales de los niveles de un factor. En otro caso, los efectos principales se confunden con los efectos promedio de interacciones.

Análogamente

$$\boxed{\sum_{j=1}^b k_j \bar{\tau}_{.j} = \sum_{j=1}^b k_j \left( \beta_j + \overline{(\alpha\beta)}_{.j} \right) = \sum_{j=1}^b k_j \beta_j^*, \quad \text{con } \beta_j^* = \beta_j + \overline{(\alpha\beta)}_{.j}} \quad (3.9)$$

También en ocasiones interesa comparar los efectos de los niveles de un factor aisladamente en cada nivel del otro factor. Así, en el Ejemplo 3.1 podría ser de interés elegir el mejor método de enseñanza para cada profesor. Si interesa comparar los efectos del factor  $B$  para cada nivel del factor  $A$ , los contrastes son de la forma:

$$\boxed{\sum_{j=1}^b c_j \tau_{ij}, \quad \text{con } \sum_{j=1}^b c_j = 0 \text{ para cada } i = 1, \dots, a} \quad (3.10)$$

Los contrastes de la forma (3.10) se denominan *contrastos simples* en los niveles de  $B$ . Como caso especial se encuentran las *diferencias apareadas simples* del factor  $B$ :

$$\boxed{\tau_{ih} - \tau_{ij} \text{ para cada } i = 1, \dots, a} \quad (3.11)$$

En el modelo de efectos principales se asume que no existe interacción y por consiguiente los contrastes de efectos principales para  $A$  y  $B$  son respectivamente:

$$\boxed{\sum_{i=1}^a c_i \bar{\tau}_{.i} = \sum_{i=1}^a c_i \alpha_i, \quad \text{con } \sum_{i=1}^a c_i = 0 \quad \text{y} \quad \sum_{j=1}^b k_j \bar{\tau}_{.j} = \sum_{j=1}^b k_j \beta_j, \quad \text{con } \sum_{j=1}^b k_j = 0} \quad (3.12)$$

### 3.4.2 Contrastes como Listas de Coeficientes

A veces, en lugar de escribir un contraste explícitamente, resulta más cómodo e incluso más conveniente referirse a él a través de un listado de sus coeficientes. En un modelo de dos vías la lista puede incluir los coeficientes de los parámetros  $\alpha_i^*$ ,  $\beta_j^*$  y  $(\alpha\beta)_{ij}$  o los coeficientes de los parámetros  $\tau_{ij}$ . Veamos un ejemplo.

**Ejemplo 3.1** En el ejemplo del experimento con las pilas (página 17) se ha realizado un gráfico de interacción y las líneas dibujadas distan de ser paralelas (véase Figura 3.3).

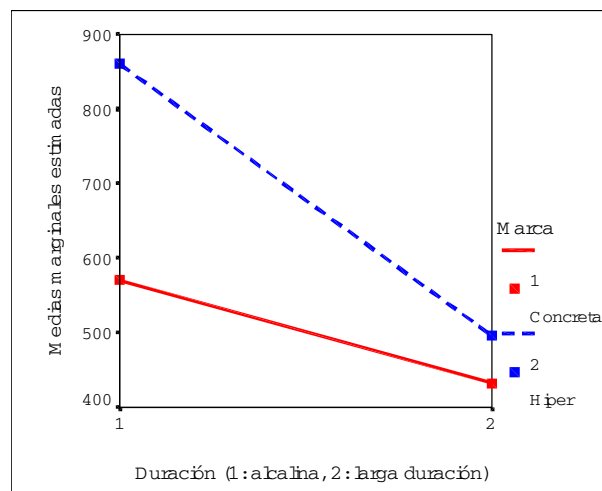


Figura 3.3: Gráfico de interacción para el ejemplo de las pilas.

Sin embargo la ausencia de paralelismo bien pudiera deberse a la propia variabilidad de los datos y no a un hipotético efecto interacción. Para estudiar más en detalle el problema se necesita analizar el contraste de interacción:

$$\tau_{11} - \tau_{12} - \tau_{21} + \tau_{22} = (\alpha\beta)_{11} - (\alpha\beta)_{12} - (\alpha\beta)_{21} + (\alpha\beta)_{22}$$

que puede representarse por la lista de coeficientes:  $[1, -1, -1, 1]$ .

El contraste que compara las duraciones medias de las pilas alcalinas y de larga duración es:

$$\bar{\tau}_2 - \bar{\tau}_1 = \frac{1}{2}(\tau_{21} + \tau_{22}) - \frac{1}{2}(\tau_{11} + \tau_{12}) = \alpha_2^* - \alpha_1^*$$

con lista de coeficientes  $[-1, 1]$ , en términos de los efectos  $\alpha_1^*$  y  $\alpha_2^*$ , o con lista  $\frac{1}{2}[-1, -1, 1, 1]$ , en términos de los efectos  $\tau_{11}$ ,  $\tau_{12}$ ,  $\tau_{21}$  y  $\tau_{22}$ .

Obsérvese que cada coeficiente del contraste de los efectos principales tiene a dos como divisor. En ocasiones el contraste de interacción también se divide por dos con la intención de que los correspondientes estimadores mínimo-cuadráticos de todos los contrastes tengan la misma varianza (ver la sección 3.5.1) y sus magnitudes sean más directamente comparables.

Los coeficientes de los contrastes a menudo se listan en una tabla como la Tabla 3.2.

| $ij$    | $A$ | $B$ | $AB$ |
|---------|-----|-----|------|
| 11      | -1  | -1  | 1    |
| 12      | -1  | 1   | -1   |
| 21      | 1   | -1  | -1   |
| 22      | 1   | 1   | 1    |
| Divisor | 2   | 2   | 2    |

Tabla 3.1: Listas de coeficientes de los contrastes de los efectos principales y de interacción para el ejemplo de las pilas

### 3.5 Análisis del Modelo Completo de Dos Vías

#### 3.5.1 Estimadores Mínimo-Cuadráticos

El estimador de mínimos cuadrados para  $\mu_{ij} = \mu + \tau_{ij}$  es el que minimiza la suma:

$$\min_{\mu_{ij}} \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^{n_{ij}} \overbrace{(Y_{ijt} - \mu_{ij})^2}^{\varepsilon_{ijt}^2}$$

Se obtiene:

$$\hat{\mu}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} = \frac{1}{n_{ij}} \sum_{t=1}^{n_{ij}} Y_{ijt} = \bar{Y}_{ij}. \tag{3.13}$$

de modo que

$$\hat{\mu}_{ij} = \bar{Y}_{ij} \sim N \left( \mu_{ij}, \frac{\sigma}{\sqrt{n_{ij}}} \right) \tag{3.14}$$

Estimadores mínimo cuadráticos de contrastes de interés asociados a este modelo son:

**Para contrastes de interacción:** Es decir, para funciones del tipo  $\sum_{i=1}^a \sum_{j=1}^b d_{ij} \tau_{ij}$  con  $\sum_{i=1}^a d_{ij} = 0$

para cada  $j$  y  $\sum_{j=1}^b d_{ij} = 0$  para cada  $i$ :

$$\boxed{\sum_{i=1}^a \sum_{j=1}^b d_{ij} \bar{Y}_{ij} \sim N \left( \sum_{i=1}^a \sum_{j=1}^b d_{ij} \tau_{ij}, \sigma \sqrt{\sum_{i=1}^a \sum_{j=1}^b \frac{d_{ij}^2}{n_{ij}}} \right)} \quad (3.15)$$

Por ejemplo, el estimador mínimo cuadrático del contraste de interacción

$$(\tau_{sh} - \tau_{uh}) - (\tau_{sq} - \tau_{uq})$$

viene dado por

$$\bar{Y}_{sh.} - \bar{Y}_{uh.} - \bar{Y}_{sq.} + \bar{Y}_{uq.}$$

con varianza

$$\sigma^2 \left( \frac{1}{n_{sh}} + \frac{1}{n_{uh}} + \frac{1}{n_{sq}} + \frac{1}{n_{uq}} \right)$$

**Para contrastes de efectos principales:** Es decir, para funciones del tipo  $\sum_{i=1}^a c_i \bar{\tau}_i$  con

$\sum_{i=1}^a c_i = 0$  y  $\sum_{j=1}^b k_j \bar{\tau}_{.j}$  con  $\sum_{j=1}^b k_j = 0$ , se obtiene respectivamente:

$$\boxed{\sum_{i=1}^a c_i \left( \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij} \right) \sim N \left( \sum_{i=1}^a c_i \bar{\tau}_i, \sigma \sqrt{\sum_{i=1}^a \sum_{j=1}^b \frac{c_i^2}{b^2 n_{ij}}} \right)} \quad (3.16)$$

$$\boxed{\sum_{j=1}^b k_j \left( \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij} \right) \sim N \left( \sum_{j=1}^b k_j \bar{\tau}_{.j}, \sigma \sqrt{\sum_{i=1}^a \sum_{j=1}^b \frac{k_j^2}{a^2 n_{ij}}} \right)} \quad (3.17)$$

Si los tamaños muestrales  $n_{ij}$  son iguales entonces las expresiones anteriores se simplifican notablemente:

$$\sum_{i=1}^a c_i \left( \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij} \right) = \sum_{i=1}^a c_i \bar{Y}_{i.} \quad \text{y} \quad \sum_{j=1}^b k_j \left( \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij} \right) = \sum_{j=1}^b k_j \bar{Y}_{.j}.$$

### 3.5.2 Estimación de la Varianza

De la sección previa se concluye que los residuos son:

$$\boxed{e_{ijt} = y_{ijt} - \hat{y}_{ijt} = y_{ijt} - \bar{y}_{ij}.} \quad (3.18)$$

verificándose que

$$\frac{scR}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^{n_{ij}} e_{ijt}^2 = \frac{1}{\sigma^2} \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^{n_{ij}} (y_{ijt} - \bar{y}_{ij.})^2 \sim \chi_{n-ab}^2$$

de modo que un estimador insesgado de  $\sigma^2$  es:

$$\boxed{scmR = \frac{scR}{n-ab} = \frac{1}{n-ab} \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^{n_{ij}} (y_{ijt} - \bar{y}_{ij.})^2} \quad (3.19)$$

### 3.5.3 Análisis de la Varianza para el Modelo Completo

Existen tres pruebas de hipótesis estándar que generalmente interesa resolver.

[1] Que la interacción entre los factores  $A$  y  $B$  no es significativa. O sea contrastar la hipótesis nula:

$$H_0^{AB} : (\alpha\beta)_{ij} - (\alpha\beta)_{iq} - (\alpha\beta)_{sj} + (\alpha\beta)_{sq} = 0 \text{ para todo } i \neq s, j \neq q$$

Obsérvese que si  $(\alpha\beta)_{ij} - (\alpha\beta)_{iq} - (\alpha\beta)_{sj} + (\alpha\beta)_{sq} = 0$  para todo  $i \neq s, j \neq q$ , entonces sus medias sobre  $s$  y  $q$  también se anulan. Luego es equivalente escribir:

$$H_0^{AB} : (\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} = 0 \text{ para todo } i, j$$

[2] y [3] Que no existen diferencias entre los efectos principales. Es decir:

$$H_0^A : \alpha_1^* = \alpha_2^* = \dots = \alpha_a^*, \text{ con } \begin{cases} \alpha_i^* = \alpha_i + \overline{(\alpha\beta)}_{i.} \\ \overline{(\alpha\beta)}_{i.} = \frac{1}{b} \sum_{j=1}^b (\alpha\beta)_{ij} \forall i \end{cases}$$

$$H_0^B : \beta_1^* = \beta_2^* = \dots = \beta_b^*, \text{ con } \begin{cases} \beta_j^* = \beta_j + \overline{(\alpha\beta)}_{.j} \\ \overline{(\alpha\beta)}_{.j} = \frac{1}{a} \sum_{i=1}^a (\alpha\beta)_{ij} \forall j \end{cases}$$

Deduciremos la resolución de estos contrastes suponiendo que los tamaños muestrales son iguales ( $n_{ij} = r, \forall ij$ ), dado que en otro caso no existen expresiones algebraicas claras y se tienen que aplicar correcciones. Obsérvese que con  $n_{ij} = r, \forall ij, n = abr$  y  $scmR = scR/ab(r-1)$ .

**Contrastando el efecto interacción (con tamaños muestrales iguales)** Si existe un efecto interacción significativo no resulta relevante contrastar la existencia de efectos principales distintos de cero. Por ello, la primera prueba de hipótesis a realizar es:

$$\left\{ \begin{array}{l} H_0^{AB} : (\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} = 0 \text{ para todo } i, j \\ H_1^{AB} : \text{Existen } i, j \text{ tales que } (\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} \neq 0 \end{array} \right.$$

Como en la sección 2.5.1 (página 29), la técnica del análisis de la varianza para resolver la prueba de hipótesis anterior se basa en comparar la suma de cuadrados residual bajo el modelo completo (dado por (3.2)),  $scR$ , con la suma de cuadrados residual bajo el modelo reducido que se obtiene cuando  $H_0^{AB}$  es verdad,  $scR_0^{AB}$ . A la diferencia entre ambas sumas de cuadrados:

$$\boxed{scAB = scR_0^{AB} - scR} \quad (3.20)$$

se la denomina **suma de cuadrados de la interacción** o **suma de cuadrados explicada por diferencias entre los efectos interacción**. La idea será rechazar  $H_0^{AB}$  cuando la cantidad obtenida en (3.20) sea “grande”.

A continuación se esquematiza como obtener estas sumas de cuadrados. Obsérvese que el modelo completo se puede reescribir en la forma:

$$\begin{aligned} y_{ijt} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijt} \\ &= \mu^* + \alpha_i^* + \beta_j^* + \left[ (\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} \right] + \varepsilon_{ijt} \end{aligned}$$

donde  $\mu^* = \mu - \overline{(\alpha\beta)}_{..}$ . Entonces, el modelo reducido puede escribirse:

$$y_{ijt} = \mu^* + \alpha_i^* + \beta_j^* + \varepsilon_{ijt}$$

que tiene el mismo formato que el modelo de dos vías de efectos principales. Tal y como se enuncia en la sección dedicada al análisis del modelo de dos vías de efectos principales, el estimador por mínimos cuadrados de  $\mu^* + \alpha_i^* + \beta_j^*$  es  $\bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}$  y, por tanto, la suma de cuadrados residual para el modelo reducido es:

$$scR_0^{AB} = \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r \left( y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2$$

Sumando y restando el término  $\bar{y}_{ij.}$ ,

$$\begin{aligned} scR_0^{AB} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r \left( (y_{ijt} - \bar{y}_{ij.}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) \right)^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{ij.})^2 + \sum_{i=1}^a \sum_{j=1}^b r (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \end{aligned}$$

De donde se concluye que:

$$scAB = scR_0^{AB} - scR = \sum_{i=1}^a \sum_{j=1}^b r \left( \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2 \quad (3.21)$$

Se verifica:

- Bajo  $H_0^{AB}$ :  $\left\{ \begin{array}{l} \text{(i)} \frac{SC(AB)}{\sigma^2} \sim \chi_{(a-1)(b-1)}^2 \\ \text{(ii)} \frac{SCM(AB)}{SCMR} = \frac{SC(AB)/(a-1)(b-1)}{SCR/ab(r-1)} \sim F_{(a-1)(b-1), ab(r-1)} \end{array} \right.$
- Por tanto:  $\text{Rechazar } H_0^{AB} \text{ al nivel } \alpha \text{ si } \hat{F} = \frac{scmAB}{scmR} > F_{(a-1)(b-1), ab(r-1), \alpha}$

**Contrastando los efectos principales de A (con tamaños muestrales iguales)** Al contrastar que el factor  $A$  no tiene influencia sobre la respuesta, uno puede pensar en contrastar que los niveles de  $A$  (promediados sobre los niveles de  $B$ ) tienen el mismo efecto sobre la respuesta, es decir:

$$H_0^A : \alpha_1^* = \alpha_2^* = \dots = \alpha_a^*$$

o bien en contrastar la hipótesis de que la respuesta depende tan sólo del nivel particular del factor  $B$  (solución que aportan muchos paquetes estadísticos), es decir,

$$H_0^{A+AB} : H_0^A \text{ y } H_0^{AB} \text{ son ambas correctas}$$

Nótese que son hipótesis distintas. Los modelos reducidos serían:

$$\text{Bajo } H_0^A : Y_{ijt} = \mu^{**} + \beta_j^* + \left[ (\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} \right] + \varepsilon_{ijt}$$

$$\text{Bajo } H_0^{A+AB} : Y_{ijt} = \mu^* + \beta_j^* + \varepsilon_{ijt}$$

Ahora bien, en este manual se ha sugerido contrastar los efectos principales sólo cuando  $H_0^{AB}$  sea cierta y, en tal caso, las dos hipótesis nulas son equivalentes y sus modelos reducidos iguales.

El estimador por mínimos cuadrados de  $\mu^* + \beta_j^*$  (o sea de la  $E(Y_{ijt})$  bajo  $H_0^A$ ) es  $\bar{y}_{ij.} - \bar{y}_{i..} + \bar{y}_{...}$  y, por tanto, la suma de cuadrados residual para el modelo reducido es:

$$\begin{aligned} scR_0^A &= \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r \left( y_{ijt} - \bar{y}_{ij.} + \bar{y}_{i..} - \bar{y}_{...} \right)^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r \left( y_{ijt} - \bar{y}_{ij.} \right)^2 + br \sum_{i=1}^a \left( \bar{y}_{i..} - \bar{y}_{...} \right)^2 \\ &= scR + br \sum_{i=1}^a \left( \bar{y}_{i..} - \bar{y}_{...} \right)^2 \end{aligned}$$

De donde se concluye que:

$$scA = scR_0^A - scR = br \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 \quad (3.22)$$

A la  $scA$  se la denomina **suma de cuadrados entre niveles del factor A** o **suma de cuadrados explicada por diferencias entre los niveles del factor A**.

Se verifica:

- Bajo  $H_0^A$ :  $\frac{SC(A)}{\sigma^2} \sim \chi_{(a-1)}^2$  y  $\frac{SCM(A)}{SCMR} \sim F_{(a-1), ab(r-1)}$

- Por tanto: Rechazar  $H_0^A$  al nivel  $\alpha$  si  $\hat{F} = \frac{scmA}{scmR} > F_{(a-1), ab(r-1), \alpha}$

**Contrastando los efectos principales de B (con tamaños muestrales iguales)** Análogos razonamientos son válidos para contrastar que el factor  $B$  no tiene influencia

$$H_0^B : \beta_1^* = \beta_2^* = \dots = \beta_b^*$$

Concluyendo que:

$$scB = scR_0^B - scR = ar \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 \quad (3.23)$$

Al  $scB$  se le denomina **suma de cuadrados entre niveles del factor B** o **suma de cuadrados explicada por diferencias entre los niveles del factor B** y se tiene

- Bajo  $H_0^B$ :  $\frac{SC(B)}{\sigma^2} \sim \chi_{(b-1)}^2$  y  $\frac{SCM(B)}{SCMR} \sim F_{(b-1), ab(r-1)}$

- Por tanto: Rechazar  $H_0^B$  al nivel  $\alpha$  si  $\hat{F} = \frac{scmB}{scmR} > F_{(b-1), ab(r-1), \alpha}$

### Descomposición de la Variabilidad

La suma de cuadrados global se puede descomponer de la forma:

$$\begin{array}{c}
 \text{Suma de} \\
 \text{Cuadrados} \\
 \text{Global (scG)} \\
 \underbrace{\sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{...})^2}_{g.l. = abr-1} = \underbrace{br \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2}_{g.l. = a-1} + \underbrace{ar \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2}_{g.l. = b-1} \\
 \\
 \underbrace{r \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2}_{g.l. = (a-1)(b-1)} + \underbrace{\sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{ij.})^2}_{g.l. = ab(r-1)} \\
 \text{Suma de} \\
 \text{Cuadrados} \\
 \text{Explicada por AB (scAB)} \\
 \text{Suma de} \\
 \text{Cuadrados} \\
 \text{Residual (scR)}
 \end{array}$$

Téngase en cuenta que cuando los tamaños muestrales no son iguales, las variables aleatorias  $SC(A)$ ,  $SC(B)$  y  $SC(AB)$  no son independientes, las expresiones para el cálculo de  $scA$ ,  $scB$  y  $scAB$  son más complicadas y  $scA + scB + scAB + scR \neq scG$ .

### 3.6 Análisis del Modelo de Efectos Principales de Dos Vías

#### 3.6.1 Estimadores Mínimo-Cuadráticos

El modelo de efectos principales es un submodelo del modelo completo en el sentido de que solamente puede describir situaciones similares a las descritas en los gráficos de la Figura 3.1, pero nunca a los gráficos de la Figura 3.2. Otra vez los estimadores mínimo cuadráticos de los distintos parámetros del modelo no son fáciles de obtener cuando los tamaños muestrales no son iguales. Por ello, nos volveremos a centrar en el supuesto  $n_{ij} = r \forall ij$ .

El estimador de mínimos cuadrados para  $E(Y_{itj}) = \mu_{ij} = \mu + \alpha_i + \beta_j$  es:

$$\boxed{\hat{\mu}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...}} \tag{3.24}$$

El estimador de mínimos cuadrados del contraste para los efectos del factor  $A$ , es decir, para funciones del tipo  $\sum_{i=1}^a c_i \alpha_i$  con  $\sum_{i=1}^a c_i = 0$ , viene dado por:

$$\sum_{i=1}^a c_i \hat{\alpha}_i = \sum_{i=1}^a c_i (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j) = \sum_{i=1}^a c_i (\bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...}) \Rightarrow$$

| <b>CUADRO DEL ANÁLISIS DE LA VARIANZA</b><br>— MODELO COMPLETO DE DOS VÍAS —  |  |                  |                                       |                      |
|---|--|------------------|---------------------------------------|----------------------|
| <i>Fuente de Variación</i>  | <i>Suma de Cuadrados</i>   | <i>g.l.</i>      | <i>scm</i>                            | $\hat{F}$            |
| Factor <i>A</i>   | $scA = br \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2$  | $a - 1$          | $scmA = \frac{scA}{a - 1}$            | $\frac{scmA}{scmR}$  |
| Factor <i>B</i>   | $scB = ar \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2$  | $b - 1$          | $scmB = \frac{scB}{b - 1}$            | $\frac{scmB}{scmR}$  |
| Inter. <i>AB</i>  | $scAB = r \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$ | $(a - 1)(b - 1)$ | $scmAB = \frac{scAB}{(a - 1)(b - 1)}$ | $\frac{scmAB}{scmR}$ |
| Residual  | $scR = \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{ij.})^2$                             | $ab(r - 1)$      | $scmR = \frac{scR}{ab(r - 1)}$        |                      |
| Global  | $scG = \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{...})^2$                             | $abr - 1$        | $scmG = \frac{scG}{abr - 1}$          |                      |
| Rechazar $H_0^{AB} : (\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} = 0 \forall i, j$<br><br>en base al <i>p</i> -valor $p = P\left(\frac{scmAB}{scR} > F_{(a-1)(b-1), ab(r-1)}\right)$   |  |                  |                                       |                      |
| Si no se rechaza $H_0^{AB}$ entonces<br><br>Rechazar $H_0^A : \alpha_1^* = \alpha_2^* = \dots = \alpha_a^*$ , según $p = P\left(\frac{scmA}{scmR} > F_{a-1, ab(r-1)}\right)$<br><br>Rechazar $H_0^B : \beta_1^* = \beta_2^* = \dots = \beta_b^*$ , según $p = P\left(\frac{scmB}{scmR} > F_{b-1, ab(r-1)}\right)$ |  |                  |                                       |                      |
| Si se rechaza $H_0^{AB}$ entonces considerar el modelo de una vía: $Y_{ijt} = \mu_{ij} + \varepsilon_{ijt}$   |  |                  |                                       |                      |

Tabla 3.3: Cuadro del análisis de la varianza para un diseño completamente aleatorizado y balanceado de dos factores de efectos fijos (modelo completo).

$$\boxed{\sum_{i=1}^a c_i \hat{\alpha}_i = \sum_{i=1}^a c_i \bar{Y}_{i..} \sim N \left( \sum_{i=1}^a c_i \alpha_i, \sigma \sqrt{\sum_{i=1}^a \frac{c_i^2}{rb}} \right)} \quad (3.25)$$

Por ejemplo, para la comparación apareada de los niveles  $p$  y  $s$  del factor  $A$ ,  $\alpha_p - \alpha_s$ , se obtiene el estimador:

$$\hat{\alpha}_p - \hat{\alpha}_s = \bar{Y}_{p..} - \bar{Y}_{s..} \quad \text{Var}(\bar{Y}_{p..} - \bar{Y}_{s..}) = \frac{2\sigma^2}{rb}$$

En igual forma se muestra que el estimador de mínimos cuadrados del contraste para los efectos del factor  $B$  es:

$$\boxed{\sum_{j=1}^b k_j \hat{\beta}_j = \sum_{j=1}^b k_j \bar{Y}_{.j.} \sim N \left( \sum_{j=1}^b k_j \beta_j, \sigma \sqrt{\sum_{j=1}^b \frac{k_j^2}{ra}} \right)} \quad (3.26)$$

### 3.6.2 Estimación de la Varianza

En el modelo de efectos principales, con  $n_{ij} = r$ , para todo par  $ij$ , los residuos son:

$$\boxed{e_{ijt} = y_{ijt} - \hat{y}_{ijt} = y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}} \quad (3.27)$$

Y un estimador insesgado de  $\sigma^2$  es:

$$\boxed{scmR = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r e_{ijt}^2}{abr - a - b + 1} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2}{abr - a - b + 1}} \quad (3.28)$$

### 3.6.3 Análisis de la Varianza para el Modelo de Dos Vías de Efectos Principales

**Contrastando los efectos principales de  $A$  (con tamaños muestrales iguales)** Interesa examinar la hipótesis nula:

$$H_0^A : \alpha_1 = \alpha_2 = \dots = \alpha_a$$

frente la alternativa de existir al menos dos niveles  $s$  y  $t$  tales que  $\alpha_s \neq \alpha_t$ .

El modelo reducido bajo  $H_0^A$  es:

$$Y_{ijt} = \mu^* + \beta_j + \varepsilon_{ijt}, \quad \text{con } \mu^* = \mu + \bar{\alpha}.$$

que es idéntico al modelo completo del análisis de la varianza de una vía. En otros términos, la  $scR_0^A$  coincide con la suma de cuadrados residual del Anova de una vía, con la salvedad de que el subíndice  $j$  debe reemplazarse aquí por el par de subíndices  $jt$ , resultando:

$$scR_0^A = \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{.j.})^2$$

de donde se deduce que la suma de cuadrados residual para testar  $H_0^A$  es:

$$scA = scR_0^A - scR = \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{.j.})^2 - \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r ((y_{ijt} - \bar{y}_{.j.}) - (\bar{y}_{i..} - \bar{y}_{...}))^2 \Rightarrow$$

$$scA = br \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 \tag{3.29}$$

Se verifica:

- Bajo  $H_0^A$ :  $\frac{SC(A)}{\sigma^2} \sim \chi_{(a-1)}^2$  y  $\frac{SCM(A)}{SCMR} \sim F_{(a-1),abr-a-b+1}$

- Por tanto: Rechazar  $H_0^A$  al nivel  $\alpha$  si  $\hat{F} = \frac{scmA}{scmR} > F_{(a-1),abr-a-b+1,\alpha}$

### Descomposición de la Variabilidad

La suma de cuadrados global se puede descomponer de la forma:

$$\underbrace{\sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{...})^2}_{\text{Suma de Cuadrados Global (scG)} \quad g.l. = abr-1} = \underbrace{br \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2}_{\text{Suma de Cuadrados Explicada por A (scA)} \quad g.l. = a-1} + \underbrace{ar \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2}_{\text{Suma de Cuadrados Explicada por B (scB)} \quad g.l. = b-1} + \underbrace{\sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2}_{\text{Suma de Cuadrados Residual (scR)} \quad g.l. = abr-a-b+1}$$

| CUADRO DEL ANÁLISIS DE LA VARIANZA<br>— MODELO DE EFECTOS PRINCIPALES DE DOS VÍAS — |  |                   |                                      |                     |
|---|--|-------------------|--------------------------------------|---------------------|
| Fuente de Variación   | Suma de Cuadrados  | <i>g.l.</i>       | <i>scm</i>                           | $\hat{F}$           |
| Factor A  | $scA = br \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2$  | $a - 1$           | $scmA = \frac{scA}{a - 1}$           | $\frac{scmA}{scmR}$ |
| Factor B  | $scB = ar \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2$  | $b - 1$           | $scmB = \frac{scB}{b - 1}$           | $\frac{scmB}{scmR}$ |
| Residual  | $scR = \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$ | $abr - a - b + 1$ | $scmR = \frac{scR}{abr - a - b + 1}$ |                     |
| Global  | $scG = \sum_{i=1}^a \sum_{j=1}^b \sum_{t=1}^r (y_{ijt} - \bar{y}_{...})^2$                                 | $abr - 1$         | $scmG = \frac{scG}{abr - 1}$         |                     |

Rechazar  $H_0^A : \alpha_1 = \alpha_2 = \dots = \alpha_a$ , según  $p = P\left(\frac{scmA}{scmR} > F_{a-1, abr-a-b+1}\right)$

Rechazar  $H_0^B : \beta_1 = \beta_2 = \dots = \beta_b$ , según  $p = P\left(\frac{scmB}{scmR} > F_{b-1, abr-a-b+1}\right)$

Tabla 3.4: Cuadro del análisis de la varianza para un diseño completamente aleatorizado y balanceado de dos factores de efectos fijos (modelo de efectos principales).

### 3.7 Comparaciones Múltiples

En el modelo completo, el planteamiento de las comparaciones múltiples es idéntico al realizado con el diseño de una vía completamente aleatorizado. Así, los métodos de Bonferroni, Scheffé, Tukey y Dunnett son igualmente válidos pero ahora con  $I = ab$  tratamientos (los tratamientos cruzados de cada factor tratamiento  $A$  y  $B$ ) y, naturalmente, con la suma de cuadrados residual,  $scmR$ , dada por (3.19).

Es conveniente considerar este planteamiento ya que la mejor combinación de tratamientos no tiene porque coincidir con la combinación de los aparentemente mejores niveles de  $A$  y  $B$ . En el gráfico (d) de la Figura 3.2 se observa que los mejores resultados los obtiene el profesor 2 con el método de enseñanza 1 mientras que, en media, el mejor profesor es el 3.

En el modelo completo carece de interés, por lo general, realizar comparaciones múltiples

relativas a contrastes de efectos principales. Desde luego las conclusiones serán de difícil interpretación si existe interacción significativa.

En otro caso, los métodos estudiados (Bonferroni, Tukey, ...) se aplican sobre los niveles de cada factor por separado utilizando la varianza residual del modelo completo.

En el modelo de efectos principales se procede como en este último caso pero empleando la correspondiente varianza residual (expresión (3.28)).

### 3.8 Tamaños Muestrales

En la sección 2.6 se desarrolló un algoritmo que permite calcular el tamaño muestral común  $r$  con que muestrear todos los niveles de un factor tratamiento, en orden a realizar el test Anova en el modelo de una vía con una potencia predeterminada  $\Pi(\Delta)$ , dados  $\Delta$  (mínima diferencia que se desea detectar),  $I$  (número de niveles del tratamiento),  $\alpha$  (nivel de significación) y  $\sigma^2$ . Esencialmente el procedimiento calculaba  $r$  conforme a:

$$r = \frac{2I\sigma^2\phi^2}{\Delta^2}$$

El algoritmo es válido también para el modelo completo sin más que reemplazar  $I$  por  $ab$  y, consecuentemente, los grados de libertad en el algoritmo son ahora  $\nu_1 = ab - 1$  y  $\nu_2 = ab(r - 1)$ .

Más aún, si el interés se centra en los tamaños muestrales precisos para obtener una potencia preespecificada en las pruebas de efectos principales, entonces las modificaciones son:

$$r_A = \frac{2a\sigma^2\phi^2}{b\Delta_A^2} \quad (3.30)$$

para el factor  $A$ . O sea, muestrear cada nivel del factor tratamiento  $A$  con  $r_A$  observaciones si se desea detectar, con un nivel de significación y potencias dadas, una diferencia mínima  $\Delta_A$  entre los efectos  $\alpha_i$  (o entre los  $\alpha_i^*$  en el modelo completo).

Análogamente

$$r_B = \frac{2b\sigma^2\phi^2}{a\Delta_B^2}$$

para el factor  $B$ .

Como antes, los grados de libertad son  $\nu_1 = a - 1$ , para el factor  $A$ ,  $\nu_1 = b - 1$ , para el factor  $B$ , y  $\nu_2 = ab(r - 1)$  para ambos factores si el modelo es el modelo completo o  $\nu_2 = abr - a - b + 1$  para ambos factores si el modelo es el modelo de efectos principales. En caso de que  $r_A \neq r_B$  debería seleccionarse el mayor valor.

### 3.9 Experimentos Pequeños

Muchas veces y por distintos motivos sólo es posible obtener una observación por celda ( $r = 1$ ). Obviamente, perdida la capacidad de seleccionar tamaños muestrales, los contrastes pueden resultar menos potentes y los intervalos de confianza más amplios de lo deseado.

El primer problema serio se plantea cuando se pretende utilizar un modelo completo. En tal caso, al necesitar estimar los efectos interacción, el número de grados de libertad del error resulta  $ab(r - 1) = 0$  y, en consecuencia,  $\sigma^2$  no puede estimarse.

En un diseño con una observación por celda se puede plantear un modelo completo que permita contrastar el efecto interacción únicamente en alguna de las tres siguientes situaciones:

1. se conoce previamente  $\sigma^2$ ,
2. el efecto interacción puede ser adecuadamente modelizado con menos de  $(a - 1)(b - 1)$  grados de libertad,
3. a pesar de que el número de tratamientos es grande, solo interesan un número pequeño de contrastes.

La situación 2 puede abordarse mediante un análisis con contrastes ortogonales tal y como se describe en la siguiente sección.

#### 3.9.1 Análisis Basado en Contrastes Ortogonales

En el apartado de *Análisis basado en contrastes ortogonales* de la sección 2.7.2, dos contrastes se denominaron ortogonales si y solo si sus respectivos estimadores mínimo-cuadráticos eran incorrelados (ver (2.25)).

Asimismo se estableció que, dados  $I$  tratamientos, existe un máximo de  $I - 1$  contrastes mutuamente ortogonales (que se llama **conjunto completo de contrastes ortogonales**), de tal modo que, por la independencia de los respectivos estimadores, la suma de cuadrados de un contraste puede descomponerse en la suma de las sumas de cuadrados de cada uno de los contrastes de un conjunto completo.

Esta idea puede aplicarse a los contrastes de interacción rompiendo la suma de cuadrados  $scAB$  en las  $(a - 1)(b - 1)$  sumas de cuadrados correspondientes a cada contraste particular de interacción:

$$scAB = \sum_{h=1}^{(a-1)(b-1)} scI(h)$$

donde  $scI(h)$  representa la suma de cuadrados de la interacción concreta  $h$ :

$$\sum_{i=1}^a \sum_{j=1}^b d_{ij}^{(h)} (\alpha\beta)_{ij} \text{ con } \sum_{i=1}^a d_{ij}^{(h)} = 0 \text{ para cada } j \text{ y } \sum_{j=1}^b d_{ij}^{(h)} = 0 \text{ para cada } i$$

definida por:

$$scI(h) = \frac{\left( \sum_{i=1}^a \sum_{j=1}^b d_{ij}^{(h)} y_{ij} \right)^2}{\sum_{i=1}^a \sum_{j=1}^b d_{ij}^{(h)2}}$$

Así, cuando existe un único dato por celda y se *sospecha* que  $e$  contrastes ortogonales de interacción específicos son probablemente no significativos, las sumas de cuadrados correspondientes a estos  $e$  contrastes se pueden agrupar para obtener un estimador de la varianza del error basado en  $e$  grados de libertad:

$$scR = \sum_{h=1}^e scI(h) \quad \text{y} \quad scmR = scR/e$$

utilizando a continuación este estimador para contrastar la existencia de interacción: se rechaza la hipótesis  $H_0^{AB}$  a un nivel  $\alpha$  siempre que:

$$\frac{scAB / ((a-1)(b-1) - e)}{scR/e} > F_{(a-1)(b-1)-e, e, \alpha}$$

$$\text{con } scAB = \sum_{h=e+1}^{(a-1)(b-1)} scI(h)$$

### 3.10 Comprobación de la Idoneidad del Modelo

Las hipótesis implícitas tanto en el modelo completo como en el modelo de efectos principales son que las variables aleatorias error tienen igual varianza, son mutuamente independientes y normalmente distribuidas. La estrategia y los métodos de chequeo de los residuos estudiadas en el anterior capítulo son igualmente válidas aquí.

Ahora los residuos son:

$$e_{ijt} = y_{ijt} - \hat{y}_{ijt}$$

con

$$\hat{y}_{ijt} = \hat{\tau}_{ij} = \hat{\alpha}_i + \hat{\beta}_j + (\widehat{\alpha\beta})_{ij} = \bar{y}_{ij}.$$

para el modelo completo, y

$$\hat{y}_{ijt} = \hat{\tau}_{ij} = \hat{\alpha}_i + \hat{\beta}_j = y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

para el modelo de efectos principales.

Y, los residuos estandarizados se definen como:

$$e_{ijt}^* = \frac{e_{ijt}}{\sqrt{scmR}}$$

con la  $scmR$  dependiendo del modelo que se trate.

### 3.11 Problemas Propuestos

**Problema 1** (Liming Cai, Tong Li, Nishant y Andre van der Kouwe, 1996) Se realiza un experimento para comparar los tiempos de reacción de un sujeto ante estímulos audibles y visuales. Se presenta el estímulo a través de un ordenador personal y se pide al sujeto que pulse una tecla cualquiera cuando lo reciba, midiendo el tiempo que transcurre entre ambas acciones. Además, al sujeto se le avisaba con antelación, también mediante una entrada audible o visual, de que el estímulo iba a tener lugar. Se consideraron dos factores de interés: el *factor A* era el tipo de estímulo y sus niveles *audible* o *visual*, codificados respectivamente como 1 y 2; y el *factor B* era el tiempo que transcurría entre las apariciones del aviso y del estímulo, tiempo controlado por los experimentadores y que se fijó en tres niveles: 5, 10 y 15 segundos, codificados de 1 a 3. Se tomaron tres observaciones a un mismo sujeto para cada combinación de tratamientos. Las observaciones se recogen en el archivo de datos ESTIMULOS.SAV.

El análisis preplanificado para este experimento consistió en ajustar un modelo completo de dos vías completamente aleatorizado y realizar la prueba de hipótesis de no interacción. Si ésta se rechazase, el contraste preplanificado de interés consistiría en realizar las comparaciones apareadas de todas las combinaciones de tratamientos mediante el procedimiento de Tukey a un 1% de significación. En otro caso se emplearía este procedimiento, también con un 1% de significación, para comparar por separado los niveles de cada factor dos a dos. Realizar el análisis estadístico del experimento y extraer conclusiones de acuerdo a las intenciones de los experimentadores.

En realidad éste fue un experimento piloto. ¿Qué modelo matemático deberán usar para el experimento completo?

**Problema 2** (Pascale Quester, 1986) Se investiga la eficacia de dos diferentes marcas de disolventes de esmalte de uñas (*factor A*) para eliminar de la ropa manchas producidas por esmaltes de este tipo. Además de las dos marcas de disolvente se seleccionaron tres marcas diferentes de esmaltes (*factor B*). El experimento consistió en dejar caer, desde una altura de 20 cm, una gota de esmalte sobre una pieza de ropa y, a continuación, se sumergió la pieza de ropa en un balde con disolvente y se midió el tiempo (en minutos) que tardó en disolverse completamente el esmalte. El diseño se realizó completamente aleatorizado y tomando cinco observaciones sobre cada combinación de tratamientos. Los datos figuran en el archivo DISOLVENTE.SAV.

Un experimento piloto realizado previamente había sugerido que el efecto interacción entre marca de esmalte y marca de disolvente no era significativo y que el experimento producía resultados satisfactorios. Realizar el análisis estadístico del experimento con un modelo de dos vías de efectos principales.

**Problema 3** Un ingeniero diseña una batería para usar en un dispositivo que será sometido a variaciones extremas de temperatura. En primera instancia, el único parámetro del diseño que él puede seleccionar es el material de la cubierta de la batería, para lo cual dispone de tres alternativas. Cuando el dispositivo se manufactura y se envía al campo, el ingeniero no tiene control sobre los extremos de temperatura a que será expuesto, si bien experiencias previas le hacen sospechar que la temperatura influye en la duración efectiva de la batería. Sin embargo, sí es posible controlar la temperatura en el laboratorio de desarrollo de productos para los fines de ensayo.

El ingeniero decide probar los tres materiales de la cubierta a tres niveles distintos de temperatura (15, 70, y 125  $F$ ). Se contabilizan las horas de duración de cuatro baterías por cada combinación de tratamientos. Los resultados figuran en el archivo BATERIAS.SAV.

1. ¿Qué efecto tienen el tipo de material y la temperatura sobre la duración de la batería?
2. ¿Existe una elección del material que proporcione una duración uniformemente larga, sin importar la temperatura?

**Problema 4** Se están investigando los efectos que sobre la resistencia del papel producen el porcentaje de concentración de fibra de madera (hardwood) en la pulpa, la presión del tanque y el tiempo de cocción de la pulpa. Se seleccionan tres niveles de concentración de madera y de presión, y dos niveles de tiempo de cocción. Se realiza un experimento completamente aleatorizado con dos réplicas y se registran los datos que se recogen en el archivo RESISTENCIA.SAV.

1. Analizar los datos y obtener conclusiones.
2. Obtener los gráficos de residuos apropiados y comentar la adecuación del modelo.

**Problema 5** Se realizó un estudio para determinar si las condiciones de humedad tienen algún efecto sobre la fuerza que se requiere para romper piezas de plástico engomado. Se probaron tres tipos de plástico ( $A$ ,  $B$  y  $C$ ) utilizando cuatro diferentes niveles de humedad (30%, 50%, 70% y 90%). Los resultados figuran en el archivo de datos PLASTICO.SAV.

1. Realizar el análisis estadístico del experimento suponiendo que se trata de un modelo de efectos fijos. Contrastar la hipótesis de no interacción entre humedad y tipo de plástico al 5% de significación.
2. Utilizando únicamente los plásticos  $A$  y  $B$  y las estimaciones de la varianza del apartado (1), contrastar de nuevo la presencia de la interacción entre humedad y tipo de plástico.
3. Utilizar una comparación de un solo grado de libertad y el valor de la  $scmR$  del primer apartado para comparar, al nivel de significación del 5%, la fuerza que se requiere a la humedad del 30% contra las humedades del 50%, 70% y 90%.

4. Utilizando únicamente el plástico  $C$  y el valor de la  $scmR$  del primer apartado repetir el apartado anterior.



## Capítulo 4

# Análisis de la Covarianza

### 4.1 Introducción

En los capítulos 2 y 3 se han estudiado diseños completamente aleatorizados y la técnica del análisis de la varianza para comparar los efectos sobre la respuesta de uno o más factores tratamiento. En los modelos expuestos hasta aquí no se contemplan más fuentes de variabilidad planificada que las debidas a los tratamientos. Sin embargo, en línea con lo argumentado en el capítulo 1, cuando el experimentador es consciente de la existencia de factores “nuisance” que suponen una fuente de variación muy importante, debe tratar de controlarlos mediante su planificación en el experimento. Si previamente a la realización del experimento los niveles de dichos factores son susceptibles de ser medidos y seleccionados, entonces podría controlarse su efecto a través de factores bloque (que se verán en el capítulo 5). El problema es que no siempre es posible controlar y seleccionar los niveles del factor nuisance. Incluso a veces, sólo es posible medirlo una vez el experimento ya está en marcha. En tal caso el factor se denomina **covariable**.

Por ejemplo, supóngase que se desea comparar el efecto de varias dietas sobre la ganancia en peso de cochinitos de un mes. La respuesta (peso al final del período experimental) dependerá probablemente del peso en el inicio del período experimental y, a su vez, estos pesos serán tremendamente variables. Una posibilidad de controlar esta variabilidad inicial en el experimento es usar un diseño en bloques. Es decir, distribuir los cochinitos por grupos de pesos iniciales semejantes (bloques) y comparar después los efectos de las dietas dentro de los bloques. Otra posibilidad es utilizar un diseño completamente aleatorizado con respuesta la ganancia en peso a través del período experimental. Esta segunda alternativa conlleva perder información. Por ejemplo es muy probable que cochinitos con pesos iniciales grandes experimenten una ganancia en peso mayor que los cochinitos pequeños y viceversa. Sería entonces adecuado introducir en el modelo el peso inicial como una nueva variable (covariable) que ayude a explicar el peso final.

En las secciones siguientes se estudia el modelo matemático apropiado para un diseño completamente aleatorizado que incluye uno o más factores tratamiento y una o más covariables.

## 4.2 Modelos Matemáticos

Consideremos un experimento planificado de acuerdo a un diseño completamente aleatorizado, con la intención de comparar los efectos de  $I$  niveles de un factor tratamiento de interés sobre una variable respuesta  $Y$ . Supongamos que  $Y$  depende del valor  $x$  de una covariable que puede medirse antes o durante el experimento. Supongamos además que existe una relación lineal entre  $E(Y)$  y  $x$  con la misma pendiente para cada tratamiento.

El modelo matemático adecuado para representar la situación experimental descrita es:

$$\underbrace{Y_{it}}_{\text{aleatorio}} = \overbrace{\mu + \tau_i + \beta x_{it}}^{\text{determinista}} + \underbrace{\varepsilon_{it}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{it} \text{ i.i.d. según una } N(0, \sigma) \quad (4.1)$$

para  $i = 1, \dots, I$  y  $t = 1, \dots, n_i$ . En (4.1),  $\mu$  es una constante,  $\tau_i$  representa la desviación, positiva o negativa, de esta constante cuando se observa el tratamiento  $i$  (podría ser sustituido por los parámetros de la interacción y de los efectos principales caso de existir más de un factor tratamiento) y, finalmente,  $\beta x_{it}$  modeliza la relación lineal existente entre la respuesta y la covariable ( $x_{it}$  representa el valor de la covariable en la  $t$ -ésima observación del nivel  $i$  del factor tratamiento).

Para el análisis posterior del modelo es muy importante que los valores de la covariable no dependan del nivel del factor tratamiento. Si así fuese carecería de interpretación la comparación de los efectos de los tratamientos en un valor  $x$  común de la covariable.

Una forma alternativa muy habitual de escribir el modelo (4.1) es:

$$\boxed{Y_{it} = \mu^* + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \varepsilon_{it}} \quad (4.2)$$

en el cual los valores de la covariable han sido centrados. Los dos modelos son equivalentes: la pendiente  $\beta$  tiene la misma interpretación en ambos,  $\mu + \tau_i$  denota la respuesta media cuando  $x_{it} = 0$  en (4.1) y  $\mu^* + \tau_i$ , con  $\mu^* = \mu + \beta\bar{x}_{..}$ , denota la respuesta media cuando  $x_{it} = \bar{x}_{..}$  en (4.2). En lo que sigue nos centraremos únicamente en el formato (4.2) pero refiriéndonos a la constante del modelo ( $\mu^*$ ) con la notación habitual:  $\mu$ .

El modelo anterior puede generalizarse en varias direcciones. Una primera extensión sería considerar que el efecto de la covariable sobre la respuesta fuese no lineal, por ejemplo, polinómico. Entonces se reemplazaría el término  $\beta x_{it}$  en (4.1) por una función polinómica de orden  $p$ ,  $\beta_1 x_{it} + \beta_2 x_{it}^2 + \dots + \beta_p x_{it}^p$ , que tratara de modelizar adecuadamente la forma común de la curva respuesta en cada tratamiento. Este nuevo modelo no será abordado en esta memoria, si bien su análisis es fácilmente generalizable siguiendo los fundamentos del análisis de regresión polinómica.

En el supuesto de que exista más de una covariable, otra posible extensión del modelo es reemplazar  $\beta x_{it}$  en (4.1) por una función polinómica apropiada de todas las covariables. Por ejemplo, para dos covariables  $x_1$  y  $x_2$ , podría emplearse una función de segundo orden del tipo:

$$\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

El análisis de este tipo de modelos puede encuadrarse en el estudio de modelos de **superficies de respuesta**.

### 4.3 Estimadores Mínimo-Cuadráticos

Obsérvese que bajo el modelo (4.2) el valor esperado de la media muestral de la respuesta en el tratamiento  $i$ :

$$E[\bar{Y}_i] = \mu + \tau_i + \beta(\bar{x}_i - \bar{x}_{..})$$

es un estimador de la respuesta media del tratamiento  $i$ -ésimo **cuando el valor de la covariable  $x$  es  $\bar{x}_i$** . O sea que, por ejemplo, la diferencia  $\bar{y}_s - \bar{y}_r$  no estima el valor de  $\tau_s - \tau_r$  (a menos claro está que todas las  $\bar{x}_i$  hayan resultado ser iguales) y por tanto no debe utilizarse para comparar los efectos de los tratamientos.

Esta consideración previa permite ya intuir que, para este modelo, las medias muestrales no serán los estimadores mínimo-cuadráticos. En efecto, se prueba que los valores  $\hat{\mu} + \hat{\tau}_i$  que minimizan la suma de los cuadrados de los errores:

$$\sum_{i=1}^I \sum_{t=1}^{n_i} e_{it}^2 = \sum_{i=1}^I \sum_{t=1}^{n_i} [y_{it} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}(x_{it} - \bar{x}_{..})]^2$$

vienen dados por:

$$\boxed{\hat{\mu} + \hat{\tau}_i = \bar{y}_i - \hat{\beta}(\bar{x}_i - \bar{x}_{..}), \quad i = 1, \dots, I} \quad (4.3)$$

con

$$\hat{\beta} = \frac{sp_{xy}^*}{sc_{xx}^*} \quad (4.4)$$

$$sp_{xy}^* = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) \quad (4.5)$$

$$sc_{xx}^* = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_i)^2 \quad (4.6)$$

Por consiguiente, los estimadores por mínimos cuadrados  $\hat{\mu} + \hat{\tau}_i$  estiman la respuesta media del tratamiento  $i$ -ésimo en el valor de la covariable igual a  $\bar{x}_i$ . A los estimadores  $\hat{\mu} + \hat{\tau}_i$  se

les conoce como **medias ajustadas**, dado que corrigen la respuesta media  $\bar{y}_i$  por la cantidad  $\hat{\beta}(\bar{x}_i - \bar{x}_{..})$ , lo que equivale a medir la respuesta con el mismo valor de la covariable. Por supuesto ahora sí se verifica:  $E(\hat{\mu} + \hat{\tau}_i) = \mu + \tau_i$ .

El siguiente ejemplo justifica la necesidad de utilizar las medias ajustadas.

**Ejemplo 4.1** Supóngase que el gráfico de dispersión de la Figura 4.1 representa a un conjunto de observaciones procedentes de un modelo como el (4.2) con un factor tratamiento con dos niveles y una covariable. Las doce observaciones con valor de la covariable 20, 30, 40 o 50 proceden del primer tratamiento y las doce restantes, valores de la covariable igual a 70, 80, 90 o 100, del segundo.

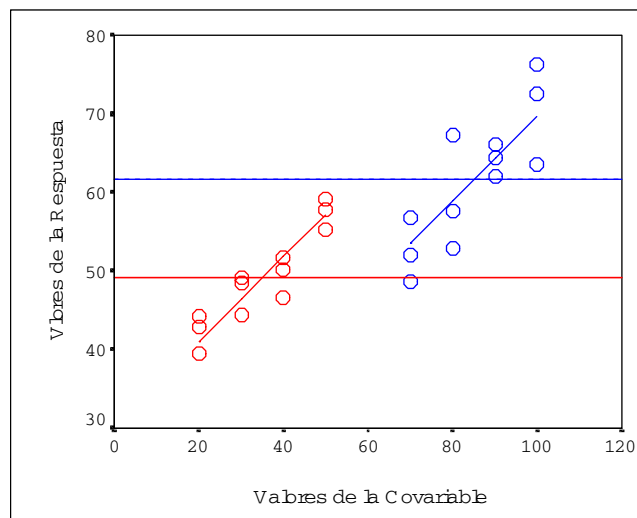


Figura 4.1: Sesgo en medias no ajustadas.

Usando las ecuaciones (4.3) a (4.6) se concluye que el modelo ajustado es:

$$\hat{y}_{it} = \hat{\mu} + \hat{\tau}_i + 0'5372(x_{it} - 60)$$

donde  $\hat{\mu} + \hat{\tau}_1 = 62'5416$  y  $\hat{\mu} + \hat{\tau}_2 = 48'2516$ .

Por tanto el tratamiento 1 es el que tiene un mayor efecto sobre la respuesta (en la Figura 4.1 esto se traduce en que de las dos líneas paralelas ajustadas, la que va por encima es la del tratamiento 1).

Esto contradice a las medias sin ajustar, ya que:

$$\bar{y}_1 = 49'11 < 61'68 = \bar{y}_2.$$

Naturalmente la contradicción radica en que las medias sin ajustar están sesgadas por los valores relativos de  $\bar{x}_1$  y  $\bar{x}_2$ .

## 4.4 Análisis de la Covarianza

Para un diseño completamente aleatorizado y el modelo (4.2), la técnica habitual del análisis de la varianza se conoce con el nombre de **análisis de la covarianza**. Como en los capítulos previos, el estadístico de prueba para resolver:

$$\begin{cases} H_0 : \tau_1 = \tau_2 = \dots = \tau_I \\ H_1 : \text{Existen al menos dos de los } \tau_i \text{'s que difieren} \end{cases}$$

se basa en comparar las sumas de cuadrados de los errores bajo los modelos completo (aquel que se tiene cuando  $H_0$  no es cierta) y reducido (aquel que se tiene bajo  $H_0$ ).

El modelo reducido es:

$$Y_{it} = \mu + \tau + \beta(x_{it} - \bar{x}_{..}) + \varepsilon_{it} \quad (4.7)$$

Se observa que (4.7) responde a la expresión clásica de un modelo de regresión lineal simple, con constante  $\beta_0 = \mu + \tau$ , pendiente  $\beta_1 = \beta$  y variable regresora  $x_{it}$  centrada. Por tanto, los estimadores mínimo-cuadráticos de los parámetros del modelo reducido (4.7) son:

$$\hat{\mu} + \hat{\tau} = \bar{y}_{..} \quad \text{y} \quad \hat{\beta} = \frac{sp_{xy}}{sc_{xx}} \quad (4.8)$$

con

$$sp_{xy} = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_{..}) y_{it} = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_{..}) (y_{it} - \bar{y}_{..}) \quad (4.9)$$

$$sc_{xx} = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_{..})^2 \quad (4.10)$$

de modo que la suma de cuadrados residual bajo el modelo reducido se corresponde con la suma de cuadrados residual de un modelo de regresión lineal simple y viene dada por:

$$\begin{aligned} scR_0 &= \sum_{i=1}^I \sum_{t=1}^{n_i} \left( y_{it} - \hat{\mu} - \hat{\tau} - \hat{\beta} (x_{it} - \bar{x}_{..}) \right)^2 \\ &= \sum_{i=1}^I \sum_{t=1}^{n_i} \left( y_{it} - \bar{y}_{..} - \frac{sp_{xy}}{sc_{xx}} (x_{it} - \bar{x}_{..}) \right)^2 \\ &= \boxed{sc_{yy} - \frac{(sp_{xy})^2}{sc_{xx}}} \end{aligned} \quad (4.11)$$

donde  $sc_{yy} = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{..})^2$ . Además, igual que en una regresión lineal simple, el número de grados de libertad del error es  $n - 2$  (número de observaciones menos uno por la estimación de la constante y otro por la de la pendiente).

Por otro lado, bajo el modelo completo (modelo (4.2)), los estimadores mínimo-cuadráticos vienen dados por (4.3) y (4.4), de donde se deduce la siguiente suma de cuadrados residual:

$$\begin{aligned}
 scR &= \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}(x_{it} - \bar{x}_{..}))^2 \\
 &= \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{i.} + \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) - \hat{\beta}(x_{it} - \bar{x}_{..}))^2 \\
 &= \sum_{i=1}^I \sum_{t=1}^{n_i} ((y_{it} - \bar{y}_{i.}) - \hat{\beta}(x_{it} - \bar{x}_{i.}))^2 \\
 &= sc_{yy}^* - \hat{\beta}(sp_{xy}^*) = \boxed{sc_{yy}^* - \frac{(sp_{xy}^*)^2}{sc_{xx}^*}} \tag{4.12}
 \end{aligned}$$

donde  $sc_{yy}^* = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{i.})^2$  y  $sp_{xy}^*$  y  $sc_{xx}^*$  vienen dados por (4.5) y (4.6) respectivamente. Los grados de libertad de  $scR$  son  $n - I - 1$ , o sea, uno menos que en un análisis de la varianza de una vía debido a tener que estimar el parámetro  $\beta$ .

Finalmente, de (4.11) y (4.12) se tiene:

$$\boxed{scT = scR_0 - scR = \left( sc_{yy} - \frac{(sp_{xy})^2}{sc_{xx}} \right) - \left( sc_{yy}^* - \frac{(sp_{xy}^*)^2}{sc_{xx}^*} \right)} \tag{4.13}$$

La suma de cuadrados  $scT$  tiene  $(n - 2) - (n - I - 1) = I - 1$  grados de libertad y se denomina **suma de cuadrados entre niveles del factor tratamiento** o **suma de cuadrados explicada por diferencias entre los niveles del factor tratamiento**.

Se verifica:

- Bajo  $H_0$ :  $\frac{SCT}{\sigma^2} \sim \chi^2_{(I-1)}$  y  $\frac{SCMT}{SCMR} \sim F_{I-1, n-I-1}$

- Por tanto:  $\boxed{\text{Rechazar } H_0 \text{ al nivel } \alpha \text{ si } \hat{F} = \frac{scmT}{scmR} > F_{I-1, n-I-1, \alpha}}$

También es interesante valorar la utilidad de haber incluido la covariable en el modelo. Es decir, contrastar las hipótesis:

$$\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$$

El modelo reducido para esta prueba es el modelo del análisis de la varianza de una vía:  $Y_{it} = \mu + \tau_i + \varepsilon_{it}$ . En el capítulo 2 se dedujo que la correspondiente suma de cuadrados residual es:

$$scR_0 = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_i)^2 = sc_{yy}^*$$

con  $I - 1$  grados de libertad. La suma de cuadrados residual para el modelo completo se calcula mediante (4.12). Por tanto:

$$sc\beta = scR_0 - scR = \frac{(sp_{xy}^*)^2}{sc_{xx}^*} = \hat{\beta}^2 sc_{xx}^* \tag{4.14}$$

La diferencia entre los grados de libertad es  $(n - I) - (n - I - 1) = 1$ , de modo que suma de cuadrados y suma de cuadrados medios coinciden en este caso.

Se verifica:

- Bajo  $H_0$ :  $\frac{SC\beta}{\sigma^2} \sim \chi^2_{(1)}$  y  $\frac{SCM\beta}{SCMR} \sim F_{1, n-I-1}$

- Por tanto: Rechazar  $H_0$  al nivel  $\alpha$  si  $\hat{F} = \frac{scm\beta}{scmR} > F_{1, n-I-1, \alpha}$

## 4.5 Comprobación de la Idoneidad del Modelo

Además de las suposiciones habituales sobre las variables error, el análisis del modelo incluyendo una covariable presupone la existencia de una relación lineal entre la covariable y la respuesta media, con la misma pendiente para cada tratamiento. En consecuencia, el análisis de los residuos debe comenzar con un chequeo sobre la bondad del ajuste de nuestro modelo.

Los gráficos de dispersión de los residuos frente a la covariable por cada tratamiento permiten intuir la existencia o no de linealidad. Si para algún tratamiento el gráfico muestra alguna tendencia no lineal, el modelo no es apropiado. Comprobado que la relación lineal parece funcionar, todavía cabe preguntarse si las pendientes con cada tratamiento son semejantes. Una prueba de hipótesis sobre la igualdad de pendientes se realiza comparando el ajuste obtenido con nuestro modelo y el ajuste que proporciona un modelo que asume pendientes distintas, es decir:

$$Y_{it} = \mu + \tau_i + \beta_i (x_{it} - \bar{x}_{..}) + \varepsilon_{it}$$

Una vez comprobado que no hay una falta de ajuste significativa, se continua chequeando el resto de suposiciones sobre el error tal y como se indicó en los capítulos previos.

| CUADRO DEL ANÁLISIS DE LA COVARIANZA<br>— MODELO DE UNA VÍA Y UNA COVARIABLE —  |   |             |   |                         |
|---|---|-------------|---|-------------------------|
| Fuente de Variación   | Suma de Cuadrados   | g.l.        | scm   | $\hat{F}$               |
| Tratamientos  | $scT = \left( sc_{yy} - \frac{(sp_{xy})^2}{sc_{xx}} \right) - \left( sc_{yy}^* - \frac{(sp_{xy}^*)^2}{sc_{xx}^*} \right)$ | $I - 1$     | $\frac{scmT}{I - 1} = \frac{scT}{I - 1}$              | $\frac{scmT}{scmR}$     |
| Covariable  | $sc\beta = \frac{(sp_{xy}^*)^2}{sc_{xx}^*}$   | 1           | $\frac{scm\beta}{sc\beta} = \frac{scm\beta}{sc\beta}$ | $\frac{scm\beta}{scmR}$ |
| Residual  | $scR = \left( sc_{yy}^* - \frac{(sp_{xy}^*)^2}{sc_{xx}^*} \right)$  | $n - I - 1$ | $\frac{scmR}{n - I - 1} = \frac{scR}{n - I - 1}$      |                         |
| Global  | $scG = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{..})^2$   | $n - 1$     | $\frac{scmG}{n - 1} = \frac{scG}{n - 1}$              |                         |
| <p>Rechazar <math>H_0 : \tau_1 = \tau_2 = \dots = \tau_I</math>, según <math>p = P \left( \frac{scmT}{scmR} &gt; F_{I-1, n-I-1} \right)</math></p> <p>Rechazar <math>H_0 : \beta = 0</math>, según <math>p = P \left( \frac{scm\beta}{scmR} &gt; F_{1, n-I-1} \right)</math></p>  |   |             |   |                         |
| $sc_{xx} = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_{..})^2$ $sc_{yy} = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{..})^2$ $sp_{xy} = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_{..})(y_{it} - \bar{y}_{..})$ $sc_{xx}^* = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_i)^2$ $sc_{yy}^* = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_i)^2$ $sp_{xy}^* = \sum_{i=1}^I \sum_{t=1}^{n_i} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)$ |   |             |   |                         |

Tabla 4.1: Cuadro del análisis de la covarianza para un diseño completamente aleatorizado con un factor tratamiento de efectos fijos y una covariable.

### 4.6 Contrastes Individuales y Comparaciones Múltiples

Según (4.3), el estimador mínimo cuadrático de cualquier contraste individual de tratamientos,

$\sum_{i=1}^I c_i \tau_i$ , es:

$$\sum_{i=1}^I c_i \hat{\tau}_i = \sum_{i=1}^I c_i (\hat{\mu} + \hat{\tau}_i) = \sum_{i=1}^I c_i (\bar{Y}_i - \hat{\beta}(\bar{x}_i - \bar{x}_{..})) = \sum_{i=1}^I c_i (\bar{Y}_i - \hat{\beta}\bar{x}_i) \tag{4.15}$$

dado que el término  $\sum_{i=1}^I c_i \hat{\beta} \bar{x}_i$  es cero por serlo la suma de los coeficientes  $c_i$ .

Por otro lado se prueba que  $Var(\hat{\beta}) = \frac{\sigma^2}{sc_{xx}^*}$  y que  $Cov(\bar{Y}_i, \hat{\beta}) = 0$ , de modo que:

$$\begin{aligned} Var\left(\sum_{i=1}^I c_i \hat{\tau}_i\right) &= Var\left(\sum_{i=1}^I c_i \bar{Y}_i\right) + Var\left(\hat{\beta} \sum_{i=1}^I c_i \bar{x}_i\right) \\ &= \sigma^2 \left[ \left(\sum_{i=1}^I \frac{c_i^2}{n_i}\right) + \left(\frac{1}{sc_{xx}^*}\right) \left(\sum_{i=1}^I c_i \bar{x}_i\right)^2 \right] \end{aligned} \quad (4.16)$$

De (4.15), (4.16) y la normalidad e independencia de los estimadores  $\bar{Y}_i$  y  $\hat{\beta}$  se deduce que los estimadores de los contrastes individuales verifican:

$$\boxed{\sum_{i=1}^I c_i \hat{\tau}_i = \sum_{i=1}^I c_i (\bar{Y}_i - \hat{\beta} \bar{x}_i) \sim N\left(\sum_{i=1}^I c_i \tau_i, \sigma \sqrt{\left(\sum_{i=1}^I \frac{c_i^2}{n_i}\right) + \left(\frac{1}{sc_{xx}^*}\right) \left(\sum_{i=1}^I c_i \bar{x}_i\right)^2}\right)} \quad (4.17)$$

Es claro a partir de (4.16) que la varianza del contraste se estima mediante:

$$\widehat{Var}\left(\sum_{i=1}^I c_i \hat{\tau}_i\right) = scmR \left[ \left(\sum_{i=1}^I \frac{c_i^2}{n_i}\right) + \left(\frac{1}{sc_{xx}^*}\right) \left(\sum_{i=1}^I c_i \bar{x}_i\right)^2 \right] \quad (4.18)$$

y como  $SCR/\sigma^2$  tiene una distribución  $\chi^2$  con  $n - I - 1$  grados de libertad:

$$\frac{\sum_{i=1}^I c_i \hat{\tau}_i - \sum_{i=1}^I c_i \tau_i}{\sqrt{\widehat{Var}\left(\sum_{i=1}^I c_i \hat{\tau}_i\right)}} \sim t_{n-I-1}$$

De modo que un intervalo de confianza al  $100(1 - \alpha)\%$  para  $\sum_{i=1}^I c_i \tau_i$  es

$$\boxed{\sum_{i=1}^I c_i \tau_i \in \left(\sum_{i=1}^I c_i \hat{\tau}_i \mp t_{n-I-1, \alpha/2} \sqrt{scmR \left[ \left(\sum_{i=1}^I \frac{c_i^2}{n_i}\right) + \left(\frac{1}{sc_{xx}^*}\right) \left(\sum_{i=1}^I c_i \bar{x}_i\right)^2 \right]}\right)} \quad (4.19)$$

En cuanto a las comparaciones múltiples comentar que los métodos de Bonferroni y Scheffé son válidos en este contexto del análisis de la covarianza. Sin embargo, dado que las medias

ajustadas  $\hat{\mu} + \hat{\tau}_i = \bar{Y}_i - \hat{\beta}(\bar{x}_i - \bar{x}_{..})$  no son independientes, no está claro que los métodos de Tukey y de Dunnett sean aplicables aquí. Por ejemplo, se cree que el método de comparaciones múltiples de Tukey todavía mantiene el nivel de significación global sin embargo no está probado.

## 4.7 Problemas Propuestos

**Problema 1** (C.R. Hicks, 1965, *Industrial Quality Control*) Se disponía de doce soportes muy delgados de acero (unidades experimentales) que se enviaron a chapear con zinc a tres empresas distintas (cuatro a cada una de ellas). La variable respuesta fue el espesor de cada laminado de zinc en unidades adecuadas. Previamente, se midió el espesor de las láminas de acero y esta medida se consideró como una covariable en el modelo. Los datos se encuentran en el archivo ZINC.SAV.

1. Construir un diagrama de dispersión de la nube (respuesta, covariable) ¿Qué relación parece existir entre el grosor del chapeado de zinc y el grosor de la lámina de acero antes de ser chapeada? ¿Parece apropiado un modelo de análisis de la covarianza? ¿Parece existir un “efecto empresa”?
2. Ajustar el modelo de análisis de la covarianza a los datos.
3. Evaluar las suposiciones estructurales del modelo.
4. Completar el cuadro de análisis de la covarianza.
5. Ajustar un modelo de análisis de la varianza de una vía ignorando la covariable y completar el correspondiente cuadro de análisis de la varianza.
6. Comparar la conducta de ambos modelos. ¿Cuál presenta un error cuadrático medio más pequeño? ¿Qué modelo proporciona una mayor evidencia de que las empresas proporcionan resultados distintos?

**Problema 2** Retómese el ajuste de un modelo de análisis de la varianza de una vía realizado para el ejemplo de los tiempos de inflado de globos (ver el Ejemplo 2.4 en la sección 2.10.3). Los residuos estandarizados que se generan muestran una clara tendencia a decrecer con el tiempo (ver Figura 2.8). Considérese entonces el ajuste que proporciona un modelo de análisis de la covarianza considerando como covariable el orden de inflado de los globos. Analizar los resultados. Los datos están en el archivo GLOBOS.SAV.

**Problema 3** (Bortnick, S., Hoffman, M, Lewis, K.K. y Williams, C., 1996) Cuatro estudiantes planificaron y ejecutaron un experimento piloto para comparar los efectos de dos factores tratamiento, marca y dibujo, sobre la capacidad de absorción de servilletas de papel. Seleccionaron tres tipos de marcas de servilletas (factor *A* con tres niveles) y, de cada marca, tomaron servilletas blancas y dibujadas (factor *B* con dos niveles, 1: blanca y 2: dibujada). Sobre una superficie lisa colocaron dos pares de libros y, suspendida entre ellos,

situaban una servilleta de papel. Entonces dejaban caer gotas de agua sobre la servilleta hasta que el agua la comenzaba a filtrar goteando sobre la superficie en la que se apoyaban los libros. Se tomaron las siguientes medidas: el tiempo, en segundos, que transcurrió hasta poder evaluar cada observación (o sea, hasta que se filtró la primera gota) y el número de gotas de agua absorbidas por pulgada al cuadrado de servilleta (este dato se empleó como medida de absorción de la servilleta). Además, se consideró como covariable la tasa de gotas caídas en relación al tiempo transcurrido (en número de gotas por segundo). Los datos están en el archivo ABSORCION.SAV.

1. Construir un diagrama de dispersión de la absorción frente a la covariable etiquetando cada punto de la nube con el tratamiento aplicado. ¿Parece apropiado un modelo de análisis de la covarianza? ¿Parecen existir diferencias entre los efectos tratamiento?
2. Completar el cuadro de análisis de la covarianza.
3. Analizar los residuos gráficamente.
4. Ajustar un modelo de análisis de la covarianza de dos vías. Contrastar los efectos principales y el efecto interacción.



# Capítulo 5

## Diseños en Bloques

### 5.1 Generalidades

#### 5.1.1 Introducción

Un diseño en bloques es apropiado cuando el objetivo del experimento es comparar los efectos de diferentes tratamientos promediados sobre un rango de condiciones experimentales distintas.

Bloquear un experimento consiste en distribuir las unidades experimentales en subgrupos tales que unidades experimentales pertenecientes a un mismo subgrupo deben ser similares y pueden ser analizadas en condiciones experimentales semejantes, en tanto que unidades experimentales ubicadas en subgrupos distintos darán lugar probablemente a respuestas muy diferentes aún cuando sean asignadas a un mismo tratamiento.

Cada uno de estos conjuntos de unidades experimentales similares se denomina **bloque** o **nivel del factor de bloqueo** o **factor bloque**.

El análisis de un diseño en bloques incluye la comparación de diferentes tratamientos aplicados a las unidades experimentales de un mismo bloque. De este modo se pretende simultáneamente:

1. evitar que grandes diferencias entre las unidades experimentales enmascaren diferencias reales entre los tratamientos y,
2. medir los efectos de los tratamientos en condiciones experimentales distintas.

**Ejemplo 5.1** Se desea determinar si cuatro puntas distintas producen diferencias en las lecturas de un equipo para medir la dureza. La máquina funciona presionando la punta sobre una probeta de metal y determinando la dureza de la probeta a partir de la profundidad de la marca que se produce. El experimentador ha decidido obtener cuatro observaciones para cada punta. Sólo existe un factor –tipo de punta– y el

diseño de un factor completamente aleatorizado consiste en asignar al azar a cada uno de los  $4 \times 4 = 16$  ensayos a *una unidad experimental*, o sea, a una probeta de metal, y tomar las lecturas de dureza correspondientes. Se requerirían entonces 16 unidades experimentales para ejecutar el experimento.

En principio existe un serio problema con este diseño: si las probetas son ligeramente distintas en cuanto a dureza, como ocurriría si proviniesen de diferentes vaciados, las unidades experimentales contribuirían a incrementar la variabilidad observada en las lecturas de la dureza. Para evitar este problema y sustraer del error experimental la variabilidad originada por las probetas, podría probarse cada punta una vez en cada una de cuatro probetas distintas. Esta última propuesta conlleva el bloqueo de las unidades experimentales: cada probeta es un bloque.

En ocasiones los niveles de un factor bloque son los valores de una covariable que ha sido medida previamente al experimento, de modo que esos valores se emplean para agrupar las unidades experimentales.

Sin embargo, lo más frecuente es formar los bloques en base a características que no pueden ser convenientemente medidas (hecho por el que, precisamente, los bloques son muy populares y muy habitualmente utilizados). Por ejemplo:

- en agricultura cabe esperar que parcelas de tierra contiguas tengan características comunes,
- en experimentos industriales cabe esperar que las piezas producidas por una misma máquina sean similares,
- en experimentos médicos o biológicos los experimentadores son conscientes de que medidas tomadas sobre el mismo individuo probablemente resulten semejantes en tanto que sobre individuos diferentes las respuestas también difieran, etc.

### 5.1.2 Acerca del Tamaño de los Bloques

Aunque es perfectamente posible que cada bloque contenga un número distinto de unidades experimentales, en este manual *sólo se estudian diseños en bloques con igual tamaño*.

En lo que sigue se denotará por:

- $b$  al número de bloques del factor bloque considerado
- $k$  al número común de unidades experimentales en cada bloque

En ocasiones el tamaño de los bloques viene determinado de forma natural por el propio experimento.

**Ejemplo 5.2** Se dispone de distintas masas de pan e interesa saber cuál de ellas es preferible en base a alguna característica medible una vez cocido el pan durante

un cierto tiempo. Se va a utilizar un horno y se desea planificar el experimento controlando los grados de temperatura (factor bloque). Las unidades experimentales serán porciones de igual tamaño de las distintas masas de pan. Suponiendo que la temperatura del horno es aproximadamente constante en cada bandeja, los bloques “naturales” serán las bandejas: cada bandeja garantiza que las unidades experimentales ubicadas sobre ella son *semejantes* en términos de temperatura. Luego el número  $b$  de bloques coincidirá con el número de bandejas y el número  $k$  de unidades experimentales en cada bloque (tamaño del bloque) será el número de masas de pan que entran en cada bandeja.

Otras veces sin embargo es el experimentador quién debe decidir el tamaño de los bloques. En experimentos industriales por ejemplo es muy habitual planificar un experimento considerando como bloques períodos concretos de tiempo como precaución ante el cambio de las condiciones experimentales. La división de los períodos de tiempo suele venir sugerida por un experimento piloto.

Es muy importante advertir que:

Bloquear sin ser necesario conduce a pruebas de hipótesis menos potentes y a intervalos de confianza más amplios que aquellos que se obtendrían mediante un diseño completamente aleatorizado.

### 5.1.3 Definiciones

Habiendo decidido el tamaño de bloque  $k$  y habiendo agrupado las unidades experimentales similares en bloques, solo resta asignar las unidades experimentales a los distintos niveles de los factores tratamiento.

La peor asignación posible es aquella que aplica cada tratamiento a todas las unidades experimentales de un mismo bloque y, por tanto, tratamientos distintos a bloques distintos. Tal asignación de los tratamientos no permite determinar si las posibles diferencias se deben a los bloques o a los tratamientos. Se dice entonces que los efectos bloque y tratamiento están **confundidos**.

Por el contrario, la mejor asignación posible consiste en aplicar todos y cada uno de los tratamientos al mismo número de unidades experimentales en cada bloque. Para ello es necesario que, si  $I$  denota el número de tratamientos, entonces:

$$k = s \times I, \quad s = 1, 2, 3, \dots$$

Si  $s = 1 \Leftrightarrow k = I$ , el diseño se denomina **Diseño en Bloques Completos** o **Diseño en Bloques Completamente Aleatorizados**.

Si  $s > 1$ , el diseño se denomina **Diseño General en Bloques Completos**.

Si  $k \neq s \times I$ , se habla de **Diseño en Bloques Incompletos**, si bien en muchas referencias bibliográficas esta denominación se emplea para designar específicamente el caso particular en que  $k < I$  (o sea, no hay unidades experimentales suficientes en cada bloque para cubrir todo el abanico de tratamientos considerado).

## 5.2 Diseños en Bloques Completos

### 5.2.1 Diseño en Bloques Completamente Aleatorizados

Un *diseño en bloques completamente aleatorizados* es un diseño con  $I$  tratamientos (que podrían ser combinaciones cruzadas de niveles de factores tratamiento) y  $n = bI$  unidades experimentales agrupadas en  $b$  bloques de tamaño  $k = I$ , de modo que unidades en un mismo bloque son semejantes y unidades en bloques distintos son sustancialmente diferentes.

Además, las  $k = I$  unidades experimentales de cada bloque son asignadas a los tratamientos al azar y de tal manera que cada tratamiento se aplica a una única unidad experimental por bloque. O sea, cada tratamiento aparece una vez en cada bloque y  $b$  veces en el diseño.

**Ejemplo 5.3** Volviendo al Ejemplo 5.2 relativo a la cocción del pan, supóngase que se dan las siguientes circunstancias:

- son cuatro las masas de pan diferentes que interesa comparar y, por tanto, existen  $I = 4$  tratamientos que se codifican mediante 1, 2, 3 y 4,
- el horno dispone de tres bandejas a distintas alturas (se consideran entonces  $b = 3$  bloques),
- se decide usar el horno dos veces y, equivalentemente, se obtendrán dos réplicas del experimento.

Si se desea planificar el experimento a través de un diseño en bloques completamente aleatorizados es necesario que el tamaño de cada bloque  $k$  sea igual a 4. Así, se distinguirán en cada bandeja cuatro posiciones distintas codificadas como sigue: FI (al fondo a la izquierda), FD (al fondo a la derecha), DI (delante a la izquierda) y DD (delante a la derecha).

|    |    |
|----|----|
| FI | FD |
| DI | DD |

Para proceder con el principio de aleatorización se tiene que aleatorizar cada bloque por separado. Por ejemplo, se generan al azar cuatro dígitos de dos cifras (74,11,39,68) cuyo orden define una permutación para los tratamientos: (2,3,4,1). Este procedimiento podría dar lugar a una asignación de los tratamientos a los bloques como la que figura en la Tabla 5.1.

| Bloque | Réplica | Bandeja | FI | FD | DI | DD |
|--------|---------|---------|----|----|----|----|
| 1      | 1       | 1       | 2  | 3  | 4  | 1  |
| 2      |         | 2       | 1  | 2  | 3  | 4  |
| 3      |         | 3       | 4  | 3  | 2  | 1  |
| 4      | 2       | 1       | 2  | 4  | 3  | 1  |
| 5      |         | 2       | 2  | 4  | 1  | 3  |
| 6      |         | 3       | 3  | 2  | 4  | 1  |

Tabla 5.1: Ejemplo de un diseño en bloques completamente aleatorizados.

La Tabla 5.1 muestra que la masa de pan 1 se observará cuatro veces en DD mientras que la masa 2 nunca se observará en esa posición. Es decir, el diseño no será apropiado cuando existan diferencias reales entre las temperaturas de distintas posiciones de una misma bandeja.

### 5.2.2 Diseño General en Bloques Completos

Un *diseño general en bloques completos* es un diseño con  $I$  tratamientos y  $n = bIs$  unidades experimentales agrupadas en  $b$  bloques de tamaño  $k = Is$ , de modo que unidades en un mismo bloque son semejantes y unidades en bloques distintos son sustancialmente diferentes.

Además, las  $k = Is$  unidades experimentales de cada bloque son asignadas aleatoriamente a los  $I$  tratamientos con la única restricción de que cada tratamiento será aplicado a  $s$  unidades experimentales del bloque.

**Ejemplo 5.4** (Bist, P. y otros en 1995) Varios investigadores planificaron un experimento orientado a comparar las intensidades de luz de tres marcas de bombillas diferentes codificadas del 1 al 3. La marca 3 era más barata que las otras dos. Otro factor de interés (factor tratamiento) a considerar en el experimento era la capacidad de la bombilla. Concretamente se controlaron dos posibles cantidades de paso de corriente, distinguiendo 50% y 100%.

Se disponía por tanto de  $I = 6$  tratamientos (3 intensidades  $\times$  2 capacidades), que fueron codificados como sigue:

|      |   |   |   |
|------|---|---|---|
|      | 1 | 2 | 3 |
| 50%  | 1 | 2 | 3 |
| 100% | 4 | 5 | 6 |

Interesaba comparar los seis tratamientos utilizando bombillas de 60 y 100 vatios. Obviamente comparar la iluminación entre 60 y 100 vatios carece de interés. Tener en cuenta ambas potencias respondía a la única intención de bloquear el experimento. Como además el experimento se ejecutaría en dos días distintos se consideraron dos bloques determinados por las combinaciones:

$$B1: [ \text{Día 1 y 60 W} ] \quad \text{y} \quad B2: [ \text{Día 2 y 100 W} ]$$

Se tomaron cuatro observaciones en cada uno de los seis tratamientos de cada bloque.

### 5.2.3 Análisis del Diseño en Bloques Completamente Aleatorizados

El modelo matemático estándar para un diseño en bloques completamente aleatorizados es:

$$\underbrace{Y_{hi}}_{\text{aleatorio}} = \overbrace{\mu + \theta_h + \tau_i}^{\text{determinista}} + \underbrace{\varepsilon_{hi}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{hi} \text{ i.i.d. según una } N(0, \sigma) \quad (5.1)$$

para  $h = 1, \dots, b$  e  $i = 1, \dots, I$ , donde  $\mu$  es una constante,  $\theta_h$  representa el efecto (positivo o negativo) sobre la respuesta debido a que se observa el bloque  $h$ ,  $\tau_i$  representa el efecto (positivo o negativo) sobre la respuesta debido a que se observa el nivel  $i$  del factor tratamiento e  $Y_{hi}$  y  $\varepsilon_{hi}$  son las variables aleatorias que miden respectivamente la respuesta y su error asociado en la observación del  $i$ -ésimo tratamiento en el  $h$ -ésimo bloque.

El modelo (5.1) se conoce como **modelo tratamiento-bloque** y su estructura sugiere de inmediato dos reflexiones importantes:

[1] No existe la posibilidad de interacción en este modelo.

Obsérvese que si se incluyesen términos representando los efectos interacción entre bloques y tratamientos no habría grados de libertad suficientes para estimar la varianza del error. En general, en casi todos los experimentos bloqueados la suposición de no interacción entre bloques y tratamientos es muy razonable. En cualquier caso, de sospechar su existencia es preciso planificar un diseño general en bloques completos (que permite incrementar el tamaño muestral pero no el número de parámetros a estimar).

[2] El modelo tratamiento-bloque es similar al modelo de dos vías de efectos principales para dos factores tratamiento en un diseño completamente aleatorizado con una observación por celda.

Tal es así que no causa sorpresa la analogía que muestran los cuadros Anova asociados a ambos modelos.

Sin embargo existe una diferencia sustancial. En un diseño completamente aleatorizado, los tratamientos, y equivalentemente los niveles de ambos factores tratamiento, son asignados aleatoriamente a las unidades experimentales. Por el contrario, en un diseño en bloques como el estudiado en esta sección, aunque las observaciones son también tomadas sobre todas las combinaciones de tratamientos y bloques, sólo los niveles del factor tratamiento son asignados aleatoriamente a las unidades experimentales. En efecto, téngase presente que la división de las unidades experimentales para formar los bloques es determinista. Precisamente por ello, existe una fuerte controversia sobre si es apropiado o no contrastar la igualdad de los efectos bloque. Al fin y al cabo, en el diseño que se está analizando los bloques representan fuentes de variación “nuisance”, esto es, no existe interés alguno en la posible significación del efecto bloque (de ser así se contradice la naturaleza del bloque frente al factor tratamiento). Más aún, es muy poco probable que sea factible utilizar los mismos bloques en una hipotética réplica del experimento.

Por todo ello, en la práctica:

Carece de interés plantearse la hipótesis nula de igualdad de los efectos bloque. No se trata de buscar significación estadística. La comparación entre la suma de cuadrados medios atribuibles a los bloques y la suma de cuadrados medios residual es relevante al objeto de concluir si bloquear el experimento resultó o no beneficioso.

En efecto, si la suma de cuadrados medios atribuibles a los bloques es considerablemente mayor que la suma de cuadrados medios residual, habrá resultado útil bloquear en el sentido de que tal acción derivó en una reducción del tamaño del error experimental. En otro caso, bloquear es contraproducente, dado que genera pruebas de hipótesis menos potentes e intervalos de confianza más amplios.

Por supuesto, en este último caso no es correcto transformar el diseño sistemáticamente en un diseño completamente aleatorizado, interpretando el factor bloque como un factor tratamiento, pues sus niveles no han sido aleatorizados. Como mucho los resultados permiten aumentar el conocimiento para planificaciones posteriores de experimentos semejantes.

Para contrastar la igualdad de los efectos de los tratamientos el criterio es análogo al empleado con el modelo de dos vías de efectos principales. En cualquier caso la Tabla 5.2 recoge el cuadro de análisis de la varianza para el modelo tratamiento-bloque.

| CUADRO DEL ANÁLISIS DE LA VARIANZA                   |   |                  |                                     |                     |
|--|---|------------------|-------------------------------------|---------------------|
| — MODELO TRATAMIENTO-BLOQUE PARA BLOQUES COMPLETOS — |   |                  |                                     |                     |
| — Una observación por celda —                        |   |                  |                                     |                     |
| Fuente de Variación                                  | Suma de Cuadrados   | g.l.             | scm                                 | $\hat{F}$           |
| Bloques  | $scB = I \sum_{h=1}^b (\bar{y}_{h.} - \bar{y}_{..})^2$                                    | $b - 1$          | $scmB = \frac{scB}{b - 1}$          | —                   |
| Tratamientos   | $scT = b \sum_{i=1}^I (\bar{y}_{.i} - \bar{y}_{..})^2$                                    | $I - 1$          | $scmT = \frac{scT}{I - 1}$          | $\frac{scmT}{scmR}$ |
| Residual   | $scR = \sum_{h=1}^b \sum_{i=1}^I (y_{hi} - \bar{y}_{h.} - \bar{y}_{.i} + \bar{y}_{..})^2$ | $(b - 1)(I - 1)$ | $scmR = \frac{scR}{(b - 1)(I - 1)}$ |                     |
| Global   | $scG = \sum_{h=1}^b \sum_{i=1}^I (y_{hi} - \bar{y}_{..})^2$                               | $bI - 1$         | $scmG = \frac{scG}{bI - 1}$         |                     |

Rechazar  $H_0 : \tau_1 = \tau_2 = \dots = \tau_I$ , según  $p = P \left( \frac{scmT}{scmR} > F_{I-1, (b-1)(I-1)} \right)$

Tabla 5.2: Análisis de la varianza para un diseño en bloques completamente aleatorizados.

### 5.2.4 Análisis del Diseño General en Bloques Completos

En esta sección se presentan los modelos matemáticos para un diseño general en bloques completos, es decir, un diseño en bloques completos con  $s > 1$  observaciones en cada combinación de tratamiento y bloque. Ahora si es posible estimar los efectos interacción y se enuncian dos posibles modelos en función de que se incluyan o no los efectos interacción.

El **modelo tratamiento-bloque sin interacción** que responde a la expresión:

$$\underbrace{Y_{hit}}_{\text{aleatorio}} = \overbrace{\mu + \theta_h + \tau_i}^{\text{determinista}} + \underbrace{\varepsilon_{hit}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{hit} \text{ i.i.d. según una } N(0, \sigma) \tag{5.2}$$

para  $h = 1, \dots, b, i = 1, \dots, I$  y  $t = 1, \dots, s$ .

Y el **modelo tratamiento-bloque con interacción** que incluye los efectos de interacción entre los bloques y los tratamientos:

$$\underbrace{Y_{hit}}_{\text{aleatorio}} = \overbrace{\mu + \theta_h + \tau_i + (\theta\tau)_{hi}}^{\text{determinista}} + \underbrace{\varepsilon_{hit}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{hit} \text{ i.i.d. según una } N(0, \sigma) \quad (5.3)$$

para  $h = 1, \dots, b, i = 1, \dots, I$  y  $t = 1, \dots, s$ .

Es clara la analogía entre:

- el modelo tratamiento-bloque (5.2) para un diseño general en bloques completos y el modelo (3.3) de dos vías de efectos principales para el diseño completamente aleatorizado con dos factores tratamiento y  $s$  observaciones por celda,
- el modelo tratamiento-bloque con interacción (5.3) para un diseño general en bloques completos y el modelo (3.2) completo de dos vías para el diseño completamente aleatorizado con dos factores tratamiento y  $s$  observaciones por celda.

Por ello, las tablas del análisis de la varianza también son análogas (ver Tablas 5.3 y 5.4).

| <b>CUADRO DEL ANÁLISIS DE LA VARIANZA</b><br>— MODELO TRATAMIENTO-BLOQUE PARA BLOQUES COMPLETOS —<br>— Con $s$ observaciones por celda y sin interacción — |  |                   |   |                     |
|--|--|-------------------|---|---------------------|
| Fuente de Variación  | Suma de Cuadrados  | <i>g.l.</i>       | <i>scm</i>                                | $\hat{F}$           |
| Bloques  | $scB =$<br>$Is \sum_{h=1}^b (\bar{y}_{h..} - \bar{y}_{...})^2$                                       | $b - 1$           | $scmB =$<br>$\frac{scB}{b - 1}$           | —                   |
| Tratamientos   | $scT =$<br>$bs \sum_{i=1}^I (\bar{y}_{.i.} - \bar{y}_{...})^2$                                       | $I - 1$           | $scmT =$<br>$\frac{scT}{I - 1}$           | $\frac{scmT}{scmR}$ |
| Residual   | $scR =$<br>$s \sum_{h=1}^b \sum_{i=1}^I (y_{hit} - \bar{y}_{h..} - \bar{y}_{.i.} + \bar{y}_{...})^2$ | $bIs - b - I + 1$ | $scmR =$<br>$\frac{scR}{bIs - b - I + 1}$ |                     |
| Global   | $scG =$<br>$\sum_{h=1}^b \sum_{i=1}^I \sum_{t=1}^s (y_{hit} - \bar{y}_{...})^2$                      | $bIs - 1$         | $scmG =$<br>$\frac{scG}{bIs - 1}$         |                     |
| Rechazar $H_0 : \tau_1 = \tau_2 = \dots = \tau_I$ , según $p = P\left(\frac{scmT}{scmR} > F_{I-1, bIs-b-I+1}\right)$                                       |  |                   |   |                     |

Tabla 5.3: Análisis de la varianza para un diseño general en bloques completos sin interacción y con tamaño de bloque  $k = Is$ .

| <b>CUADRO DEL ANÁLISIS DE LA VARIANZA</b><br><b>— MODELO TRATAMIENTO-BLOQUE PARA BLOQUES COMPLETOS —</b><br><b>— Con <math>s</math> observaciones por celda y con interacción —</b>  |  |                  |                                       |                      |
|--|--|------------------|---------------------------------------|----------------------|
| <i>Fuente de Variación</i>   | <i>Suma de Cuadrados</i>   | <i>g.l.</i>      | <i>scm</i>                            | $\hat{F}$            |
| Bloques  | $scB = Is \sum_{h=1}^b (\bar{y}_{h..} - \bar{y}_{...})^2$  | $b - 1$          | $scmB = \frac{scB}{b - 1}$            | —                    |
| Tratamientos   | $scT = bs \sum_{i=1}^I (\bar{y}_{.i.} - \bar{y}_{...})^2$  | $I - 1$          | $scmT = \frac{scT}{I - 1}$            | $\frac{scmT}{scmR}$  |
| Interacción  | $scBT = s \sum_{h=1}^b \sum_{i=1}^I (\bar{y}_{hi.} - \bar{y}_{h..} - \bar{y}_{.i.} + \bar{y}_{...})^2$ | $(b - 1)(I - 1)$ | $scmBT = \frac{scBT}{(b - 1)(I - 1)}$ | $\frac{scmBT}{scmR}$ |
| Residual   | $scR = \sum_{h=1}^b \sum_{i=1}^I \sum_{t=1}^s (y_{hit} - \bar{y}_{hi.})^2$                             | $bI(s - 1)$      | $scmR = \frac{scR}{bI(s - 1)}$        |                      |
| Global   | $scG = \sum_{h=1}^b \sum_{i=1}^I \sum_{t=1}^s (y_{hit} - \bar{y}_{...})^2$                             | $bIs - 1$        | $scmG = \frac{scG}{bIs - 1}$          |                      |
| Rechazar $H_0^{BT} : (\theta\tau)_{hi} - \overline{(\theta\tau)}_{h.} - \overline{(\theta\tau)}_{.i} + \overline{(\theta\tau)}_{..} = 0 \forall h, i$<br>en base al $p$ -valor $p = P\left(\frac{scmBT}{scR} > F_{(b-1)(I-1), bI(s-1)}\right)$ |  |                  |                                       |                      |
| Si no se rechaza $H_0^{BT}$ entonces<br>Rechazar $H_0 : \tau_1^* = \tau_2^* = \dots = \tau_I^*$ , según $p = P\left(\frac{scmT}{scmR} > F_{I-1, bI(s-1)}\right)$   |  |                  |                                       |                      |

Tabla 5.4: Análisis de la varianza para un diseño general en bloques completos con interacción y con tamaño de bloque  $k = Is$ .

En el modelo con interacción se deberá contrastar el efecto interacción en primer lugar. Si resulta significativo entonces no resulta de interés contrastar la igualdad de los efectos tratamiento. Como en la sección previa resulta interesante una evaluación de la suma de cuadrados media debida al bloqueo en relación con la residual a los efectos de valorar la oportunidad o inconveniencia de haber bloqueado el experimento.

### 5.2.5 Tamaños Muestrales

En un diseño en bloques completamente aleatorizados está predeterminado el tamaño de los bloques ( $k = I$ ). Por tanto, para determinar de antemano el tamaño muestral que se precisa para obtener una potencia determinada en el análisis de la varianza, basta preguntarse por el número de bloques a considerar y, equivalentemente, por el tamaño muestral necesario para obtener esa potencia en un diseño completamente aleatorizado de una vía (véase la similitud de ambos en la Tabla 5.5).

| Completamente Aleatorizado |      |      |     |        |      | Bloques Completamente Aleatorizados |                   |      |        |     |      |        |
|----------------------------|------|------|-----|--------|------|-------------------------------------|-------------------|------|--------|-----|------|--------|
| Factor de interés          |      |      |     |        |      | Bloque                              | Factor de interés |      |        |     |      |        |
| 1(1)                       | 1(2) | 1(3) | ... | 1(I-1) | 1(I) | B.1                                 | 1(2)              | 1(3) | 1(1)   | ... | 1(I) | 1(I-1) |
| 2(1)                       | 2(2) | 2(3) | ... | 2(I-1) | 2(I) | B.2                                 | 2(1)              | 2(3) | 2(I-1) | ... | 2(2) | 2(I)   |
| ⋮                          | ⋮    | ⋮    | ... | ⋮      | ⋮    | ⋮                                   | ⋮                 | ⋮    | ⋮      | ... | ⋮    | ⋮      |
| b(1)                       | b(2) | b(3) | ... | b(I-1) | b(I) | B.b                                 | b(2)              | b(3) | b(1)   | ... | b(I) | m(I-1) |

Tabla 5.5: Similitud entre los diseños de una vía completamente aleatorizado y en bloques completamente aleatorizados.

En un diseño general en bloques completos existen  $n = bIs$  unidades experimentales divididas en  $b$  bloques de tamaño  $Is$ . Una vez más  $I$  está dado. Pero, por otro lado, la selección de  $b$  y  $s$  estará sujeta a: la necesidad de acomodar equilibradamente las observaciones en los bloques, el presupuesto y garantizar una potencia dada.

Para seleccionar  $b$  y  $s$  buscando una potencia predeterminada se empleará el mismo método que se utilizó en el modelo de dos vías de efectos principales para determinar las observaciones precisas por celda al comparar los tratamientos de **uno** sólo de los factores de interés (ver expresión (3.30) en la sección 3.8). Así se tiene:

$$bs \geq \frac{2I\sigma^2\phi^2}{\Delta^2}$$

Se fija  $b$  o  $s$  y se calcula el otro dependiendo de si el tamaño de los bloques o el número de bloques respectivamente pueda ser grande.

## 5.3 Diseños en Bloques Incompletos

### 5.3.1 Definiciones y Consideraciones Generales

En muchos experimentos que se planifican utilizando diseños en bloques aleatorizados, el tamaño de los bloques está limitado por diversas causas (capacidad del equipo, posibilidad de evaluación del material experimental similar, cansancio del observador, ...).

Cuando el número de tratamientos es muy grande, esta limitación puede impedir, por ejemplo, que cada tratamiento sea asignado a unidades experimentales de todos los bloques. En tal caso es necesario planificar un **diseño en bloques incompletos**.

La pregunta inmediata que uno se hace en tal situación es ¿cuántos tratamientos se observarán dentro de cada bloque? En primer lugar indicar que en este capítulo únicamente se describen **diseños equireplicados**. Es decir, diseños donde cada tratamiento se observa un mismo número  $r$  de veces en el experimento.

Denótese entonces por  $n_{hi}$  el número de veces que el tratamiento  $i$  se asigna al bloque  $h$ . En general, es más adecuado observar tantos tratamientos diferentes como sea posible en cada bloque. Así la varianza de los estimadores de los contrastes relativos a los tratamientos tenderá a ser menor en promedio.

Cuando el tamaño de bloque  $k$  sea menor que el número de tratamientos  $I$ , la estrategia anterior conduce a tomar:

$$n_{hi} = 1 \text{ o } 0 \text{ para todo } h \text{ y para todo } i$$

Un diseño verificando esta condición se denomina **diseño binario**.

Más aún, para muchos propósitos los mejores diseños binarios son aquellos en los cuales todos los pares de tratamientos ocurren juntos en el mismo, o aproximadamente el mismo, número de bloques. Esto presenta la ventaja de generar intervalos de confianza para los contrastes apareados de igual, o aproximadamente igual, longitud.

**Plan Experimental y Aleatorización.** Al diseñar un experimento con bloques incompletos, se distinguen dos etapas:

- Un primer paso consiste en planificar una distribución de las etiquetas de los tratamientos dentro de los bloques **del modo más uniforme que sea posible**. Esto conduce a lo que se suele denominar **plan experimental**.

**Ejemplo 5.5** La Tabla 5.6 muestra el plan experimental para un diseño en bloques incompletos con  $b = 8$  bloques de tamaño  $k = 3$  e  $I = 8$  tratamientos observados un número igual de veces,  $r = 3$ , en el experimento.

| Bloque |   |   |   | Bloque |   |   |   |
|--------|---|---|---|--------|---|---|---|
| I      | 1 | 3 | 8 | V      | 5 | 7 | 4 |
| II     | 2 | 4 | 1 | VI     | 6 | 8 | 5 |
| III    | 3 | 5 | 2 | VII    | 7 | 1 | 6 |
| IV     | 4 | 6 | 3 | VIII   | 8 | 2 | 7 |

Tabla 5.6: Plan experimental para un diseño en bloques incompletos con  $b = 8$ ,  $k = 3$ ,  $I = 8$  y  $r = 3$ .

Por *distribución más uniforme que sea posible* se entiende que, por ejemplo, en el plan experimental descrito en la Tabla 5.6, ninguna etiqueta aparezca más de una vez en cada bloque y cada pareja de etiquetas aparezca en cada bloque una o cero veces.

- El siguiente paso es el proceso de **aleatorización**, que consistiría en:
  1. Asignar al azar las etiquetas de los bloques del plan experimental a los niveles del factor bloque.
  2. Asignar al azar las unidades experimentales de cada bloque a las etiquetas de aquellos tratamientos ubicados en el correspondiente bloque.

**Ejemplo 5.6** Considérese un experimento que se ejecuta para comparar la resistencia a la tensión de  $I = 7$  aleaciones distintas de un metal. Supóngase además que sólo es posible tomar tres observaciones diarias y que el experimento debe estar finalizado en un plazo de siete días. Dado que distinto personal técnico puede trabajar en el laboratorio en días diferentes y la temperatura del laboratorio no será exactamente igual a lo largo de los siete días, cabe considerar a cada uno de los siete días como un bloque. Así se tiene un diseño en bloques incompletos con  $b = 7$  bloques de tamaño  $k = 3$  e  $I = 7$  tratamientos.

El plan experimental que se muestra en la primera columna de la Tabla 5.7 podría ser válido. Obsérvese que es binario, ya que cada tratamiento o no aparece o aparece una única vez en cada bloque, y es 3-replicado, o sea todos los tratamientos se observan exactamente tres veces. Se puede comprobar además que cada combinación de dos etiquetas de tratamiento aparece en un bloque exactamente una vez, de modo que las etiquetas de tratamientos están uniformemente distribuidas por los bloques.

| Etiqu.<br>Bloque | Plan  | Día | Aleat.<br>Paso 1 | Aleat.<br>Paso 2 |
|------------------|-------|-----|------------------|------------------|
| I                | 1 2 4 | 1   | 6 7 2            | 2 7 6            |
| II               | 2 3 5 | 2   | 2 3 5            | 5 3 2            |
| III              | 3 4 6 | 3   | 3 4 6            | 4 3 6            |
| IV               | 4 5 7 | 4   | 1 2 4            | 2 1 4            |
| V                | 5 6 1 | 5   | 5 6 1            | 1 5 6            |
| VI               | 6 7 2 | 6   | 4 5 7            | 4 5 7            |
| VII              | 7 1 3 | 7   | 7 1 3            | 7 3 1            |

Tabla 5.7: Plan experimental y aleatorización de un diseño en bloques incompletos con  $b = 7$ ,  $k = 3$ ,  $I = 7$  y  $r = 3$ .

Ahora se procede a aleatorizar el plan siguiendo los dos pasos indicados. En primer lugar es preciso asignar al azar las etiquetas de los bloques a los días. Por ejemplo generamos números aleatorios de dos dígitos y les asociamos las etiquetas de bloque:

Dígitos aleatorios: 71 36 65 93 92 02 97  
 Etiquetas bloque: I II III IV V VI VII

A continuación, reordenando los números en orden ascendente, se reordenan las etiquetas de bloque y se asignan al orden natural de los días:

Etiquetas bloque: VI II III I V IV VII  
 Días: 1 2 3 4 5 6 7

El resultado de este primer paso de la aleatorización se muestra en la cuarta columna de la Tabla 5.7.

En un segundo paso y de modo similar se realiza la asignación al azar de los slots de tiempo a las etiquetas de tratamiento. El resultado figura en la última columna de la Tabla 5.7.

La importancia de seleccionar un plan experimental con una distribución de las etiquetas de los tratamientos entre los bloques lo más uniforme posible es doble:

- por un lado, intentar asegurar que todos los contrastes de tratamientos sean estimables, y,
- por otro, tratar que los estadísticos de los contrastes apareados (pairwise) tengan varianzas similares.

El diseño de la Tabla 5.8 carece de tal uniformidad: algunos bloques contienen a todas las etiquetas pares de los tratamientos y el resto contiene a las impares.

| Bloque |   |   | Bloque |       |
|--------|---|---|--------|-------|
| I      | 1 | 3 | V      | 5 7 1 |
| II     | 2 | 4 | VI     | 6 8 2 |
| III    | 3 | 5 | VII    | 7 1 3 |
| IV     | 4 | 6 | VIII   | 8 2 4 |

Tabla 5.8: Plan experimental de un diseño en bloques incompletos con  $b = 8$ ,  $k = 3$ ,  $I = 8$  y  $r = 3$ .

La consecuencia inmediata de un diseño como el de la Tabla 5.8 es que los contrastes pairwise que incluyan un tratamiento etiquetado con un número par y otro etiquetado con un número impar no son estimables. Un tal diseño se denomina **diseño desconectado**.

Un diseño se dice **conectado** cuando cada tratamiento puede unirse con otro tratamiento cualquiera a través de las líneas de un grafo de conexión como el que aparece en la derecha de la Figura 5.1, donde se han unido mediante una línea cada par de tratamientos que ocurren juntos en cualquier bloque del diseño.

Figura 5.1: Ejemplos de un diseño desconectado (a la izquierda) y de un diseño conectado (a la derecha).

Todos los contrastes de efectos tratamiento son estimables en un diseño si y solo si el diseño está conectado.

Aunque los diseños desconectados pueden ser útiles en casos concretos de una sólo réplica, jamás deben utilizarse en experimentos que contengan un mínimo de dos observaciones por tratamiento.

Los tres tipos de diseños en bloques incompletos más simples y, frecuentemente más eficientes, son:

- **diseños en bloques incompletos balanceados,**
- **diseños divisibles en grupos,**
- **diseños cíclicos.**

### 5.3.2 Diseños en Bloques Incompletos Balanceados

Un *diseño en bloques incompletos balanceados* es un diseño con  $I$  tratamientos, cada uno de los cuales es replicado  $r$  veces, y con  $kb$  unidades experimentales agrupadas en  $b$  bloques de tamaño  $k < I$ , de tal forma que unidades experimentales en un mismo bloque son semejantes y unidades experimentales en distintos bloques son sustancialmente diferentes.

Además, el plan experimental del diseño satisface las siguientes condiciones:

1. El diseño es binario.
2. Cada par de etiquetas de tratamientos aparecen juntas en  $\lambda$  bloques, siendo  $\lambda$  un entero fijo.

Por último, la aleatorización del diseño se realiza siguiendo los dos pasos mencionados en la sección 5.3.1 y añadiendo un tercer paso que consistiría en asignar aleatoriamente las etiquetas de los tratamientos del plan a los niveles concretos del factor tratamiento.

Los diseños en bloques incompletos balanceados tienen muchas propiedades de interés. Así:

- todos los contrastes de tratamientos son estimables,
- todas las comparaciones de pares de efectos de tratamientos se estiman con la misma varianza,
- tienden a proporcionar los intervalos de confianza más cortos.

La Tabla 5.7 muestra un diseño en bloques incompletos balanceados para  $b = 7$ ,  $k = 3$ ,  $I = 7$  y  $r = 3$  (obsérvese que en este caso  $\lambda = 1$ ). Otro ejemplo, previo a la aleatorización, con  $b = 14$ ,  $k = 4$ ,  $r = 7$ ,  $I = 8$  y  $\lambda = 3$  se da en la Tabla 5.9.

| Bloque | Tratam. |   |   |   | Bloque | Tratam. |   |   |   |
|--------|---------|---|---|---|--------|---------|---|---|---|
| I      | 1       | 2 | 3 | 4 | VIII   | 2       | 3 | 5 | 8 |
| II     | 1       | 2 | 5 | 6 | IX     | 2       | 3 | 6 | 7 |
| III    | 1       | 2 | 7 | 8 | X      | 2       | 4 | 5 | 7 |
| IV     | 1       | 3 | 5 | 7 | XI     | 2       | 4 | 6 | 8 |
| V      | 1       | 3 | 6 | 8 | XII    | 3       | 4 | 5 | 6 |
| VI     | 1       | 4 | 5 | 8 | XIII   | 3       | 4 | 7 | 8 |
| VII    | 1       | 4 | 6 | 7 | XIV    | 5       | 6 | 7 | 8 |

Tabla 5.9: Plan experimental de un diseño en bloques incompletos balanceados con  $b = 14$ ,  $k = 4$ ,  $I = 8$ ,  $r = 7$  y  $\lambda = 3$ .

El mayor inconveniente de un diseño en bloques incompletos balanceados es que estos tipos de diseños no existen para determinadas selecciones de  $I$ ,  $k$ ,  $b$  y  $r$ . Se puede probar que, por ejemplo, no existe un diseño de este tipo para  $I = b = 8$  y  $r = k = 3$ .

Las tres condiciones necesarias (no suficientes) para la existencia de un diseño en bloques incompletos balanceado son:

$$\begin{aligned} Ir &= bk \\ r(k-1) &= \lambda(I-1) \\ b &\geq I \end{aligned}$$

### 5.3.3 Diseños Divisibles en Grupos

Un *diseño divisible en grupos* es un diseño con  $I = gl$  tratamientos ( $g > 1$  y  $l > 1$ ), cada uno de los cuales es replicado  $r$  veces, y con  $kb$  unidades experimentales agrupadas en  $b$  bloques de tamaño  $k < I$ , de tal forma que unidades experimentales en un mismo bloque son semejantes y unidades experimentales en distintos bloques son sustancialmente diferentes.

Además, el plan experimental del diseño satisface las siguientes condiciones:

1. Las  $I = gl$  etiquetas de los tratamientos se dividen en  $g$  grupos de  $l$  etiquetas. A las etiquetas pertenecientes a un mismo grupo se las denomina **primeros socios** y a aquellas etiquetas pertenecientes a grupos diferentes se las denomina **segundos socios**.
2. El diseño es binario.
3. Cada par de primeros socios aparecen juntos en  $\lambda_1$  bloques.
4. Cada par de segundos socios aparecen juntos en  $\lambda_2$  bloques.

Por último, la aleatorización del diseño se realiza siguiendo los dos pasos mencionados en la sección 5.3.1.

Puede probarse que los valores de  $\lambda_1$  y  $\lambda_2$  gobiernan las longitudes de los intervalos de confianza para los contrastes entre tratamientos.

Generalmente es preferible tomar  $\lambda_1$  y  $\lambda_2$  tan próximos como sea posible, pues ello garantiza que las regiones de rechazo para los contrastes dos a dos de los tratamientos sean de tamaños semejantes.

Por norma general, los diseños divisibles en grupos con  $\lambda_1$  y  $\lambda_2$  difiriendo en una unidad conducen a la mejor elección de un diseño en bloques incompletos cuando no existe un diseño en bloques incompletos balanceados.

El diseño de la Tabla 5.6 es un ejemplo de un diseño divisible en grupos con  $g = 4$  grupos de  $l = 2$  etiquetas: (1, 5), (2, 6), (3, 7) y (4, 8), siendo  $\lambda_1 = 0$  y  $\lambda_2 = 1$ .

Un segundo ejemplo con  $g = 4$  grupos de  $l = 3$  etiquetas:

$$(1, 2, 3) \quad (4, 5, 6) \quad (7, 8, 9) \quad (10, 11, 12)$$

y  $\lambda_1 = 3$  y  $\lambda_2 = 1$ , puede verse en la Tabla 5.10.

| Bloques | Tratamientos |   |   |    |    |    |
|---------|--------------|---|---|----|----|----|
| I       | 1            | 2 | 3 | 4  | 5  | 6  |
| II      | 1            | 2 | 3 | 7  | 8  | 9  |
| III     | 1            | 2 | 3 | 10 | 11 | 12 |
| IV      | 4            | 5 | 6 | 7  | 8  | 9  |
| V       | 4            | 5 | 6 | 10 | 11 | 12 |
| VI      | 7            | 8 | 9 | 10 | 11 | 12 |

Tabla 5.10: Plan experimental de un diseño grupos divisibles con  $b = k = 6$ ,  $I = 12$ ,  $r = 3$ ,  $\lambda_1 = 3$  y  $\lambda_2 = 1$ .

Las 4 condiciones necesarias (no suficientes) para garantizar la existencia de un diseño grupos divisibles son:

$$\begin{aligned}
 Ir &= bk \\
 r(k - 1) &= \lambda_1(l - 1) + \lambda_2l(g - 1) \\
 r &\geq \lambda_1 \\
 rk &\geq \lambda_2I
 \end{aligned}$$

Es importante tener en cuenta que los diseños divisibles en grupos con  $\lambda_2 = 0$  deben evitarse dado que en ellos no todos los contrastes son estimables (por ejemplo, el diseño desconectado de la Tabla 5.8 es en realidad un diseño divisible en grupos, con grupos (1, 3, 5, 7) y (2, 4, 6, 8) y con  $\lambda_1 = 2$  y  $\lambda_2 = 0$ ).

### 5.3.4 Diseños Cíclicos

Un *diseño cíclico* es un diseño con  $I$  tratamientos, cada uno de los cuales es replicado  $r$  veces, y con  $kb$  unidades experimentales agrupadas en  $b = I$  bloques de tamaño  $k < I$ , de tal forma que unidades experimentales en un mismo bloque son semejantes y unidades experimentales en distintos bloques son sustancialmente diferentes.

Además, el plan experimental del diseño satisface las siguientes condiciones:

1. El primer bloque, llamado **bloque inicial**, se forma con una selección de  $k$  etiquetas de tratamientos.
2. El segundo bloque se obtiene permutando cíclicamente las etiquetas de los tratamientos del bloque inicial, es decir, reemplazando la etiqueta número 1 por la número 2, la 2 por la 3, ..., la  $I - 1$  por la  $I$  y la  $I$  por la 1. Cada uno de los restantes bloques se generan del inmediatamente anterior siguiendo el mismo procedimiento. El algoritmo termina al formar el  $I$ -ésimo bloque.

Por último, la aleatorización del diseño se realiza siguiendo los dos pasos mencionados en la sección 5.3.1.

El diseño divisible en grupos de la Tabla 5.6 es también un diseño cíclico con bloque inicial (1, 2, 4). Otros dos ejemplos pueden verse en la Tabla 5.11.

| Diseño 1: $I = 7$ |         |   |   | Diseño 2: $I = 6$ |        |         |   |   |   |
|-------------------|---------|---|---|-------------------|--------|---------|---|---|---|
| Bloque            | Tratam. |   |   |                   | Bloque | Tratam. |   |   |   |
| I                 | 1       | 2 | 3 | 6                 | I      | 1       | 2 | 3 | 6 |
| II                | 2       | 3 | 4 | 7                 | II     | 2       | 3 | 4 | 1 |
| III               | 3       | 4 | 5 | 1                 | III    | 3       | 4 | 5 | 2 |
| IV                | 4       | 5 | 6 | 2                 | IV     | 4       | 5 | 6 | 3 |
| V                 | 5       | 6 | 7 | 3                 | V      | 5       | 6 | 1 | 4 |
| VI                | 6       | 7 | 1 | 4                 | VI     | 6       | 1 | 2 | 5 |
| VII               | 7       | 1 | 2 | 5                 |        |         |   |   |   |

Tabla 5.11: Dos ejemplos de planes experimentales de diseños cíclicos con bloque inicial (1, 2, 3, 6),  $k = 4$  y  $r = 4$ . En el primero (a la izquierda)  $I = 7$  mientras que en el segundo (a la derecha)  $I = 6$ .

Un diseño cíclico tendrá, como mucho,  $I/2$  diferentes varianzas para los estimadores de los contrastes pairwise.

Algunos diseños cíclicos tienen bloques duplicados, en cuyo caso los repetidos pueden ser ignorados y son así especialmente útiles cuando se requieren menos de  $I$  bloques. En otro caso es preferible usar  $I$  bloques.

### 5.3.5 Análisis de un Diseño General en Bloques Incompletos

La formulación del modelo matemático estándar para un diseño general en bloques incompletos es exactamente igual que la del modelo en bloques completos, con la salvedad de que no toda combinación bloque-tratamiento pertenece al diseño. O sea, el modelo es:

$$\boxed{\underbrace{Y_{hi}}_{\text{aleatorio}} = \overbrace{\mu + \theta_h + \tau_i}^{\text{determinista}} + \underbrace{\varepsilon_{hi}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{hi} \text{ i.i.d. según una } N(0, \sigma)} \quad (5.4)$$

para  $h = 1, \dots, b$ ,  $i = 1, \dots, I$ , **con (h,i) en el diseño**.

De modo que, en (5.4),  $\mu$  es una constante,  $\theta_h$  representa el efecto (positivo o negativo) sobre la respuesta debido a que se observa el bloque  $h$ ,  $\tau_i$  representa el efecto (positivo o negativo) sobre la respuesta debido a que se observa el nivel  $i$  del factor tratamiento e  $Y_{hi}$  y  $\varepsilon_{hi}$  son las variables aleatorias que miden respectivamente la respuesta y su error asociado en la observación del  $i$ -ésimo tratamiento en el  $h$ -ésimo bloque.

El modelo (5.4) se conoce como **modelo tratamiento-bloque para un diseño en bloques incompletos**. Obsérvese que:

[1] Dado que no todos los tratamientos aparecen en cada bloque resulta complicado chequear la hipótesis de no interacción entre bloques y tratamientos. Por ello, este modelo es solo recomendable cuando existen evidencias serias de no interacción.

[2] Los estimadores mínimo cuadráticos de los parámetros del modelo relativos a los tratamientos deben incluir un *ajuste por bloques*. Esto es intuitivamente razonable debido a que algunos tratamientos pudieran ser observados en *mejores bloques* que otros.

En otras palabras, pudiera ocurrir que el estimador por mínimos cuadrados de, por ejemplo, el contraste  $\tau_p - \tau_i = 0$  no fuese el estimador “*desajustado*”  $\bar{Y}_{.p} - \bar{Y}_{.i}$  (estimador mínimo-cuadrático en un diseño en bloques completos).

Por ejemplo, en el diseño de la Tabla 5.7, donde  $k = 3$ , los tratamientos 2 y 7 se observan en los bloques 1, 2 y 4 (el 2) y 1, 6 y 7 (el 7). Entonces:

$$\begin{aligned} E(\bar{Y}_{.2} - \bar{Y}_{.7}) &= E\left(\frac{1}{3}(Y_{12} + Y_{22} + Y_{42}) - \frac{1}{3}(Y_{17} + Y_{67} + Y_{77})\right) \\ &= \frac{1}{3}(3\mu + \theta_1 + \theta_2 + \theta_4 + 3\tau_2 - 3\mu - \theta_1 - \theta_6 - \theta_7 - 3\tau_7) \\ &= (\tau_2 - \tau_7) + \frac{1}{3}(\theta_2 + \theta_4 - \theta_6 - \theta_7) \end{aligned}$$

De hecho, al resolver el problema de minimización de la suma de los cuadrados de los errores en un diseño en bloques incompletos arbitrario, los estimadores mínimo-cuadráticos  $\hat{\tau}_i$  que se obtienen son los estándar pero ajustados por bloques. El sistema de ecuaciones que resulta es:

$$r(k-1)\hat{\tau}_i - \sum_{p \neq i} \lambda_{pi} \hat{\tau}_p = kQ_i, \quad i = 1, \dots, I \quad (5.5)$$

donde:

- $\lambda_{pi}$  denota el número de bloques que contienen a los tratamientos  $p$  e  $i$ ,
- $Q_i$  se denomina *ajuste total del tratamiento  $i$ -ésimo* y viene dado por:

$$Q_i = T_i - \frac{1}{k} \sum_{h=1}^b n_{hi} B_h \tag{5.6}$$

con

$$T_i = \sum_{h=1}^b n_{hi} y_{hi} = \text{suma de todas las observaciones del tratamiento } i,$$

$$B_h = \sum_{i=1}^I n_{hi} y_{hi} = \text{suma de todas las observaciones en el bloque } h.$$

De modo que el término  $\sum_{h=1}^b n_{hi} B_h$  contabiliza la suma de todas las observaciones pertenecientes a bloques conteniendo al tratamiento  $i$ -ésimo.

Resolver el sistema de ecuaciones (5.5) resulta en general muy complicado. En cambio, tiene soluciones particulares sencillas con diseños en bloques incompletos balanceados y con diseños divisibles en grupos.

### Descomposición de la Variabilidad

Se demuestra que si  $G$  denota el doble sumatorio  $\sum_{h=1}^b \sum_{i=1}^I n_{hi} y_{hi}$ , entonces:

$$\underbrace{\underbrace{\text{Suma de Cuadrados Residual}}_{scR}}_{g.l. = bk - I - b + 1} = \underbrace{\underbrace{\left( \sum_{h=1}^b \sum_{i=1}^I n_{hi} y_{hi}^2 - \frac{1}{bk} G^2 \right)}_{\text{Suma de Cuadrados Global}}}_{g.l. = bk - 1} - \underbrace{\underbrace{\left( \frac{1}{k} \sum_{h=1}^b B_h^2 - \frac{1}{bk} G^2 \right)}_{\text{Suma de Cuadrados Explicada por Bloques}}}_{g.l. = b - 1} - \underbrace{\underbrace{\left( \sum_{i=1}^I Q_i \hat{\tau}_i \right)}_{\text{Suma de Cuadrados Explicada por Tratam. Ajustados}}}_{g.l. = I - 1}$$

### Análisis de la Varianza

Cuando la hipótesis nula  $H_0^T : \{\tau_i = \tau, \forall i\}$  es cierta, el modelo reducido que se obtiene a partir del modelo matemático tratamiento-bloque dado en (5.4) es semejante a un modelo de análisis de la varianza de una vía donde los tratamientos serían los bloques; es decir:

$$Y_{hi} = \mu^* + \theta_h + \varepsilon_{hi}$$

siendo  $\mu^* = \mu + \tau$ . Nótese que aquí existen  $k$  observaciones en cada uno de los  $b$  bloques. Por consiguiente, la suma de cuadrados residual en el modelo reducido,  $scR_0$ , será semejante a la suma de cuadrados residual en un diseño completamente aleatorizado de una vía, reemplazando  $n_i$  por  $k$  para todo  $i$ . Adaptando las correspondientes expresiones resulta que:

$$\begin{aligned} scR_0 &= \sum_{h=1}^b \sum_{i=1}^I n_{hi} y_{hi}^2 - \frac{1}{k} \sum_{h=1}^b B_h^2 \\ &= \left( \sum_{h=1}^b \sum_{i=1}^I n_{hi} y_{hi}^2 - \frac{1}{bk} G^2 \right) - \left( \frac{1}{k} \sum_{h=1}^b B_h^2 - \frac{1}{bk} G^2 \right) \end{aligned}$$

De modo que, la suma de cuadrados para los tratamientos ajustados por los bloques es:

$$scT_{Ajust.} = scR_0 - scR = \sum_{i=1}^I Q_i \hat{\tau}_i \quad (5.7)$$

donde  $\hat{\tau}_i$  son las soluciones del sistema (5.5) y  $Q_i$  están dados en (5.6). Como el número de grados de libertad para los tratamientos es  $I - 1$ , los grados de libertad para el error son:  $(bk - 1) - (b - 1) - (I - 1) = bk - b - I + 1$ .

Así, se rechaza la hipótesis nula  $H_0^T : \{\tau_i = \tau, \forall i\}$  a un nivel de significación  $\alpha$  siempre que

$$\frac{scmT_{Ajust.}}{scmR} = \frac{scT_{Ajust.}/(I - 1)}{scR/(bk - b - I + 1)} > F_{I-1, bk-b-I+1, \alpha}$$

En el supuesto de que se precisase evaluar la significación de los bloques para tenerlo en cuenta en experimentos futuros, la cantidad *suma de cuadrados explicada por bloques* dada en la descomposición de variabilidad anterior, **no** es el valor adecuado para resolver el contraste de significación de los bloques. El motivo es bien simple: tal cantidad no ha sido ajustada por el hecho de que cada bloque no contiene una observación de cada tratamiento. De igual forma que se ha hecho para los tratamientos habría que calcular la *suma de cuadrados explicada por bloques ajustados*. Algunos paquetes estadísticos proporcionan este valor bajo el epígrafe *suma de cuadrados "ajustada" o "Tipo III"*. La fórmula es similar a la de la suma de cuadrados explicada por tratamientos ajustados sin más que intercambiar los papeles de bloques y tratamientos. O sea:

$$scB_{Ajust.} = scG - scR - scT_{NoAjust.} = scG - scR - \left( \frac{1}{r} \sum_{i=1}^I T_i^2 - \frac{1}{Ir} G^2 \right) \quad (5.8)$$

| <b>CUADRO DEL ANÁLISIS DE LA VARIANZA<br/>— MODELO TRATAMIENTO-BLOQUE DE UN<br/>DISEÑO GENERAL EN BLOQUES INCOMPLETOS —</b>  |  |                  |  |                              |
|--|--|------------------|--|------------------------------|
| <i>Fuente de Variación</i>   | <i>Suma de Cuadrados</i>   | <i>g.l.</i>      | <i>scm</i>   | $\hat{F}$                    |
| Bloques  | $scB = \frac{1}{k} \sum_{h=1}^b B_h^2 - \frac{G^2}{bk}$                                    | $b - 1$          | —  | —                            |
| Tratam. Ajust.   | $scT_{Ajust.} = \sum_{i=1}^I Q_i \hat{\tau}_i$   | $I - 1$          | $\frac{scmT_{Ajust.}}{I - 1} = \frac{scT_{Ajust.}}{I - 1}$ | $\frac{scmT_{Ajust.}}{scmR}$ |
| Residual   | $scR = scG - scB - scT_{Ajust.}$   | $bk - b - I + 1$ | $\frac{scmR}{bk - b - I + 1} = \frac{scR}{bk - b - I + 1}$ | —                            |
| Global   | $scG = \sum_{h=1}^b \sum_{i=1}^I n_{hi} y_{hi}^2 - \frac{G^2}{bk}$                         | $bk - 1$         | $\frac{scmG}{bk - 1} = \frac{scG}{bk - 1}$                 | —                            |
| Bloques Ajust.   | $scB_{Ajust.} = scG - scR - \left( \frac{1}{r} \sum_{h=1}^b T_h^2 - \frac{G^2}{k} \right)$ | $b - 1$          | $\frac{scmB_{Ajust.}}{b - 1} = \frac{scB_{Ajust.}}{b - 1}$ | —                            |
| Rechazar $H_0^T : \tau_1 = \tau_2 = \dots = \tau_I$ , según $p = P \left( \frac{scmT_{Ajust.}}{scmR} > F_{I-1, bk-b-I+1, \alpha} \right)$<br><br>Comparar $scmB_{Ajust.}$ con $scmR$ al objeto de evaluar el peso de los bloques |  |                  |  |                              |

Tabla 5.12: Cuadro del análisis de la varianza para un diseño en bloques incompletos binario con  $b$  bloques de tamaño  $k$  con  $I$  tratamientos  $r$  veces replicados.

### Pruebas de Hipótesis Múltiples

En relación con las pruebas de hipótesis de rango múltiple indicar tan solo que, con la estimación adecuada de la varianza y los grados de libertad apropiados, se tiene:

- Los métodos de Scheffé y Bonferroni son válidos para cualquier diseño en bloques incompletos.

- El procedimiento de Tukey sólo debe emplearse en diseños en bloques incompletos balanceados (en otro caso se sospecha que es conservador, si bien no se ha probado).
- Por último, el método de Dunnett es válido únicamente en diseños en bloques incompletos balanceados.

### 5.3.6 Análisis de un Diseño en Bloques Incompletos Balanceados

La solución mínimo cuadrática para un diseño general en bloques incompletos viene dada por la solución al sistema de ecuaciones (5.5). En el caso particular de un diseño en bloques incompletos balanceados  $\lambda_{pi} = \lambda$  para todo  $p$  e  $i$ , puesto que cada par de etiquetas de tratamientos aparecen juntas en un igual número  $\lambda$  de bloques. Entonces (5.5) pasa a ser de la forma:

$$r(k-1)\hat{\tau}_i - \lambda \sum_{p \neq i} \hat{\tau}_p = kQ_i, \quad i = 1, \dots, I \quad (5.9)$$

Suponiendo que  $\sum_{i=1}^I \hat{\tau}_i = 0$ , se sigue

$$(r(k-1) + \lambda) \hat{\tau}_i = kQ_i, \quad i = 1, \dots, I$$

Además, una condición necesaria para la existencia de un diseño en bloques incompletos balanceados es que  $r(k-1) = \lambda(I-1)$ , o equivalentemente,  $r(k-1) + \lambda = \lambda I$ , que permite concluir:

$$\lambda I \hat{\tau}_i = kQ_i \quad \Rightarrow \quad \boxed{\hat{\tau}_i = \frac{k}{\lambda I} Q_i, \quad i = 1, \dots, I} \quad (5.10)$$

Por tanto, el estimador mínimo cuadrático del contraste  $\sum_{i=1}^I c_i \tau_i$ , con  $\sum_{i=1}^I c_i = 0$ , y su varianza son:

$$\sum_{i=1}^I c_i \hat{\tau}_i = \frac{k}{\lambda I} \sum_{i=1}^I c_i Q_i \quad \text{y} \quad \text{Var} \left( \sum_{i=1}^I c_i \hat{\tau}_i \right) = \frac{k}{\lambda I} \sigma^2 \sum_{i=1}^I c_i^2 \quad (5.11)$$

Las pruebas de hipótesis múltiples proporcionarán intervalos de confianza de la forma:

$$\sum_{i=1}^I c_i \tau_i \in \left( \frac{k}{\lambda I} \sum_{i=1}^I c_i Q_i \mp \omega \sqrt{\frac{k}{\lambda I} scm R \sum_{i=1}^I c_i^2} \right)$$

con  $\omega$  calculado en base a  $bk - b - I + 1$  grados de libertad.

### 5.3.7 Análisis de un Diseño Divisible en Grupos

Ahora las  $I$  etiquetas de tratamientos se dividen en  $g$  grupos de  $l$  etiquetas y de acuerdo al plan experimental de este tipo de diseños:

$$\lambda_{pi} = \begin{cases} \lambda_1 & \text{si los tratamientos } p \text{ e } i \text{ son primeros socios} \\ \lambda_2 & \text{si los tratamientos } p \text{ e } i \text{ son segundos socios} \end{cases}$$

Se prueba que en tal caso:

$$\hat{\tau}_i = \left( \frac{k}{(r(k-1) + \lambda_1) I \lambda_2} \right) \left[ (I \lambda_2 + (\lambda_1 - \lambda_2)) Q_i + (\lambda_1 - \lambda_2) \sum_{(1)} Q_p \right],$$

donde  $\sum_{(1)} Q_p$  denota la suma de los  $Q_p$  correspondientes a etiquetas de tratamientos que son primeros socios de la etiqueta del tratamiento  $i$ .

Se demuestra también que el estimador mínimo cuadrático de un contraste estimable  $\sum_{i=1}^I c_i \tau_i$  tiene varianza:

$$Var \left( \sum_{i=1}^I c_i \hat{\tau}_i \right) = \sum_{i=1}^I c_i^2 Var(\hat{\tau}_i) + 2 \sum_{i=1}^{I-1} \sum_{p=i+1}^I c_i c_p Cov(\hat{\tau}_i, \hat{\tau}_p)$$

donde

$$Var(\hat{\tau}_i) = \frac{k [I \lambda_2 + (\lambda_1 - \lambda_2)] \sigma^2}{I \lambda_2 [I \lambda_2 + l (\lambda_1 - \lambda_2)]}$$

$$Cov(\hat{\tau}_i, \hat{\tau}_p) = \begin{cases} \frac{k (\lambda_1 - \lambda_2) \sigma^2}{I \lambda_2 [I \lambda_2 + l (\lambda_1 - \lambda_2)]}, & \text{si } p \text{ e } i \text{ son primeros socios} \\ 0, & \text{si } p \text{ e } i \text{ son segundos socios} \end{cases}$$

Por último, la varianza del estimador mínimo cuadrático del contraste pairwise  $\tau_i - \tau_p$  es:

$$Var(\hat{\tau}_i - \hat{\tau}_p) = \begin{cases} \frac{2k \sigma^2}{[I \lambda_2 + l (\lambda_1 - \lambda_2)]}, & \text{si } p \text{ e } i \text{ son primeros socios} \\ \frac{2k [I \lambda_2 + (\lambda_1 - \lambda_2)] \sigma^2}{I \lambda_2 [I \lambda_2 + l (\lambda_1 - \lambda_2)]}, & \text{si } p \text{ e } i \text{ son segundos socios} \end{cases}$$

### 5.3.8 Análisis de un Diseño Cíclico

Los diseños cíclicos son diseños en bloques incompletos que pueden o no poseer las propiedades de los diseños en bloques incompletos balanceados o los diseños divisibles en grupos. En cualquiera de los dos primeros casos el análisis ya ha sido abordado y, en otro supuesto, los estimadores mínimo cuadráticos no tienen expresiones sencillas recomendándose su análisis a través de algún paquete estadístico.

### 5.3.9 Tamaños Muestrales

Se trata de determinar el número  $b$  de bloques apropiado para alcanzar una potencia determinada en el test Anova, una vez que se han fijado el número de tratamientos  $I$  y el tamaño  $k$  de los bloques. El proceder adecuado para un diseño en bloques incompletos balanceados se describe en los ejemplos que siguen, donde se calculan  $b$  y el número de réplicas para cada tratamiento  $r = bk/I$ . Como entre los diseños en bloques incompletos el de bloques incompletos balanceados es el más eficiente, cualquier otro diseño en bloques incompletos requerirá en general valores de  $b$  y  $r$  algo más grandes.

**Ejemplo 5.7** Supóngase que se desea utilizar el método de Tukey para las comparaciones pairwise en un experimento que se diseñará mediante un diseño en bloques incompletos con  $I = 5$  tratamientos y tamaño de bloque  $k = 3$ . De experiencias previas se conoce que es muy poco probable que la varianza residual del experimento supere las 2 unidades al cuadrado. Los investigadores desean que los intervalos de confianza simultáneos obtenidos mediante las pruebas de Tukey tengan longitudes no superiores a tres unidades (si el diseño es en bloques incompletos balanceados todos tendrán idéntica longitud).

Por tratarse de un diseño en bloques incompletos balanceados se ha de verificar que  $\lambda = r(k - 1)/(I - 1)$  y, por tratarse de un diseño en bloques,  $b = Ir/k$ . Por consiguiente los grados de libertad del error pueden expresarse como:

$$g.l.(\text{error}) = bk - b - I + 1 = Ir - \frac{Ir}{k} - I + 1 = \frac{Ir(k - 1) - k(I - 1)}{k} = \frac{10r - 12}{3}$$

La mínima diferencia significativa ( $msd$ ) al nivel de significación global  $\alpha$  con la prueba de Tukey viene dada por:

$$msd = \frac{q_{I, g.l.(\text{error}), \alpha}}{\sqrt{2}} \sqrt{2 \left( \frac{k(I - 1)}{rI(k - 1)} \right) scmR} = q_{5, g.l.(\text{error}), 0'05} \sqrt{\frac{12}{5r}},$$

con  $g.l.(\text{error}) = (10r/3) - 4$ . Como se requiere que  $msd \leq 1'5$ , se concluye que

$$r \geq 1'0667 q_{5, g.l.(\text{error}), 0'05}^2$$

Mediante sucesivas pruebas de ensayo y error se demuestra que la desigualdad anterior sería válida con alrededor de 17–18 observaciones por tratamiento, es decir, 85–90 observaciones en total. Ello a su vez requeriría 28–30 bloques de tamaño 3. Dado que existe un diseño en bloques incompletos balanceados con parámetros  $I = 5$ ,  $k = 3$ ,  $b = 10$  y  $r = 6$ , bastaría repetir el plan experimental entero de este diseño tres veces para obtener un diseño en bloques incompletos balanceados con  $r = 18$ . En tal caso se obtendrían intervalos de confianza simultáneos para los contrastes pairwise por el procedimiento de Tukey de longitud:

$$2q_{5,56,0'05} \sqrt{12/(15 \times 8)} \approx 2'92 < 3$$

**Ejemplo 5.8** Supóngase ahora que se desea contrastar, al 5% de significación, la hipótesis nula  $H_0 : \{\text{todos los } \tau_i \text{ son iguales}\}$ , de modo que se detecte una diferencia real en los efectos tratamiento de  $\Delta = 1$  unidad con probabilidad 0'95, utilizando un diseño en bloques incompletos balanceados con  $I = 5$  y  $k = 3$ .

El estimador mínimo-cuadrático,  $\hat{\tau}_i - \hat{\tau}_p$ , del contraste de la comparación apareada  $\tau_i - \tau_p$  (así como su varianza) en un diseño en bloques incompletos balanceados se deduce de la expresión general dada en (5.11) haciendo  $c_i = 1$ ,  $c_p = -1$  y  $c_j = 0, \forall j \neq i, p$ . En particular su varianza resulta:

$$Var(\hat{\tau}_i - \hat{\tau}_p) = 2 \frac{k}{\lambda I} \sigma^2 = 2 \left( \frac{k(I-1)}{rI(k-1)} \right) \sigma^2 = \frac{12}{5r} \sigma^2 \quad (5.12)$$

El número  $r$  de observaciones necesarias por tratamiento se calcula entonces de un modo similar a (3.30), reemplazando  $a$  por  $I$  y  $2\sigma^2/b$  por la varianza (5.12); es decir

$$r = \frac{2I\sigma^2\phi^2}{\Delta^2} \left( \frac{k(I-1)}{I(k-1)} \right)$$

Por ejemplo, si se cree que  $\sigma^2$  no supera la unidad al cuadrado; entonces  $r = 12\phi^2$ . Entoces se usan las tablas de potencia del Apéndice para encontrar  $\phi^2$ . Los grados de libertad del numerador son  $\nu_1 = I - 1 = 4$  y  $\nu_2$  son los grados de libertad del error dados por  $(10r - 12)/3$ . Las pruebas de ensayo y error en base a las tablas conducen a que el valor de  $r$  apropiado es 48. Como existe un diseño en bloques incompletos balanceados con  $I = 5$ ,  $k = 3$ ,  $b = 10$  y  $r = 6$ , basta repetir el plan experimental entero de este diseño ocho veces para obtener un diseño en bloques incompletos balanceados con  $r = 48$ .

## 5.4 Diseños con Dos Factores Bloque sin Interacción

### 5.4.1 Definiciones y Consideraciones Generales: Diseños Fila-Columna

En esta sección se considera un experimento con dos factores bloque y un factor tratamiento de tal forma que se puede asumir que no existen interacciones entre todos ellos. En tal caso, supuesta la ausencia de interacciones, el diseño puede planificarse con una única observación para cualquier combinación de niveles de cada factor bloque (en adelante celda).

El plan se concibe como una **matriz**, o sea, una tabla, con los niveles de un factor bloque encabezando las filas de la matriz y los niveles del otro factor bloque encabezando las columnas. A continuación, se ubica una y sólo una etiqueta de tratamiento en cada posición o celda de la matriz.

Estos diseños se conocen con el nombre de **diseños fila-columna**, denominando a los factores bloque **bloque fila** y **bloque columna**, respectivamente.

En muchos diseños fila-columna **la totalidad** de los contrastes de tratamientos son estimables. En particular, esta afirmación es correcta en los diseños en **Cuadrado Latino** y en los diseños **Youden**, dos de los diseños fila-columna más populares.

**Plan Experimental y Aleatorización.** Igual que en los diseños en bloques incompletos, una vez se determinan los números de filas, columnas y tratamientos, existen dos etapas en la planificación de un experimento fila-columna:

- Determinar un plan experimental como el de la Tabla 5.13.

|             |    | Bloque<br>Columna |    |    |    |
|-------------|----|-------------------|----|----|----|
|             |    | C1                | C2 | C3 | C4 |
| Bloque Fila | F1 | 1                 | 3  | 6  | 7  |
|             | F2 | 2                 | 4  | 7  | 1  |
|             | F3 | 3                 | 5  | 1  | 2  |
|             | F4 | 4                 | 6  | 2  | 3  |
|             | F5 | 5                 | 7  | 3  | 4  |
|             | F6 | 6                 | 1  | 4  | 5  |
|             | F7 | 7                 | 2  | 5  | 6  |

Tabla 5.13: Plan experimental de un diseño fila-columna con  $b = 7$  niveles del bloque fila,  $c = 4$  niveles del bloque columna y  $I = 7$  niveles del factor tratamiento replicados  $r = 4$  veces.

• Aleatorizar la asignación de etiquetas del plan a los niveles reales de los factores tratamiento, fila y columna. Esto se realiza como sigue:

- i) Las etiquetas de las filas del plan se asignan al azar a los niveles del bloque fila.
- ii) Las etiquetas de las columnas del plan se asignan al azar a los niveles del bloque columna.
- iii) Las etiquetas de los tratamientos del factor tratamiento en el plan se asignan al azar a los niveles reales del factor tratamiento.

#### 5.4.2 Diseños en Cuadrado Latino

Un *cuadrado latino*  $I \times I$  es una disposición de  $I$  letras latinas en una matriz  $I \times I$  de tal modo que todas las letras aparecen una vez en cada fila y una vez en cada columna.

Por ejemplo, un cuadrado latino  $3 \times 3$  está dado en la Tabla 5.14.

|   |   |   |
|---|---|---|
| A | B | C |
| B | C | A |
| C | A | B |

Tabla 5.14: Cuadrado latino  $3 \times 3$ .

Un *diseño en cuadrado latino* es un diseño con  $I$  tratamientos e  $I^2$  unidades experimentales agrupadas en  $b = I$  bloques fila y  $c = I$  bloques columna, de tal forma que unidades experimentales en un mismo bloque fila son semejantes, unidades experimentales en un mismo bloque columna son semejantes y unidades experimentales en distintos bloques fila y distintos bloques columna son sustancialmente diferentes.

Además, el plan experimental del diseño es un cuadrado latino  $I \times I$ .

Por último, la aleatorización del diseño se realiza siguiendo los dos pasos mencionados en la sección 5.4.1.

Obsérvese que si en un diseño en cuadrado latino ignoramos el bloque columna, tendremos un diseño en bloques completamente aleatorizado (el bloque fila sería el factor bloque) y, análogamente, si ignoramos el bloque fila, tendremos un diseño en bloques completamente aleatorizados (el bloque columna sería el factor bloque). Además se trata de un diseño equireplicado ya que cada tratamiento aparece un mismo número  $I$  de veces en el diseño.

Un cuadrado latino es un **cuadrado latino estándar** cuando las letras de la primera fila y de la primera columna están dispuestas en orden alfabético.

Un cuadrado latino es un **cuadrado latino cíclico** si las letras de cada fila se generan cíclicamente de la anterior según el orden alfabético.

Por ejemplo, la Tabla 5.14 muestra un diseño en cuadrado latino  $3 \times 3$  estándar y cíclico.

Existe un único diseño en cuadrado latino  $3 \times 3$  estándar, sin embargo hay cuatro diseños en cuadrado latino  $4 \times 4$  estándar cuyos planes experimentales se muestran en la Tabla 5.15.

| Cuadro 1 |   |   |   | Cuadro 2 |   |   |   | Cuadro 3 |   |   |   | Cuadro 4 |   |   |   |
|----------|---|---|---|----------|---|---|---|----------|---|---|---|----------|---|---|---|
| A        | B | C | D | A        | B | C | D | A        | B | C | D | A        | B | C | D |
| B        | C | D | A | B        | A | D | C | B        | A | D | C | B        | D | A | C |
| C        | D | A | B | C        | D | A | B | C        | D | B | A | C        | A | D | B |
| D        | A | B | C | D        | C | B | A | D        | C | A | B | D        | C | B | A |

Tabla 5.15: Cuatro posibles diseños en cuadrado latino  $4 \times 4$  estándar.

Obviamente, para cualquier número de tratamientos  $I$  existe siempre al menos un diseño en cuadrado latino estándar cíclico.

Los diseños en cuadrado latino son utilizados a menudo cuando se evalúa información proporcionada por distintos tratamientos sobre los mismos individuos en diferentes períodos de tiempo. Por ejemplo, se desean comparar los efectos de  $I$  drogas y se toman como filas del cuadrado los individuos sometidos a los tratamientos y como columnas los  $I$  períodos de tiempo durante los cuales cada individuo recibe la droga y proporciona el resultado.

Para que cada cuadrado latino tenga la misma posibilidad de ser elegido como plan experimental, se suele elegir al azar de entre todos los posibles.

### Replicar un diseño en cuadrado latino.

Existen dos vías para aumentar el tamaño del experimento replicando un cuadrado latino.

La primera da lugar al llamado **diseño en cuadrado latino  $s$ -replicado**, que consiste en unir verticalmente  $s$  cuadrados latinos  $I \times I$ , apilando sucesivamente uno debajo de otro. Naturalmente, usar este diseño requiere que los bloques columna sean de tamaño  $Is$ , de modo que: el número de veces que se repite cada tratamiento será  $r = Is$ , el número de bloques fila será  $Is$  y el número de bloques columna será  $c$ .

Los dos planes de la Tabla 5.16 son dos ejemplos de diseños en cuadrado latino 2-replicados.

| Plan 1 |   |   | Plan 2 |   |   |
|--------|---|---|--------|---|---|
| A      | B | C | A      | B | C |
| B      | C | A | B      | C | A |
| C      | A | B | C      | A | B |
| A      | B | C | A      | C | B |
| B      | C | A | B      | A | C |
| C      | A | B | C      | B | A |

Tabla 5.16: Dos planes experimentales distintos de diseños en cuadrados latinos 2-replicados.

Otra forma alternativa de proceder para obtener  $r = Is$  observaciones por tratamiento consiste en generar  $s$  cuadrado latino separados. La desventaja de este último procedimiento con respecto al primero es que con el mismo número de observaciones no es factible estimar interacciones. Por ejemplo, si se toman los cuadros 1 y 2 de la Tabla 5.15 y se usan por separado, se genera un diseño donde cada tratamiento se observa  $r = 8$  veces y se emplean  $b = 8$  bloques fila y  $c = 8$  bloques columna. Si, sin embargo, se apilan ambos cuadros construyendo un diseño 2-replicado, el resultado es un diseño donde de nuevo cada tratamiento se observa  $r = 8$  veces, pero empleando  $b = 8$  bloques fila y  $c = 4$  bloques columna.

### 5.4.3 Diseños Youden

Un *cuadrado Youden*  $I \times c$  es una disposición de  $I$  letras latinas en una matriz  $I \times c$ , con  $c < I$ , de tal modo que cada letra aparece una vez en cada columna y a lo sumo una vez en cada fila.

Análogamente podría definirse un cuadrado Youden  $c \times I$ , con  $c < I$ . Obsérvese que aunque se denomina cuadrado Youden no es un cuadrado. Por ejemplo, un cuadrado Youden  $4 \times 3$  es el dado en la Tabla 5.17.

| Plan |   |   |
|------|---|---|
| A    | B | C |
| B    | C | D |
| C    | D | A |
| D    | A | B |

Tabla 5.17: Un plan experimental en forma de cuadrado Youden  $4 \times 3$ .

Un *diseño Youden* es un diseño con  $I$  tratamientos e  $I^2$  unidades experimentales agrupadas en  $b = I$  bloques fila y  $c < I$  bloques columna, de tal forma que unidades experimentales en un mismo bloque fila son semejantes, unidades experimentales en un mismo bloque columna son semejantes y unidades experimentales en distintos bloques fila y distintos bloques columna son sustancialmente diferentes.

Además, los bloques columna forman un diseño en bloques completamente aleatorizado y los bloques fila forman un diseño en bloques incompletos balanceados. Cada nivel del factor tratamiento se observa  $r = c$  veces (una vez en cada columna y a lo sumo una vez en cada fila).

Así el plan experimental del diseño es un cuadrado Youden  $I \times c$ .

Por último, la aleatorización del diseño se realiza siguiendo los dos pasos mencionados en la sección 5.4.1.

Este tipo de diseño es adecuado para aquellas situaciones donde no es posible tomar un tamaño de bloque tan grande como el número de tratamientos en alguno de los dos factores de bloqueo.

Por ejemplo, el plan de la Tabla 5.17 describe un diseño Youden cíclico. Los bloques filas son en realidad un diseño en bloques incompletos balanceados cíclico donde todos los pares de tratamientos ocurren conjuntamente en  $\lambda = 2$  filas.

Cualquier diseño en bloques incompletos balanceados cíclico puede usarse como un diseño Youden cíclico.

Basta con borrar una columna de un diseño en cuadrado latino  $I \times I$  para obtener un diseño Youden con  $C = I - 1$  bloques columna.

### Replicar un diseño Youden.

Los dos procedimientos analizados para replicar un diseño en cuadrado latino son válidos para los diseños Youden.

Así, podremos analizar por separado varios cuadrados Youden de las mismas dimensiones para obtener  $r = cs$  observaciones de cada tratamiento por medio de  $b = sI$  bloques fila y  $sc$  bloques columna, o bien, apilar verticalmente  $s$  cuadrados Youden obteniendo de nuevo  $r = cs$  observaciones de cada tratamiento por medio de  $b = sI$  bloques fila y  $c$  bloques columna. Como antes, el segundo procedimiento de réplica permitiría contrastar interacciones. En cualquiera de los dos casos, los bloques fila seguirían formando un diseño en bloques incompletos balanceados y los bloques columna un diseño en bloques completos.

#### 5.4.4 Diseños Cíclicos y otros Diseños Fila-Columna

Cualquier disposición de  $I$  etiquetas de tratamiento en una tabla con  $b$  filas y  $c$  columnas puede utilizarse como un diseño fila-columna en un experimento con dos factores de bloqueo. Pero, igual que ocurría con los diseños en bloques incompletos, unos diseños fila-columna serán mejores que otros. Como con aquellos, serán preferidos los que conduzcan a intervalos de confianza para los contrastes pairwise con menor e igual longitud, es decir, los que tienen cada par de tratamientos ocurriendo conjuntamente un mismo número de veces en los bloques fila y en los bloques columna.

Esta característica la verifican los diseños en cuadrado latino, los diseños Youden y algunos diseños cíclicos.

Un *diseño fila-columna cíclico con bloques columna completos* es un diseño en el cual los bloques fila forman un diseño en bloques cíclico y los bloques columna son bloques completos.

La clase de diseños fila-columna cíclicos es enorme y presenta la ventaja de que siempre es posible encontrar uno con  $b = Is$  filas cuando no existe un diseño Youden.

Por ejemplo, no existe un diseño Youden para  $I = b = 8$  y  $c = 3$  por un motivo bien sencillo: no existe un diseño en bloques incompletos balanceados para 8 tratamientos en 8 bloques de tamaño 3, ya que  $\lambda = r(c - 1)/(I - 1) = 6/7$  no es un entero. Sin embargo, el primer plan de la Tabla 5.18 proporciona un diseño fila-columna cíclico que valdría como posible solución para estos parámetros. En efecto, el diseño se ha generado cíclicamente a partir de la primera fila, cada tratamiento ocurre  $r = 3$  veces y, aunque los pares de tratamientos (1, 5), (2, 6), (3, 7) y (4, 8) nunca ocurren juntos en un bloque fila, el resto de pares ocurren exactamente una vez en cada bloque fila, convirtiendo al diseño en razonablemente bueno.

| Plan 1 |    |    |    | Plan 2 |    |    |    |
|--------|----|----|----|--------|----|----|----|
|        | C1 | C2 | C3 |        | C1 | C2 | C3 |
| F1     | 1  | 2  | 4  | F1     | 1  | 2  | 7  |
| F2     | 2  | 3  | 5  | F2     | 2  | 4  | 1  |
| F3     | 3  | 4  | 6  | F3     | 3  | 7  | 6  |
| F4     | 4  | 5  | 7  | F4     | 4  | 8  | 2  |
| F5     | 5  | 6  | 1  | F5     | 5  | 6  | 8  |
| F6     | 6  | 7  | 2  | F6     | 6  | 3  | 5  |
| F7     | 7  | 1  | 3  | F7     | 7  | 1  | 3  |
| F8     | 1  | 2  | 4  | F8     | 8  | 5  | 4  |

Tabla 5.18: Dos planes experimentales de diseños fila-columna con bloques columna completos para  $I = b = 8$  y  $r = c = 3$ .

También se podría usar el plan 2 de la Tabla 5.18, que no es ni un diseño Youden ni un diseño cíclico. Sin embargo ya se adivina que este plan no será tan apropiado para las comparaciones apareadas de tratamientos. El tratamiento 1, por ejemplo, aparece con los tratamientos 3 y 4 en un bloque, con los tratamientos 2 y 7 en dos bloques y en ningún bloque aparece con los tratamientos 5, 6 y 8. Es decir, los tratamientos no están uniformemente distribuidos en los bloques fila.

### 5.4.5 Modelo Matemático para un Diseño Fila-Columna

Para un diseño fila-columna con  $b$  filas y  $c$  columnas, el modelo tratamiento-fila-columna para una observación del tratamiento  $i$  en el bloque fila  $h$  y en el bloque columna  $q$  es

$$\underbrace{Y_{hqi}}_{\text{aleatorio}} = \overbrace{\mu + \theta_h + \phi_q + \tau_i}^{\text{determinista}} + \underbrace{\varepsilon_{hqi}}_{\text{aleatorio}} \quad \text{con } \varepsilon_{hqi} \text{ i.i.d. según una } N(0, \sigma) \tag{5.13}$$

para  $h = 1, \dots, b, q = 1, \dots, c, i = 1, \dots, I$ , y **con (h,q,i) en el diseño**.

En principio se supone que no existe interacción entre bloques y tampoco entre bloques y tratamientos. De darse  $b > I$  o  $c > I$  o ambos existirían suficientes grados de libertad para contrastar posibles interacciones.

#### 5.4.6 Estimadores Mínimo-Cuadráticos en un Diseño Fila-Columna

##### Estimadores en un Diseño Fila-Columna General

En principio se obtienen los estimadores mínimo-cuadráticos para un diseño en fila-columna general, particularizándolos después para diseños en cuadrado latino y diseños Youden.

En adelante  $n_{hqi}$  denota al número de observaciones del tratamiento  $i$  en la celda  $(h, q)$ . Por tanto se tiene que necesariamente  $n_{hqi} = 1$  o  $0$ . Además:

$$\sum_{h=1}^b \sum_{q=1}^c n_{hqi} = r \quad \text{y} \quad \sum_{q=1}^c n_{hqi} = 1$$

$$\sum_{h=1}^b \sum_{i=1}^I n_{hqi} = b, \quad \sum_{q=1}^c \sum_{i=1}^I n_{hqi} = c, \quad \text{y} \quad \sum_{h=1}^b \sum_{q=1}^c \sum_{i=1}^I n_{hqi} = bc = rI$$

También se usará la notación:  $\sum_{q=1}^c n_{hqi} = n_{h \cdot i}$  y  $\sum_{h=1}^b n_{hqi} = n_{\cdot qi}$

Finalmente,  $T_i$ ,  $B_h$ ,  $C_q$  y  $G$  denotarán respectivamente a la suma de las observaciones en el tratamiento  $i$ , en el bloque fila  $h$ , en el bloque columna  $q$  y la suma global de todas las observaciones. Es decir:

$$T_i = \sum_{h=1}^b \sum_{q=1}^c n_{hqi} y_{hqi} \quad C_q = \sum_{h=1}^b \sum_{i=1}^I n_{hqi} y_{hqi}$$

$$B_h = \sum_{q=1}^c \sum_{i=1}^I n_{hqi} y_{hqi} \quad G = \sum_{h=1}^b \sum_{q=1}^c \sum_{i=1}^I n_{hqi} y_{hqi}$$

Para obtener soluciones únicas en el proceso de estimación mínimo-cuadrática es necesario añadir restricciones sobre los parámetros. Se impondrá que  $\sum_{i=1}^I \hat{\tau}_i = 0$ ,  $\sum_{h=1}^b \hat{\theta}_h = 0$  y  $\sum_{q=1}^c \hat{\phi}_q = 0$ . Con estas restricciones, al derivar e igualar a cero la correspondiente suma de cuadrados del error, se obtiene, después de algún álgebra, el siguiente sistema de  $I$  ecuaciones normales (con una única solución):

$$r\hat{\tau}_i - \frac{1}{c} \sum_{p=1}^I \lambda_{pi} \hat{\tau}_p - \frac{1}{b} \sum_{p=1}^I \delta_{pi} \hat{\tau}_p = Q_i, \quad i = 1, \dots, I, \quad (5.14)$$

donde:

- $\lambda_{pi} = \sum_{h=1}^b n_{h,i} n_{h,p}$  denota el número de bloques fila en los cuales los tratamientos  $p$  e  $i$  aparecen juntos.
- $\delta_{pi} = \sum_{q=1}^c n_{.qi} n_{.qp}$  denota el número de bloques columna en los cuales los tratamientos  $p$  e  $i$  aparecen juntos.
- $Q_i = T_i - \frac{1}{c} \sum_{h=1}^b n_{h,i} B_h - \frac{1}{b} \sum_{q=1}^c n_{.qi} C_q + \frac{r}{bc} G$ .

### Solución para Diseños con Bloques Columna Completos

Los tres tipos de diseños fila-columna analizados en la presente sección (cuadrado latino, Youden y cíclicos con bloques columna completos) tienen bloques fila binarios ( $c < I$ ) y bloques columna completos ( $b = Is$ ). Para estos diseños, cada etiqueta de tratamiento aparece  $s$  veces en cada bloque columna de modo que  $n_{.qi} = s$  para todo  $q$  e  $i$ , y  $\delta_{pi} = cs^2$  para todo  $p$  e  $i$ . Por ello y dado que  $\sum_{q=1}^c C_q = G$  y  $r = cs$  se concluye que

$$\frac{1}{b} \sum_{q=1}^c n_{.qi} C_q = \frac{s}{b} \sum_{q=1}^c C_q = \frac{r}{bc} G. \quad (5.15)$$

Además,  $\sum_{p=1}^I \delta_{pi} \hat{\tau}_p = cs^2 \sum_{p=1}^I \hat{\tau}_p = 0$ , debido a la restricción impuesta:  $\sum_{p=1}^I \hat{\tau}_p = 0$ .

Este hecho junto con (5.15) permite simplificar el sistema de ecuaciones normales (5.14):

$$r\hat{\tau}_i - \frac{1}{c} \sum_{p=1}^I \lambda_{pi} \hat{\tau}_p = Q_i, \quad i = 1, \dots, I \quad (5.16)$$

donde ahora  $Q_i$  se reduce a:

$$Q_i = T_i - \frac{1}{c} \sum_{h=1}^b n_{h,i} B_h. \quad (5.17)$$

Finalmente, como  $\lambda_{pp} = r$  para todo  $p$ , el sistema queda:

$$r(c-1)\hat{\tau}_i - \sum_{p \neq i}^I \lambda_{pi} \hat{\tau}_p = cQ_i, \quad i = 1, \dots, I, \quad (5.18)$$

que es exactamente idéntico al deducido para un diseño en bloques incompletos con tamaño de bloque  $k = c$  (ver (5.5)).

— Solución para diseños en cuadrado latino —

Para un diseño en cuadrado latino  $s$ -replicado (apilando  $s$  cuadrados uno sobre otro) los bloques fila también son bloques completos. De hecho  $b = r = Is$ ,  $c = I$ ,  $\lambda_{pi} = Is$ , para todo  $i$  y  $p$ , y  $n_{h,i} = 1$ , para todo  $i$  y  $h$ .

Introduciendo estos valores en la solución general (5.14) y teniendo en cuenta que  $\sum_{p=1}^I \hat{\tau}_p = 0$  se obtiene:

$$\hat{\tau}_i = \frac{1}{Is}Q_i = \frac{1}{Is}T_i - \frac{1}{I^2s}G, \quad \text{para todo } i = 1, \dots, I$$

El estimador mínimo cuadrático del contraste  $\sum_{i=1}^I c_i \tau_i$  y su varianza resultan:

$$\sum_{i=1}^I c_i \hat{\tau}_i = \frac{1}{Is} \sum_{i=1}^I c_i T_i, \quad \text{y} \quad \text{Var} \left( \sum_{i=1}^I c_i \hat{\tau}_i \right) = \frac{\sigma^2}{Is} \sum_{i=1}^I c_i^2$$

— Solución para diseños Youden —

Para un diseño Youden  $s$ -replicado con  $b = Is$  filas y  $c < I$  columnas, los bloques fila forman un diseño en bloques incompletos balanceados con tamaño de bloque fila igual a  $c$ . Entonces  $\lambda_{pi} = \lambda$ , para todo  $p \neq i$ , y, como para cualquier diseño en bloques incompletos balanceados de tamaño  $k = c$ ,  $r(c-1) = \lambda(I-1)$ . Introduciendo estos valores en la solución general y teniendo en cuenta que  $\sum_{p=1}^I \hat{\tau}_p = 0$  se obtiene:

$$\hat{\tau}_i = \frac{c}{\lambda I} Q_i, \quad \text{para todo } i = 1, \dots, I$$

El estimador mínimo cuadrático del contraste  $\sum_{i=1}^I c_i \tau_i$  y su varianza son:

$$\sum_{i=1}^I c_i \hat{\tau}_i = \frac{c}{\lambda I} \sum_{i=1}^I c_i Q_i \quad \text{y} \quad \text{Var} \left( \sum_{i=1}^I c_i \hat{\tau}_i \right) = \frac{c\sigma^2}{\lambda I} \sum_{i=1}^I c_i^2$$

#### 5.4.7 Análisis de la Varianza en un Diseño Fila-Columna con Bloques Columna Completos

La Tabla 5.19 muestra el cuadro de análisis de la varianza apropiado para contrastar la hipótesis nula  $H_0^T: \{\tau_i = \tau, \forall i\}$  en el modelo fila-columna-tratamiento dado en (5.13). Conviene en cualquier caso advertir que si los bloques columna no son completos entonces los valores de  $Q_i$  deben modificarse apropiadamente.

| <b>CUADRO DEL ANÁLISIS DE LA VARIANZA</b><br><b>MODELO FILA-COLUMNA-TRATAMIENTO DE UN</b><br><b>DISEÑO FILA-COLUMNA CON BLOQUES COLUMNA</b><br><b>COMPLETOS Y NO INTERACCIONES</b> |   |                      |  |                              |
|--|---|----------------------|--|------------------------------|
| <i>Fuente de Variación</i>   | <i>Suma de Cuadrados</i>  | <i>g.l.</i>          | <i>scm</i>   | $\hat{F}$                    |
| Fil  | $scFil = \frac{1}{b} \sum_{h=1}^b B_h^2 - \frac{G^2}{bc}$                         | $b - 1$              | —  | —                            |
| Columnas   | $scCol = \frac{1}{c} \sum_{q=1}^c C_q^2 - \frac{G^2}{bc}$                         | $c - 1$              | —  | —                            |
| Trat. Ajus.  | $scT_{Ajust.} = \sum_{i=1}^I Q_i \hat{\tau}_i$                                    | $I - 1$              | $\frac{scmT_{Ajust.}}{I - 1} = \frac{scT_{Ajust.}}{I - 1}$         | $\frac{scmT_{Ajust.}}{scmR}$ |
| Residual   | $scR = scG - scFil - scCol - scT_{Ajust.}$  | $bc - b - c - I + 2$ | $\frac{scmR}{bc - b - c - I + 2} = \frac{scR}{bc - b - c - I + 2}$ | —                            |
| Global   | $scG = \sum_{h=1}^b \sum_{q=1}^c \sum_{i=1}^I n_{hqi} y_{hqi}^2 - \frac{G^2}{bc}$ | $bc - 1$             | $\frac{scmG}{bc - 1} = \frac{scG}{bc - 1}$                         | —                            |

Rechazar  $H_0^T : \tau_1 = \tau_2 = \dots = \tau_I$ , según  $p = P \left( \frac{scmT_{Ajust.}}{scmR} > F_{I-1, bc-b-c-I+2, \alpha} \right)$

Tabla 5.19: Cuadro del análisis de la varianza para un diseño conectado fila-columna con bloques columna completos y sin interacciones.

Naturalmente los diseños en cuadrado latino, Youden y cíclicos con bloques columna completos son casos particulares de un diseño fila-columna con bloques columna completos. Por tanto los correspondientes cuadros de análisis de la varianza son como el dado en la Tabla 5.19 con las modificaciones propias de sus respectivas particularidades.

— Análisis de la Varianza para un Diseño en Cuadrado Latino —

Como ya se ha probado, en este caso:

$$\hat{\tau}_i = \frac{1}{I_s} Q_i \quad \text{con} \quad Q_i = T_i - \frac{1}{c} \sum_{h=1}^b n_{h,i} B_h = T_i - \frac{1}{I} G$$

O sea, para un diseño en cuadrado latino, la suma de cuadrados explicada por tratamientos se reescribe en la forma:

$$scT_{Ajust.} = \sum_{i=1}^I Q_i \hat{\tau}_i = \frac{1}{I_s} \sum_{i=1}^I \left( T_i - \frac{1}{I} G \right)^2$$

lo que indica que la suma de cuadrados explicada por tratamientos en un diseño en cuadrado latino ni se ajusta por los efectos de los bloque fila ni por los efectos de los bloque columna. Ello es la lógica consecuencia de que, en realidad, cada tratamiento se observa el mismo número de veces en todos los bloques fila y en todos los bloques columna. Desarrollando los paréntesis, la expresión anterior queda:

$$scT = \frac{1}{I_s} \sum_{i=1}^I T_i^2 - \frac{1}{I^2_s} G^2$$

En un diseño en cuadrado latino  $s$ -replicado:  $b = Is$  y  $c = I$ , de modo que los grados de libertad del error pueden reescribirse en la forma:

$$g.l. = bc - b - c - I + 2 = I^2s - Is - 2I + 2 = (Is - 2)(I - 1)$$

— *Análisis de la Varianza para un Diseño Youden* —

El diseño Youden  $s$ -replicado con  $I$  tratamientos tiene bloques columna completos de tamaño  $b = Is$  y bloques fila formando un diseño en bloques incompletos balanceados con tamaño de bloque  $c$ . Cada tratamiento se observa  $r = cs$  veces. También para este caso se ha calculado la solución a las ecuaciones normales, resultando:

$$\hat{\tau}_i = \frac{c}{\lambda I} Q_i \quad \text{con} \quad Q_i = T_i - \frac{1}{c} \sum_{h=1}^b n_{h,i} B_h \quad \text{y} \quad \lambda = \frac{r(c-1)}{(I-1)}$$

Como en el caso de un diseño en bloques incompletos balanceados, en un diseño Youden los estimadores de los efectos tratamiento no son independientes de los efectos de los bloques fila, siendo necesario ajustar la suma de cuadrados explicada por diferencias entre tratamientos según los bloques fila donde se encuentran ubicados. No es necesario realizar ajuste alguno por los bloques columnas ya que estos forman un diseño completo. El ajuste toma la forma:

$$scT_{Ajust.} = \sum_{i=1}^I Q_i \hat{\tau}_i = \frac{c}{\lambda I} \sum_{i=1}^I Q_i^2 \quad \text{con} \quad Q_i = T_i - \frac{1}{c} \sum_{h=1}^b n_{h,i} B_h$$

En un diseño Youden el número de grados de libertad del error es:

$$g.l. = bc - b - c - I + 2 = Is - Is - c - I + 2 = (Is - 1)(c - 1) - (I - 1)$$

### 5.4.8 Comparaciones Múltiples y Tamaños Muestrales

Comenzar esta sección precisando que en los diseños en cuadrado latino y en los diseños Youden todos los contrastes son estimables, lo cual no es cierto para cualquier diseño fila-columna.

En relación con los métodos de comparación múltiple presentados en esta memoria señalar que los de Bonferroni y Scheffé son válidos para cualquier diseño fila-columna. Así, la expresión para los intervalos de confianza simultáneos (a un  $100(1 - \alpha)\%$  de confianza) para un contraste de los efectos tratamiento,  $\sum_{i=1}^I c_i \tau_i$ , es:

$$\left( \sum_{i=1}^I c_i \hat{\tau}_i \mp \omega \sqrt{\widehat{Var} \left( \sum_{i=1}^I c_i \hat{\tau}_i \right)} \right),$$

donde el coeficiente crítico  $\omega$  es  $\omega_B = t_{bc-b-c-I+2, \alpha/(2m)}$  o  $\omega_S = \sqrt{(I-1) F_{I-1, bc-b-c-I+2, \alpha}}$ , según se trate de Bonferroni o Scheffé respectivamente.

— Diseños en Cuadrado Latino —

A partir de las expresiones deducidas en la sección (5.4.6) para el estimador y la varianza de un contraste  $\sum_{i=1}^I c_i \tau_i$  en un diseño en cuadrado latino, se concluye que la forma de un intervalo de confianza múltiple que lo contenga es:

$$\left( \frac{1}{I_s} \sum_{i=1}^I c_i T_i \mp \omega \sqrt{scmR \frac{1}{I_s} \sum_{i=1}^I c_i^2} \right),$$

donde  $\omega$  es el coeficiente crítico apropiado según el método de comparación múltiple utilizado y el nivel de confianza seleccionado.

Cualquiera de los métodos de comparación múltiple descritos (Bonferroni, Scheffé, Tukey o Dunnett) es válido para este tipo de diseño. Los grados de libertad a emplear son, en este caso,  $(I_s - 2)(I - 1)$ .

La fórmula para calcular el tamaño muestral necesario para obtener una potencia preespecificada  $\pi(\Delta)$  en la prueba Anova, dados  $I$ ,  $\alpha$  y  $\sigma^2$ , es la misma que la deducida para un diseño en bloques completos aleatorizados cambiando  $b$  por  $I$ . Es decir,  $s$  debe satisfacer:

$$s \geq \frac{2\sigma^2 \phi^2}{\Delta^2}$$

Un intervalo de confianza múltiple para un contraste viene dado en este caso por:

$$\left( \frac{I-1}{Is(c-1)} \sum_{i=1}^I c_i Q_i \mp \omega \sqrt{scmR \frac{I-1}{Is(c-1)} \sum_{i=1}^I c_i^2} \right),$$

donde  $\omega$  es el coeficiente crítico apropiado según el método de comparación múltiple utilizado y el nivel de confianza seleccionado. Cualquiera de los métodos de comparación múltiple descritos (Bonferroni, Scheffé, Tukey o Dunnett) es válido para este tipo de diseño. Los grados de libertad a emplear son  $(Is-1)(c-1) - (I-1)$ .

La metodología para calcular los tamaños muestrales apropiados con un diseño Youden  $s$ -replicado son análogos a los empleados con diseños en bloques incompletos balanceados. En particular  $s$  debe verificar:

$$cs \geq \frac{2I\sigma^2\phi^2}{\Delta^2} \left[ \frac{c(I-1)}{I(c-1)} \right]$$

lo que se simplifica a:

$$s \geq \frac{2\sigma^2\phi^2(I-1)}{\Delta^2(c-1)}$$

## 5.5 Alternativas No Paramétricas

### 5.5.1 Diseño en Bloques Completamente Aleatorizados: Prueba de Friedman

La prueba de Friedman, debida al economista Milton Friedman, es una alternativa a la prueba  $F$  del análisis de la varianza para un diseño en bloques completamente aleatorizados en ausencia de la hipótesis de normalidad. En realidad la prueba de Friedman está basada en rangos y constituye una extensión de la prueba de los signos al caso de más de dos tratamientos (de hecho ambas pruebas son equivalentes cuando el número de tratamientos es  $I = 2$ ).

La prueba de Friedman contrasta la hipótesis nula de que la totalidad de  $I$  tratamientos tienen efectos idénticos sobre la respuesta. Si esto fuese así cualquier ordenación de las observaciones (y consecuentemente de los tratamientos) en cada bloque es igualmente probable. En otro caso al menos uno de los  $I$  tratamientos tendería a producir respuestas mayores que al menos uno de los restantes.

La idea de la prueba es por tanto contrastar la credibilidad de la equiprobabilidad de las ordenaciones de los tratamientos dentro de los bloques.

Para la obtención del estadístico de contraste se siguen los siguientes pasos.

Se asignan rangos a las observaciones dentro de cada bloque. Así, para cada  $h = 1, \dots, b$ , y  $i = 1, \dots, I$ ,  $R_{hi}$  denotará el rango de la observación del tratamiento  $i$ -ésimo dentro del  $h$ -ésimo bloque.

Realizado este cálculo para todos los bloques, se suman los rangos de observaciones correspondientes a cada tratamiento. Sea  $R_{.i} = \sum_{h=1}^b R_{hi}$ , para  $i = 1, \dots, I$ . Siempre se tendrá que verificar que la suma global  $R = \sum_{i=1}^I R_{.i} = \frac{bI(I+1)}{2}$ .

La estadística de prueba es:

$$W = \frac{1}{C} \left\{ \frac{12}{bI(I+1)} \sum_{i=1}^I R_{.i}^2 - 3b(I+1) \right\} \tag{5.19}$$

donde  $C$  vale 1 si no hay empates (observaciones repetidas dentro de un bloque) o, en caso de empates:

$$C = 1 - \frac{\sum_{h=1}^b \sum_{j=1}^s T_{hj}}{(I-1)I(I+1)b}$$

siendo  $s$  el número de grupos de empates en el bloque  $h$  y  $T_{hj} = (t_{hj} - 1)t_{hj}(t_{hj} + 1)$ , con  $t_{hj}$  el número de empates acontecidos en el  $j$ -ésimo grupo de empates del  $h$ -ésimo bloque.

Pues bien, bajo  $H_0$ , el estadístico  $W$  dado por (5.19) tiene un comportamiento en distribución aproximadamente  $\chi^2$  con  $I-1$  grados de libertad. Tal aproximación es tanto más correcta cuanto mayor es el tamaño muestral. Suele considerarse adecuada la aproximación si  $bI > 25$  e  $I \geq 5$ .

Resumiendo:

Con el contraste de Friedman, rechazar la igualdad de los efectos de los tratamientos a un nivel de significación  $\alpha$  siempre que  $W > \chi_{I-1, \alpha}^2$ .

Obsérvese que, como en la prueba de Kruskal-Wallis, por definición  $C$  es siempre menor o igual que 1, de ahí que sea recomendable evaluar primero  $W$  con  $C = 1$  y, si el contraste ya resulta significativo, no será preciso calcular el valor de  $C$  puesto que seguiría resultando significativo.

**Ejemplo 5.9** Se piensa solicitar a un grupo de doce médicos su impresión particular, en una escala de 0 a 5, acerca del grado de gravedad de cuatro enfermedades (codificadas de 1 a 4). El interés del experimento radica en contrastar si existen diferencias significativas en los niveles de gravedad de las cuatro enfermedades.

Si el experimento se aleatoriza adecuadamente (selección al azar de los doce médicos y aleatorización del orden de las preguntas para cada uno de ellos) el experimento puede catalogarse como un diseño en bloques completamente aleatorizados, donde los bloques son los médicos (y así  $b = 12$ ), los tratamientos son cada una de las enfermedades ( $I = 4$ ) y las respuestas toman valores enteros entre 0 y 5, de modo que cabe no asumir normalidad.

Es entonces aplicable el contraste de Friedman según la tabla que sigue.

| Bloques   | Datos        |   |   |   | Rangos       |     |     |     | Empates                     |                       |
|---|--------------|---|---|---|--------------|-----|-----|-----|-----------------------------|-----------------------|
|   | Enfermedades |   |   |   | Enfermedades |     |     |     |                             |                       |
|   | 1            | 2 | 3 | 4 | 1            | 2   | 3   | 4   | $t_{hj}$                    | $\sum_{j=1}^s T_{hj}$ |
| Méd. 1  | 4            | 2 | 2 | 3 | 4            | 1'5 | 1'5 | 3   | 2                           | 6                     |
| Méd. 2  | 3            | 2 | 4 | 5 | 2            | 1   | 3   | 4   | 0                           | 0                     |
| Méd. 3  | 5            | 1 | 3 | 4 | 4            | 1   | 2   | 3   | 0                           | 0                     |
| Méd. 4  | 3            | 2 | 3 | 3 | 3            | 1   | 3   | 3   | 3                           | 24                    |
| Méd. 5  | 5            | 3 | 4 | 4 | 4            | 1   | 2'5 | 2'5 | 2                           | 6                     |
| Méd. 6  | 3            | 1 | 3 | 4 | 2'5          | 1   | 2'5 | 4   | 2                           | 6                     |
| Méd. 7  | 2            | 2 | 3 | 3 | 1'5          | 1'5 | 3'5 | 3'5 | 2 y 2                       | 12                    |
| Méd. 8  | 5            | 3 | 4 | 3 | 4            | 1'5 | 3   | 1'5 | 2                           | 6                     |
| Méd. 9  | 4            | 2 | 4 | 3 | 3'5          | 1   | 3'5 | 2   | 2                           | 6                     |
| Méd. 10   | 4            | 1 | 1 | 4 | 3'5          | 1'5 | 1'5 | 3'5 | 2 y 2                       | 12                    |
| Méd. 11   | 4            | 1 | 2 | 4 | 3'5          | 1   | 2   | 3'5 | 2                           | 6                     |
| Méd. 12   | 3            | 3 | 1 | 4 | 2'5          | 2'5 | 1   | 4   | 2                           | 6                     |
| $R_{.1} = 38 \quad R_{.2} = 15'5 \quad R_{.3} = 29 \quad R_{.4} = 37'5$ |              |   |   |   |              |     |     |     | $\sum_h \sum_j T_{hj} = 90$ |                       |

Se obtiene entonces  $C = 1 - \frac{90}{12 \times 3 \times 4 \times 5} = 0'875$  y:

$$W = \frac{1}{0'875} \left\{ \frac{12}{4 \times 12 \times 5} (38^2 + 15'5^2 + 29^2 + 37'5^2) - 3 \times 12 \times 5 \right\} = 18'95$$

Como  $P(\chi_3^2 > 18'95) < 0'001$  se concluye que el contraste es significativo.

La expresión de  $W$  en (5.19) es la más usual y frecuente. Un estadístico alternativo es:

$$W' = \frac{(b-1)[B - bI(I+1)^2/4]}{A - B} \tag{5.20}$$

con  $A = \sum_{h=1}^b \sum_{i=1}^I R_{hi}^2$  y  $B = \frac{1}{b} \sum_{i=1}^I R_{.i}^2$ .

Bajo  $H_0$ , el estadístico  $W'$  tiene una distribución aproximadamente  $F$  de Fisher-Snedecor con  $I - 1$  y  $(b - 1)(I - 1)$  grados de libertad. La aproximación es muy buena, especialmente para un número grande de bloques  $b$ . De hecho, estudios muy recientes debidos a Iman y Davenport (1980) muestran que la aproximación  $F$  es claramente superior a la aproximación  $\chi^2$ , de modo que es recomendable resolver el contraste de Friedman en base al estadístico  $W'$ .

Si el contraste de Friedman resulta significativo a un nivel de significación  $\alpha$  es factible aplicar el siguiente criterio para las comparaciones múltiples. Dos tratamientos  $i$  y  $p$  son considerados distintos al nivel de significación global  $\alpha$  siempre que:

$$|R_{.i} - R_{.p}| > t_{\alpha/2} \left[ \frac{2b(A - B)}{(b - 1)(I - 1)} \right]^{1/2}$$

### 5.5.2 Diseño en Bloques Incompletos Balanceados: Prueba de Durbin

Durbin propuso un estadístico para contrastar la hipótesis nula de igualdad de los efectos de tratamientos en un diseño en bloques incompletos balanceados. Se optará por la prueba de Durbin frente al Anova siempre que: (i) no se verifiquen las hipótesis de normalidad, (ii) se desee un método de análisis más sencillo y (iii) las observaciones son en realidad rangos.

Como en la sección 5.3.2 sean:  $I$  el número de tratamientos,  $k$  el número de tratamientos por bloque ( $k < I$ ),  $b$  el número de bloques,  $r$  el número de veces que aparece cada tratamiento en el diseño y  $\lambda$  el número de bloques en que cualquier par de tratamientos distintos aparecen juntos.

Para la obtención del estadístico de contraste se siguen los siguientes pasos.

Como en Friedman se asignan rangos a las observaciones dentro de cada bloque. Sea  $R_{hi}$  el rango de la observación del tratamiento  $i$ -ésimo dentro del  $h$ -ésimo bloque *siempre que el tratamiento  $i$  exista en el bloque  $h$* , o sea, siempre que  $n_{hi} = 1$ .

Después, se suman los rangos de las  $r$  observaciones correspondientes a cada tratamiento. Sea  $R_{.i} = \sum_{h=1}^b R_{hi}$ , para  $i = 1, \dots, I$ . Obsérvese que aunque el sumatorio se extiende hasta  $b$  solo existen  $r$  sumandos.

La estadística de prueba es:

$$D = \frac{12(I - 1)}{rI(k - 1)(k + 1)} \sum_{i=1}^I \left( R_{.i} - \frac{r(k + 1)}{2} \right)^2 \tag{5.21}$$

que admite la siguiente expresión equivalente:

$$D = \frac{12(I - 1)}{rI(k - 1)(k + 1)} \sum_{i=1}^I R_{.i}^2 - 3 \frac{r(I - 1)(k + 1)}{k - 1} \tag{5.22}$$

Bajo  $H_0$ , el estadístico  $D$  tiene un comportamiento en distribución aproximadamente  $\chi^2$  con  $I - 1$  grados de libertad. La bondad de la aproximación se asienta en el hecho de que el número de réplicas  $r$  de cada tratamiento en el diseño sea moderadamente grande.

Entonces:

Con el contraste de Durbin, rechazar la igualdad de los efectos de los tratamientos a un nivel de significación  $\alpha$  siempre que  $D > \chi_{I-1, \alpha}^2$ .

**Ejemplo 5.10** Un fabricante de helados tiene a la venta siete variedades de helado y desea saber si existen diferencias significativas entre las preferencias de la gente. Selecciona al azar a siete personas y cada una de ellas prueba tres variedades distintas de helado. Al acabar deben establecer un ranking de preferencias: de 1 a 3 (el rango 1 para el que más les gustó). El plan experimental y los resultados obtenidos se reflejan en el siguiente cuadro:

| Persona    | Variedad |   |   |   |   |   |   |
|------------|----------|---|---|---|---|---|---|
|            | 1        | 2 | 3 | 4 | 5 | 6 | 7 |
| 1          | 2        | 3 |   | 1 |   |   |   |
| 2          |          | 3 | 1 |   | 2 |   |   |
| 3          |          |   | 2 | 1 |   | 3 |   |
| 4          |          |   |   | 1 | 2 |   | 3 |
| 5          | 3        |   |   |   | 1 | 2 |   |
| 6          |          | 3 |   |   |   | 1 | 2 |
| 7          | 3        |   | 1 |   |   |   | 2 |
| $R_{.i} =$ | 8        | 9 | 4 | 3 | 5 | 6 | 7 |

Se trata de un diseño en bloques incompletos balanceados con  $I = 7$  tratamientos (variedades),  $b = 7$  bloques (personas),  $k = 3$  tratamientos asignados por bloque,  $r = 3$  veces que se prueba cada variedad en el experimento y  $\lambda = 1$ . Además, a tenor del tipo de respuesta (rangos), la prueba de Durbin es adecuada para contrastar la hipótesis nula de que no existe alguna variedad de helado especialmente preferida sobre el resto de variedades.

Al evaluar el estadístico de Durbin se obtiene  $\hat{D} = 12 < \chi_{6, 0.05}^2 = 12.59$ . El contraste no resulta significativo al 5% y no se rechaza la nula. Sin embargo el  $p$ -valor es bastante pequeño:  $p \approx 0.065$ .

Es importante reseñar que si el número de tratamientos es igual al número de unidades experimentales por bloque (o sea el diseño pasa a ser un diseño en bloques completamente aleatorizados), entonces la prueba de Durbin se reduce a la prueba de Friedman.

Si existen empates se asignan los rangos medios aunque entonces se modifica la distribución de  $D$  bajo  $H_0$ . En todo caso, si el número de empates no es excesivo el efecto sobre el cambio en la distribución es prácticamente nulo.

En todo caso existe un ajuste apropiado cuando existe un número importante de empates.

Si el contraste de Durbin resulta significativo a un nivel de significación  $\alpha$  es factible aplicar el siguiente criterio para las comparaciones múltiples. Dos tratamientos  $i$  y  $p$  son considerados distintos al nivel de significación global  $\alpha$  siempre que:

$$|R_i - R_p| > t_{\alpha/2} \left\{ \frac{r(k+1)(k-1)[bk(I-1) - ID]}{6(I-1)(bk - I - b + 1)} \right\}^{1/2}$$

## 5.6 Problemas Propuestos

**Problema 1** (Millen, B., Shankar, R. Christoffersen, K. y Nevathia, P., 1996) Se ejecutó un experimento para explorar la capacidad que un individuo tiene para reproducir con precisión una línea recta de una longitud dada. Se trazó una línea horizontal de cinco centímetros de longitud sobre una hoja de papel blanco. A continuación se invitaba a distintos sujetos a sentarse en una mesa que estaba cincuenta centímetros de una pared blanca donde se fijaba la hoja a la altura de los ojos de cada individuo. A cada individuo se le pedía entonces que reprodujese la línea recta sobre una hoja en la que se había dibujado un borde.

Los individuos que participaron en el experimento eran todos estudiantes universitarios de ambos sexos y con edades comprendidas entre 20 y 30 años.

Había seis tipos diferentes de bordes representando las combinaciones de tres formas geométricas: cuadrado, círculo y triángulo equilátero (niveles de un factor  $C$  codificados de 1 a 3) con dos áreas: 16 y 9 pulgadas al cuadrado (niveles de un factor  $D$  codificados como 1 y 2 respectivamente). Y es que el propósito del experimento no era tanto medir la precisión con que los sujetos se acercaban a los cinco centímetros de longitud, sino ver como afectaban a las longitudes de las líneas trazadas las formas y áreas de los bordes.

Los individuos dibujaron razonablemente bien las rectas a mano y uno de los experimentadores midió, redondeando al medio milímetro más cercano, las longitudes de cada línea. El archivo RECTAS.SAV contiene las desviaciones a los cinco centímetros de catorce de los individuos que intervinieron en el experimento.

1. Ajustar un modelo tratamiento-bloque a los datos considerando las seis combinaciones de formas y áreas como tratamientos y los individuos como bloques. Chquear las suposiciones sobre el error del modelo.
2. Obtener al menos un gráfico que permita examinar los datos.

3. Plantear contrastes sobre las seis combinaciones de tratamientos representando las siguientes comparaciones:
  - (a) diferencias en los efectos de las áreas para cada forma por separado,
  - (b) diferencia promedio en los efectos de las áreas,
  - (c) diferencia promedio en los efectos de las formas.
4. Resolver el primero de los contrastes del apartado anterior y establecer conclusiones.
5. ¿Bajo qué condiciones serían de interés los otros dos contrastes del tercer apartado? ¿Se dan esas condiciones en este experimento?

**Problema 2** (Cochran y Cox, 1957) Se ejecuta un experimento para comparar los efectos que el tiempo de almacenaje en cámaras frigoríficas de la carne de vaca (filetes) tiene sobre su calidad final. Se investigaron seis períodos de tiempo (0, 1, 2, 4, 9 y 18 días) que se codificaron de 1 a 6. Se pensó que filetes de lugares similares de ambos lados del animal serían semejantes y por ello en el experimento los bloques tendrían tamaño 2. Se eligieron 15 partes distintas del animal y, por tanto, 15 bloques.

La respuesta  $y_{hi}$  del tratamiento  $i$ -ésimo en el bloque  $h$ -ésimo era una medida objetiva de la ternura de la carne y la máxima puntuación posible era 40. El diseño y las respuestas figuran en la siguiente tabla.

| Bloques | Tratamientos |    |    |    |    |    |
|---------|--------------|----|----|----|----|----|
|         | 1            | 2  | 3  | 4  | 5  | 6  |
| 1       | 7            | 17 |    |    |    |    |
| 2       |              |    | 26 | 25 |    |    |
| 3       |              |    |    |    | 33 | 29 |
| 4       | 17           |    | 27 |    |    |    |
| 5       |              | 23 |    |    | 27 |    |
| 6       |              |    |    | 29 |    | 30 |
| 7       | 10           |    |    | 25 |    |    |
| 8       |              | 26 |    |    |    | 37 |
| 9       |              |    | 24 |    | 26 |    |
| 10      | 25           |    |    |    | 40 |    |
| 11      |              | 25 |    | 34 |    |    |
| 12      |              |    | 34 |    |    | 32 |
| 13      | 11           |    |    |    |    | 27 |
| 14      |              | 24 | 21 |    |    |    |
| 15      |              |    |    | 26 | 32 |    |

1. ¿Cuál es el valor de  $\lambda$  para este diseño en bloques incompletos balanceados?
2. ¿Qué beneficio esperaban obtener los experimentadores al emplear un diseño en bloques en lugar de un diseño completamente aleatorizado?

3. Calcular el estimador mínimo-cuadrático de  $\tau_6 - \tau_1$  y su correspondiente varianza.
4. Obtener conclusiones del experimento.

**Problema 3** (Das, M.N. y Kulkarni, G.A. *Biometrics*, 1996) Se realiza un experimento para evaluar la potencia de diferentes dosis y preparados de vitamina D. Las unidades experimentales fueron ratas agrupadas en bloques caracterizados por pertenecer a la misma camada. Había seis tratamientos. Los tratamientos codificados con 1, 2 y 3 consistían en preparados estándar de tres dosis de vitamina D, igualmente espaciadas en orden ascendente en la escala logarítmica. Los tratamientos 4, 5 y 6 son tres nuevos preparados a prueba de tres dosis de de vitamina D, igualmente espaciadas en orden descendente en la escala logarítmica. Los datos se presentan en la siguiente tabla.

| Bloques | Tratamientos |    |    |    |    |    |
|---------|--------------|----|----|----|----|----|
|         | 1            | 2  | 3  | 4  | 5  | 6  |
| 1       | 2            | 8  |    |    | 9  | 7  |
| 2       | 6            |    | 9  | 3  |    | 8  |
| 3       |              | 6  | 12 | 4  | 6  |    |
| 4       | 9            | 11 |    |    | 14 | 13 |
| 5       | 10           |    | 17 | 8  |    | 10 |
| 6       |              | 7  | 5  | 6  | 9  |    |
| 7       | 4            | 10 |    |    | 11 | 13 |
| 8       | 11           |    | 9  | 3  |    | 15 |
| 9       |              | 9  | 14 | 5  | 8  |    |
| 10      | 4            | 7  |    |    | 10 | 10 |
| 11      | 12           |    | 9  | 15 |    | 15 |
| 12      |              | 8  | 11 | 7  | 8  |    |
| 13      | 4            | 4  |    |    | 5  | 9  |
| 14      | 7            |    | 8  | 3  |    | 9  |
| 15      |              | 15 | 10 | 6  | 8  |    |
| 16      | 2            | 4  |    |    | 6  | 6  |
| 17      | 4            |    | 13 | 5  |    | 12 |
| 18      |              | 10 | 13 | 4  | 18 |    |

1. Mostrar que se trata de un diseño divisible en grupos con tres bloques repetidos seis veces. Identificar los grupos y los valores de  $\lambda_1$  y de  $\lambda_2$ .
2. Calcular la varianza para la comparación de un par de tratamientos que son primeros socios. También para dos que son segundos socios.
3. En base al anterior apartado, ¿por qué los autores habrán elegido este particular diseño?
4. Realizar el análisis estadístico y establecer conclusiones.

**Problema 4** (Wardrop, R., 1991) Un video-juego permite elegir entre cinco modos distintos de sonido y el Profesor Robert Wardrop, de la Universidad de Wisconsin, quería evaluar con cuál de los cinco jugaba mejor.

Los tres primeros modos son diferentes tipos de música de fondo junto a ciertos sonidos llamativos asociados al juego y que tratan de realzarlo. El cuarto modo tiene estos sonidos propios del juego pero carece de música de fondo. El quinto modo consiste en ejecutar el juego sin sonido alguno.

El profesor planificó un diseño en cuadrado latino. Los cinco modos de sonido constituyeron los tratamientos que se codificaron de 1 a 5. El profesor jugaría en cinco días distintos, cada día cinco veces y cambiando el orden de selección del modo de sonido de un día para otro. Es decir, los dos factores de bloqueo considerados fueron el día y el orden de los modos de sonido.

La respuesta fue la puntuación alcanzada en cada partida. El diseño y los resultados figuran en la siguiente tabla:

|       | Día   |     |       |     |       |    |       |     |       |     |
|-------|-------|-----|-------|-----|-------|----|-------|-----|-------|-----|
|       | 1     |     | 2     |     | 3     |    | 4     |     | 5     |     |
| Orden | Trat. | Y   | Trat. | Y   | Trat. | Y  | Trat. | Y   | Trat. | Y   |
| 1     | 1     | 94  | 3     | 100 | 4     | 98 | 2     | 101 | 5     | 112 |
| 2     | 3     | 103 | 2     | 111 | 1     | 51 | 5     | 110 | 4     | 90  |
| 3     | 4     | 114 | 1     | 75  | 5     | 94 | 3     | 85  | 2     | 107 |
| 4     | 5     | 100 | 4     | 74  | 2     | 70 | 1     | 93  | 3     | 106 |
| 5     | 2     | 106 | 5     | 95  | 3     | 81 | 4     | 90  | 1     | 73  |

1. Escribir un modelo apropiado.
2. Obtener algún gráfico ilustrativo.
3. Completar un análisis de la varianza.
4. Evaluar si el bloqueo fue efectivo.
5. Resolver, con un 5% de significación, los siguientes contrastes:
  - los contrastes pairwise o dos a dos,
  - el contraste “música” frente a “no música”, o sea:

$$1/3(\tau_1 + \tau_2 + \tau_3) - 1/2(\tau_4 + \tau_5)$$

- y el contraste “sonidos del juego” frente a “no sonidos del juego”, o sea:

$$1/4(\tau_1 + \tau_2 + \tau_3 + \tau_4) - \tau_5$$

6. ¿Cuáles son las conclusiones principales?

7. ¿Qué modo de sonido debería usar el profesor?

**Problema 5** Retomemos el problema anterior. Supóngase que el quinto día el ordenador no hubiese funcionado y que por consiguiente no se dispusiese de la información relativa al quinto día en la tabla anterior. Entonces

1. ¿Qué tipo de diseño resultaría?
2. Resolver los apartados anteriores en esta nueva situación.
3. ¿Se obtienen conclusiones semejantes?

**Problema 6** (Gbenado, M., Veress, A., Heimenz, L. Monroe, J. y Yu, S., 1996) En el siguiente experimento el objetivo era investigar el efecto del color sobre la percepción de una cantidad. Los sujetos que en él participaron se seleccionaron al azar entre estudiantes.

El experimento consistió en esparcir un número pequeño de caramelos (sin el envoltorio) de un color específico sobre una bandeja. Cada uno de los sujetos participantes en el experimento echaba entonces un vistazo a la bandeja durante tres segundos y a continuación se le pedía que adivinase el número de caramelos que había en la bandeja.

Los factores tratamiento de interés eran dos: el “número de caramelos sobre la bandeja” y el “color de los caramelos”. Los niveles seleccionados fueron 17, 29 y 41 para el factor tratamiento número de caramelos y amarillo, naranja y marrón para el factor tratamiento color. Los experimentadores decidieron que cada sujeto vería la totalidad de combinaciones de tratamientos (o sea 9) y optaron por planificar un diseño en cuadrado latino  $9 \times 9$ .

En este problema sólo se contempla una parte muy pequeña del estudio original. En particular se ha seleccionado un cuadrado latino replicado dos veces (aunque el número de réplicas esa en verdad mucho mayor). Los datos se muestran en la siguiente tabla donde los sujetos representan los bloques fila, las secuencias temporales en que se asignaron los tratamientos a los sujetos representan los bloques columna y las combinaciones de tratamientos, que se codificaron según:

|                     |                    |                   |
|---------------------|--------------------|-------------------|
| 11 = (17, amarillo) | 12 = (17, naranja) | 13 = (17, marrón) |
| 21 = (29, amarillo) | 22 = (29, naranja) | 23 = (29, marrón) |
| 31 = (41, amarillo) | 32 = (41, naranja) | 33 = (41, marrón) |

figuran entre paréntesis encima de la respuesta del individuo correspondiente.

| Sujeto | Orden Temporal |             |            |            |            |             |            |             |             |
|--------|----------------|-------------|------------|------------|------------|-------------|------------|-------------|-------------|
|        | 1              | 2           | 3          | 4          | 5          | 6           | 7          | 8           | 9           |
| 1      | (23)<br>4      | (22)<br>-3  | (11)<br>0  | (12)<br>-3 | (32)<br>-1 | (13)<br>-3  | (31)<br>-6 | (33)<br>-9  | (21)<br>-1  |
| 2      | (12)<br>2      | (31)<br>16  | (32)<br>21 | (21)<br>9  | (22)<br>4  | (33)<br>16  | (23)<br>9  | (11)<br>2   | (13)<br>2   |
| 3      | (21)<br>4      | (23)<br>-1  | (12)<br>7  | (13)<br>-5 | (33)<br>-1 | (11)<br>-13 | (32)<br>-9 | (31)<br>-19 | (22)<br>-16 |
| 4      | (32)<br>21     | (12)<br>4   | (22)<br>3  | (31)<br>11 | (13)<br>0  | (21)<br>4   | (11)<br>0  | (23)<br>4   | (33)<br>11  |
| 5      | (31)<br>7      | (11)<br>-2  | (21)<br>2  | (33)<br>3  | (12)<br>-3 | (23)<br>3   | (13)<br>-4 | (22)<br>-5  | (32)<br>-7  |
| 6      | (11)<br>3      | (33)<br>7   | (31)<br>14 | (23)<br>11 | (21)<br>12 | (32)<br>17  | (22)<br>10 | (13)<br>5   | (12)<br>0   |
| 7      | (22)<br>11     | (21)<br>14  | (13)<br>0  | (11)<br>1  | (31)<br>16 | (12)<br>1   | (33)<br>13 | (32)<br>14  | (23)<br>7   |
| 8      | (13)<br>7      | (32)<br>16  | (33)<br>16 | (22)<br>4  | (23)<br>4  | (31)<br>16  | (21)<br>14 | (12)<br>7   | (11)<br>2   |
| 9      | (33)<br>21     | (13)<br>2   | (23)<br>10 | (32)<br>24 | (11)<br>6  | (22)<br>13  | (12)<br>2  | (21)<br>8   | (31)<br>20  |
| 10     | (33)<br>16     | (31)<br>20  | (22)<br>6  | (21)<br>6  | (11)<br>7  | (23)<br>6   | (12)<br>2  | (13)<br>3   | (32)<br>14  |
| 11     | (12)<br>2      | (22)<br>4   | (32)<br>11 | (13)<br>2  | (21)<br>9  | (33)<br>1   | (23)<br>4  | (31)<br>-4  | (11)<br>7   |
| 12     | (13)<br>-4     | (23)<br>-11 | (33)<br>-4 | (11)<br>-3 | (22)<br>4  | (31)<br>11  | (21)<br>-1 | (32)<br>1   | (12)<br>-3  |
| 13     | (21)<br>4      | (12)<br>-1  | (11)<br>2  | (32)<br>11 | (31)<br>11 | (13)<br>-3  | (33)<br>1  | (22)<br>-1  | (23)<br>11  |
| 14     | (22)<br>2      | (13)<br>-7  | (12)<br>-9 | (33)<br>8  | (32)<br>-2 | (11)<br>-6  | (31)<br>-9 | (33)<br>4   | (21)<br>2   |
| 15     | (31)<br>21     | (32)<br>21  | (23)<br>14 | (22)<br>14 | (12)<br>4  | (21)<br>16  | (13)<br>0  | (11)<br>5   | (33)<br>11  |
| 16     | (11)<br>2      | (21)<br>9   | (31)<br>21 | (12)<br>6  | (23)<br>9  | (32)<br>18  | (22)<br>9  | (33)<br>16  | (13)<br>2   |
| 17     | (32)<br>6      | (33)<br>6   | (21)<br>-1 | (23)<br>-1 | (13)<br>2  | (22)<br>4   | (11)<br>-3 | (12)<br>-1  | (31)<br>6   |
| 18     | (23)<br>4      | (11)<br>2   | (13)<br>2  | (31)<br>11 | (33)<br>6  | (12)<br>2   | (32)<br>6  | (21)<br>-1  | (22)<br>4   |

1. El experimento se ejecutó en un hall muy transitado de la Ohio State University. Los participantes en el experimento fueron reclutados de entre los estudiantes que pasaban por allí siempre y cuando no fuesen daltónicos. Naturalmente no podían

ver como se ejecutaba el experimento previamente. ¿Crees que las conclusiones del estudio son válidas para la población en general?

2. Ajustar un modelo que incluya los efectos de los dos factores bloque, el efecto tratamiento y el efecto interacción entre el bloque tiempo y los tratamientos. Chequear después si los residuos son aproximadamente normales y homocedásticos en los tratamientos. ¿Es preferible utilizar las respuestas originales o su raíz cuadrada o alguna otra transformación?
3. Presentar una tabla de análisis de la varianza y contrastar aquellas hipótesis que se consideren de interés. Establecer conclusiones.
4. Reparametrizar el modelo en términos de efectos principales e interacción de los dos factores tratamiento. Rehacer entonces el cuadro de análisis de la varianza. ¿Qué se puede concluir del experimento?

**Problema 7** En siete hospitales de una determinada ciudad se han contabilizado los nacimientos acaecidos en los últimos doce meses. Los datos se clasificaron según las estaciones del año al objeto de contrastar si la tasa de nacimientos es constante a lo largo de las cuatro estaciones. Resolver el problema planteado según los resultados que se exponen en la siguiente tabla y que también se encuentran en el archivo BIRTHS.SAV. Explicar el tipo de diseño experimental y justificar el procedimiento analítico empleado.

| Hospital | Número de Nacimientos |           |        |       |
|----------|-----------------------|-----------|--------|-------|
|          | Invierno              | Primavera | Verano | Otoño |
| A        | 92                    | 112       | 94     | 77    |
| B        | 9                     | 11        | 10     | 12    |
| C        | 98                    | 109       | 92     | 81    |
| D        | 19                    | 26        | 19     | 18    |
| E        | 21                    | 22        | 23     | 24    |
| F        | 58                    | 71        | 51     | 62    |
| G        | 42                    | 49        | 44     | 41    |

**Problema 8** Con el propósito de investigar y controlar a los coyotes se diseña un experimento para determinar cuál de un total de cinco olores tiende a ser más agradable para estos depredadores. Los experimentadores han observado que la presencia simultánea de más de tres olores tiende a confundirlos y a generar resultados inconsistentes. Se opta entonces por dividir un enorme corral en parcelas muy separadas sobre cada una de las cuales se generan hasta tres fuentes de olor distintas. En cada parcela se suelta un coyote y se evalúa el tiempo en segundos que tarda en detectar cada olor. La dispersión de los olores en las parcelas se realizó de acuerdo a un diseño en bloques incompletos balanceados y los resultados se exponen en la siguiente tabla.

| Coyote | Olor |    |   |    |   |
|--------|------|----|---|----|---|
|        | 1    | 2  | 3 | 4  | 5 |
| 1      | 12   | 23 |   | 14 |   |
| 2      |      | 17 | 2 |    | 2 |
| 3      | 16   |    | 1 | 6  |   |
| 4      |      | 42 |   | 10 | 0 |
| 5      | 8    |    | 6 |    | 1 |
| 6      | 22   | 31 |   |    | 0 |
| 7      | 28   | 16 | 4 |    |   |
| 8      | 15   |    |   | 7  | 4 |
| 9      |      | 67 | 5 | 18 |   |
| 10     |      |    | 6 | 16 | 1 |

1. ¿Se trata en verdad de un diseño en bloques incompletos balanceados? Determinar sus características.
2. ¿Qué procedimiento analítico es más apropiado para resolver el problema de los experimentadores? ¿Por qué?
3. ¿Hay diferencias significativas en los tiempos medios de detección de los cinco olores? Si así fuere, ¿qué olores son los preferidos por los coyotes?

## Capítulo 6

# Efectos Aleatorios y Componentes de la Varianza

### 6.1 Introducción

En los capítulos precedentes, los niveles de los factores tratamiento eran específicamente seleccionados por el experimentador ya que el interés del experimento se centraba en conocer los efectos sobre la respuesta de esos niveles particulares. Como se indicó en el capítulo 1 se habla entonces de **efectos fijos**, indicando así que su representación en el modelo se corresponde con “constantes” desconocidas (parámetros). Los modelos conteniendo únicamente efectos fijos se denominan también *modelos de efectos fijos*.

Sin embargo, frecuentemente interesa al experimentador un factor que tiene un gran número de posibles niveles, de modo tal que, para utilizar en el experimento, opta por elegir una muestra de ellos al azar. En tal caso se habla de **efectos aleatorios** y en el correspondiente modelo matemático aparecen representados como variables aleatorias idénticamente distribuidas según la distribución de la población de niveles. Los modelos se denominan entonces *modelos de efectos aleatorios*. En este tipo de modelos el interés no se centra en los efectos particulares que sobre la respuesta tienen los efectos incluidos en el experimento. Bastante más allá, el interés radica en medir la variabilidad existente en la totalidad de los efectos de la población de niveles. Se trata pues de objetivos bien distintos y, por consiguiente, la planificación y análisis difiere con respecto al caso de efectos fijos.

A lo largo del capítulo se supondrá que la población de niveles es infinita o lo suficientemente grande como para ser considerada como tal. En otro caso es preciso aplicar correcciones para poblaciones finitas a toda la formulación matemática. Sin embargo, esta cuestión no será tratada en este manual (para este problema se sugiere consultar la obra de Searle y Fawcett (1970)).

## 6.2 Algunos Ejemplos

Se exponen a continuación dos ejemplos de situaciones experimentales susceptibles de ser modelizadas mediante modelos de efectos aleatorios.

**Ejemplo 6.1** Un productor de sopas de tomate envasadas desea reducir la variabilidad en el espesor de sus sopas. Las causas más probables de esta variabilidad serían la calidad de la fécula de maíz servida por un proveedor y empleada en su fabricación o la influencia de los distintos operarios que elaboran el producto. Consideremos dos posibles escenarios:

[1] Los operarios llevan mucho tiempo en la empresa y están altamente preparados para su labor, de modo que casi con total seguridad el problema radica en la fécula de maíz recibida del proveedor. El factor tratamiento es la fécula de maíz y sus niveles son todas las posibles partidas que ha suministrado, y podría suministrar en el futuro, el citado proveedor. Al menos teóricamente tal población de niveles es infinita y se puede suponer que las últimas partidas recibidas son una muestra aleatoria de todas las que podrían haber sido servidas. El efecto de la fécula sobre el espesor de la sopa es por tanto un efecto aleatorio y será modelado mediante una variable aleatoria.

[2] Se sabe que la calidad de la fécula es extremadamente consistente y homogénea, de modo que el origen de la variabilidad debe encontrarse en los operarios. La compañía es enorme y los operarios cambian con mucha frecuencia. Por este motivo los operarios seleccionados para formar parte del experimento constituyen una pequeña muestra de todos los operarios empleados en la actualidad, o que pudiesen pertenecer a la compañía en el futuro. Asumido que el grupo seleccionado es representativo de la población, puede considerársele una muestra aleatoria de la población total de operarios. Otra vez estamos ante un efecto aleatorio.

Constatado que no es preciso incluir factores bloque, podríamos usar un diseño de una vía completamente aleatorizado. Primero, en su planificación, habrá que seleccionar al azar una muestra de tamaño  $I$  de la población de niveles del factor tratamiento aleatorio y, a continuación, asignar al azar y como de costumbre las unidades experimentales a los  $I$  niveles seleccionados. Aquí, para conseguir una potencia preespecificada en las pruebas de hipótesis o unas longitudes máximas pre-determinadas en los intervalos de confianza, además del tamaño muestral de cada nivel, habrá que determinar también con antelación el valor apropiado de  $I$ .

**Ejemplo 6.2** (Cameron, J.M., 1952) El siguiente experimento está relacionado con la limpieza de la lana. Siguiendo la lista de pasos a seguir en la planificación de un experimento, tal y como se expuso en el capítulo 1, para este experimento se tendría:

**1.- Definir los objetivos del experimento.** La lana “en crudo” (tal y como se recoge) contiene distintos grados de grasa, suciedad y elementos extraños que deben ser eliminados antes de comenzar el proceso de manufactura.

Tanto el precio de adquisición como los impuestos de aduana de un cargamento de lana se basan en su peso después de haber limpiado concienzudamente la lana. En concreto interesa la variable *contenido limpio*, entendiéndose por tal la tasa, en tanto por cien, del peso de la lana limpia con respecto al peso original de la lana en crudo.

El experimento se ejecutó en orden a estimar la variabilidad en el contenido limpio de balas de lana de un mismo cargamento.

**2.- Identificar las posibles fuentes de variación.**

(i) *Factores tratamiento y niveles.* El factor tratamiento era la “bala de lana” y sus niveles la población completa de balas procedentes de un cargamento particular. Se seleccionaron al azar siete balas del cargamento, el cual era lo suficientemente grande como para considerar a esta selección una muestra aleatoria de una población infinita de balas. Así, cabe considerar al factor tratamiento “bala de lana” como un factor de efectos aleatorios.

(ii) *Unidades experimentales.* Los tiempos en que cada bala de lana fue observada.

(iii) *Factores bloque, factores ruido y covariables.* No se identificaron otras fuentes de variación.

**3.- Elegir una regla de asignación de las unidades experimentales a los tratamientos.** Se optó por un diseño completamente aleatorizado.

**4.- Especificar las medidas que se realizarán, el procedimiento experimental y anticiparse a las posibles dificultades.** Se usaría una máquina para atravesar cada bala y extraer varios trozos de lana de cada una de las siete balas seleccionadas, de modo que se dispondría de varias observaciones de contenido limpio de cada bala. Se pesaría cada trozo extraído y a continuación se restregaría la lana y se limpiaría concienzudamente para, acto seguido, proceder a pesarlo de nuevo y obtener el valor del contenido limpio. Una posible dificultad que se previó anticipadamente en todo el proceso anterior fue que el proceso automatizado de restregar la lana funcionaba bien con grandes cantidades de lana, pero podría funcionar mucho peor con los trozos pequeños que se iban a utilizar en el experimento. Obviamente ello conduciría a un error experimental de dimensión mayor que el observado en la producción rutinaria.

## 6.3 Un Efecto Aleatorio

### 6.3.1 El Modelo de Una Vía de Efectos Aleatorios

Para un diseño completamente aleatorizado con  $I$  niveles seleccionados al azar de un factor tratamiento  $T$ , el **modelo de una vía de efectos aleatorios** es

$$\boxed{\underbrace{Y_{it}}_{\text{aleatorio}} = \underbrace{\mu}_{\text{determinista}} + \underbrace{T_i + \varepsilon_{it}}_{\text{aleatorio}}} \quad (6.1)$$

para  $i = 1, \dots, I$  y para  $t = 1, \dots, n_i$  y de modo que

$$T_i \text{ son i.i.d. según una } N(0, \sigma_T) \text{ para todo } i = 1, \dots, I, \quad (6.2)$$

$$\varepsilon_{it} \text{ son i.i.d. según una } N(0, \sigma) \text{ para todo } i = 1, \dots, I \text{ y } t = 1, \dots, n_i, \quad (6.3)$$

$$\text{las } T_i \text{'s y } \varepsilon_{it} \text{'s son todas mutuamente independientes} \quad (6.4)$$

Comparando el modelo descrito por (6.1), (6.2), (6.3) y (6.4) con el modelo de una vía de efectos fijos dado por (2.2) en la sección 2.3, se observa que la forma de ambos modelos y las hipótesis sobre el error son exactamente iguales. La única diferencia estriba en la modelización del efecto tratamiento. Dado que el  $i$ -ésimo nivel del factor tratamiento  $T$  observado en el experimento ha sido seleccionado aleatoriamente de una población infinita, su efecto observado es en realidad la observación de una variable aleatoria  $T_i$ . Naturalmente, la distribución de  $T_i$  es la distribución de los efectos tratamiento en la población completa y, en (6.2), se supone que esta es normal con varianza  $\sigma_T^2$ . Por supuesto esta nueva suposición también deberá ser chequeada con los residuos del modelo. La media de la población de los efectos del tratamiento ha sido asimilada en la constante  $\mu$  y ello justifica que la media de las  $T_i$  sea cero. Pero, entonces, el parámetro de interés pasa a ser ahora  $\sigma_T^2$ . Su importancia es fundamental: si todos los efectos de *la totalidad* de niveles del factor tratamiento son iguales, entonces  $\sigma_T^2 = 0$ . Si, por el contrario, existen niveles con efectos muy diferentes, entonces  $\sigma_T^2$  es muy grande.

La suposición de independencia dada en (6.4) equivale a asumir que el factor tratamiento no tiene incidencia sobre cualquier fuente de variación que haya sido englobada en el error.

En el modelo actual se tiene:

$$E(Y_{it}) = E(\mu) + E(T_i) + E(\varepsilon_{it}) = \mu$$

$$\text{Var}(Y_{it}) = \text{Var}(T_i) + \text{Var}(\varepsilon_{it}) + 2\text{Cov}(T_i, \varepsilon_{it}) = \sigma_T^2 + \sigma^2$$

dado que, por la independencia de  $T_i$  y  $\varepsilon_{it}$ , la  $\text{Cov}(T_i, \varepsilon_{it}) = 0$ .

Por tanto se concluye que

$$\boxed{Y_{it} \sim N\left(\mu, \sqrt{\sigma_T^2 + \sigma^2}\right)} \quad (6.5)$$

Las dos componentes de la varianza de  $Y_{it}$ :  $\sigma_T^2$  y  $\sigma^2$  se denominan **componentes de la varianza**.

Obsérvese que, al contrario que en el modelo de efectos fijos, observaciones pertenecientes a un mismo tratamiento son correladas:

$$\text{Cov}(Y_{it}, Y_{is}) = \text{Cov}(\mu + T_i + \varepsilon_{it}, \mu + T_i + \varepsilon_{is}) = \text{Var}(T_i) = \sigma_T^2$$

### 6.3.2 Estimación de $\sigma^2$

Dada la similitud entre los modelos de una vía de efectos aleatorios y de efectos fijos, cabe preguntarse si la varianza residual obtenida en el capítulo 2 para el modelo de efectos fijos resulta, también en el modelo de efectos aleatorios, un estimador insesgado de  $\sigma^2$ . Veamos a continuación que en efecto así es.

En (2.8) se obtiene la expresión para la suma de cuadrados residual en un modelo de una vía de efectos fijos:

$$SCR = \sum_{i=1}^I \sum_{t=1}^{n_i} (Y_{it} - \bar{Y}_i)^2 = \sum_{i=1}^I \sum_{t=1}^{n_i} Y_{it}^2 - \sum_{i=1}^I n_i \bar{Y}_i^2$$

Con respecto a los sumandos del primer bloque de la expresión anterior se tiene:

$$E(Y_{it}^2) = \text{Var}(Y_{it}) + (E(Y_{it}))^2 = (\sigma_T^2 + \sigma^2) + \mu^2$$

Y con los del segundo bloque se tiene:

$$\begin{aligned} \bar{Y}_i &= \mu + T_i + \frac{1}{n_i} \sum_{t=1}^{n_i} \varepsilon_{it} \implies \\ \text{Var}(\bar{Y}_i) &= \sigma_T^2 + \frac{\sigma^2}{n_i} \quad \text{y} \quad E(\bar{Y}_i) = \mu \end{aligned} \quad (6.6)$$

Por tanto

$$E(\bar{Y}_i^2) = \sigma_T^2 + \frac{\sigma^2}{n_i} + \mu^2 \quad (6.7)$$

Concluyendo que, si  $n = \sum_{i=1}^I n_i$ ,

$$\begin{aligned} E(SCR) &= \sum_{i=1}^I \sum_{t=1}^{n_i} (\sigma_T^2 + \sigma^2 + \mu^2) - \sum_{i=1}^I n_i \left( \sigma_T^2 + \frac{\sigma^2}{n_i} + \mu^2 \right) \\ &= n\sigma^2 - I\sigma^2 = (n - I)\sigma^2 \end{aligned}$$

Así pues, definiendo la varianza residual como en el modelo de efectos fijos,  $SCMR = SCR/(n - I)$ , se verifica que  $E[SCMR] = E[SCR/(n - I)] = \sigma^2$  y la varianza residual es un estimador insesgado de  $\sigma^2$  también en el modelo de efectos aleatorios.

### 6.3.3 Estimación de $\sigma_T^2$

Si el valor de  $scmR$  en un modelo de efectos fijos ha funcionado bien para estimar  $\sigma^2$  en un modelo de efectos aleatorios, ¿funcionará también la varianza explicada por tratamientos,  $scmT$ , como estimador de  $\sigma_T^2$ ? Ahora la respuesta no es “del todo” afirmativa, si bien podremos “usar”  $scmT$  para encontrar un estimador apropiado.

En el capítulo 2 se define  $scmT = scT/(I - 1)$ , donde  $scT$  está dado por:

$$scT = \sum_{i=1}^I n_i \bar{y}_i^2 - n \bar{y}_{..}^2$$

con la correspondiente variable aleatoria

$$SCT = \sum_{i=1}^I n_i \bar{Y}_i^2 - n \bar{Y}_{..}^2 \tag{6.8}$$

Los mismos tipos de cálculos de la sección previa conducen a:

$$\bar{Y}_{..} = \mu + \frac{1}{n} \sum_{i=1}^I n_i T_i + \frac{1}{n} \sum_{i=1}^I \sum_{t=1}^{n_i} \varepsilon_{it}$$

Entonces

$$Var(\bar{Y}_{..}) = \frac{\sigma_T^2}{n^2} \sum_{i=1}^I n_i^2 + \frac{n}{n^2} \sigma^2 \quad y \quad E(\bar{Y}_{..}) = \mu$$

lo que junto con (6.7) permite concluir en (6.8)

$$\begin{aligned} E(SCT) &= \sum_{i=1}^I n_i \left( \sigma_T^2 + \frac{\sigma^2}{n_i} + \mu^2 \right) - n \left( \frac{\sigma_T^2}{n^2} \sum_{i=1}^I n_i^2 + \frac{\sigma^2}{n} + \mu^2 \right) \\ &= \left( n - \frac{1}{n} \sum_{i=1}^I n_i^2 \right) \sigma_T^2 + (I - 1)\sigma^2 \end{aligned}$$

Como  $SCMT = SCT / (I - 1)$ , se tiene

$$E(SCMT) = c\sigma_T^2 + \sigma^2, \quad \text{donde} \quad c = \frac{n^2 - \sum_{i=1}^I n_i^2}{n(I-1)}$$

Obsérvese que si todos los  $n_i$  son iguales e iguales a  $r$ , entonces  $n = Ir$  y  $c = r$ .

Queda claro entonces que  $SCMT$  es un estimador insesgado de  $c\sigma_T^2 + \sigma^2$ , no de  $\sigma_T^2$ . Sin embargo, se puede obtener un estimador insesgado de  $\sigma_T^2$  sin más que considerar que

$$E \left[ \frac{SCMT - SCMR}{c} \right] = \sigma_T^2 \quad (6.9)$$

La expresión (6.9) muestra que el estimador obtenido puede arrojar estimaciones negativas de  $\sigma_T^2$ , aún cuando  $\sigma_T^2$  nunca puede ser negativo. Esto ocurrirá siempre que  $scmR$  sea mayor que  $scmT$ , lo que a su vez será muy probable que suceda cuando  $\sigma_T^2$  sea muy próxima a cero. Si  $scmR$  es mucho mayor que  $scmT$  entonces nuestro modelo será muy cuestionable y probablemente no proporcione una buena descripción de los datos.

### 6.3.4 Contraste de Igualdad de los Efectos Tratamiento

En el modelo de efectos aleatorios el interés se centra en la variabilidad de la totalidad de posibles efectos del factor tratamiento y no sólo de los efectos de los niveles que intervienen en el experimento. De ahí que la prueba de hipótesis de interés sea:

$$\begin{cases} H_0^T : \sigma_T^2 = 0 \\ H_1^T : \sigma_T^2 > 0 \end{cases}$$

¿Es posible utilizar la misma estadística de prueba o medida de discrepancia para resolver esta prueba de hipótesis que la empleada para contrastar si los efectos eran iguales en un modelo de efectos fijos? En otros términos, ¿es una forma de proceder adecuada rechazar la hipótesis nula en favor de la alternativa siempre que  $\frac{scmT}{scmR} > F_{I-1, n-I, \alpha}$ ?

Las dos secciones previas nos permiten examinar el comportamiento del cociente anterior para el modelo actual de efectos aleatorios. En la sección 6.3.2 se ha probado que:

$$E(SCMR) = \sigma^2$$

y en la sección 6.3.3 se ha probado que:

$$E(SCMT) = c\sigma_T^2 + \sigma^2,$$

donde  $c = \frac{n^2 - \sum_{i=1}^I n_i^2}{n(I-1)}$ , y  $c = r$  cuando todos los  $n_i$  son iguales e iguales a  $r$ .

Entonces, si  $H_0^T$  es cierta, el valor esperado del cociente  $\frac{scmT}{scmR}$  es 1 y, si  $H_1^T$  es cierta, el cociente anterior tomará valores grandes y positivos. La situación es por tanto similar al caso de efectos fijos. El único punto que resta por comprobar es que el comportamiento en distribución bajo  $H_0^T$  del mencionado cociente sea el de una  $F$  de Fisher-Snedecor con  $I-1$  y  $n-I$  grados de libertad.

Se puede probar que

$$\frac{SCT}{c\sigma_T^2 + \sigma^2} \sim \chi_{I-1}^2 \quad (6.10)$$

$$\frac{SCR}{\sigma^2} \sim \chi_{n-I}^2 \quad (6.11)$$

y que ambos estadísticos son independientes, de modo que

$$\frac{SCT / ((c\sigma_T^2 + \sigma^2)(I-1))}{SCR / (\sigma^2(n-I))} = \frac{SCMT / (c\sigma_T^2 + \sigma^2)}{SCMR / \sigma^2} \sim F_{I-1, n-I} \quad (6.12)$$

Entonces, efectivamente, cuando  $H_0^T$  es cierta se verifica que

$$\frac{SCMT}{SCMR} \sim F_{I-1, n-I}$$

Por tanto el cuadro de análisis de la varianza para el modelo de una vía de efectos aleatorios (ver Tabla 6.1) es prácticamente igual al de efectos fijos con la diferencia de los valores esperados de las sumas de cuadrados medios.

Muchas veces resulta más interesante contrastar si la varianza de la población de tratamientos es menor o igual que alguna proporción de la varianza del error, es decir, resolver la prueba de hipótesis:

$$\begin{cases} H_0^{\gamma T} : \sigma_T^2 \leq \gamma\sigma^2 \\ H_1^{\gamma T} : \sigma_T^2 > \gamma\sigma^2 \end{cases}$$

para alguna constante  $\gamma > 0$ .

En tal caso, la estadística de prueba se modificaría como sigue:

$$\text{Rechazar } H_0^{\gamma T} \text{ a un nivel de significación } \alpha \text{ si } \frac{scmT}{scmR} > (c\gamma + 1) F_{I-1, n-I, \alpha} \quad (6.13)$$

| CUADRO DEL ANÁLISIS DE LA VARIANZA<br>— UNA VÍA —                                  |  |             |                            |                          |                     |
|--|--|-------------|----------------------------|--------------------------|---------------------|
| Fuente de Variación  | Suma de Cuadrados  | <i>g.l.</i> | <i>scm</i>                 | $E(SCM)$                 | $\hat{F}$           |
| Tratamientos   | $scT = \sum_{i=1}^I \sum_{t=1}^{n_i} (\bar{y}_i - \bar{y}_{..})^2$ | $I - 1$     | $scmT = \frac{scT}{I - 1}$ | $c\sigma_T^2 + \sigma^2$ | $\frac{scmT}{scmR}$ |
| Residual   | $scR = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_i)^2$       | $n - I$     | $scmR = \frac{scR}{n - I}$ | $\sigma^2$               |                     |
| Global   | $scG = \sum_{i=1}^I \sum_{t=1}^{n_i} (y_{it} - \bar{y}_{..})^2$    | $n - 1$     | $scmG = \frac{scG}{n - 1}$ |                          |                     |
| $c = \frac{n^2 - \sum_{i=1}^I n_i^2}{n(I - 1)}$                                    |  |             |                            |                          |                     |
| Rechazar $H_0 : \sigma_T^2 = 0$ en base al p-valor $p = P(\hat{F} > F_{I-1, n-I})$ |  |             |                            |                          |                     |

Tabla 6.1: Cuadro del análisis de la varianza para un diseño completamente aleatorizado de efectos aleatorios.

### 6.3.5 Intervalos de Confianza para las Componentes de la Varianza

Demostrado que la variable respuesta,  $Y_{it}$ , en un modelo de una vía con efectos aleatorios tiene un comportamiento en distribución normal con varianza  $\sigma_T^2 + \sigma^2$ , parece interesante encontrar un intervalo de confianza para  $\sigma_T^2$  (varianza de la población de efectos-tratamiento) o, alternativamente, para  $\frac{\sigma_T^2}{\sigma^2}$  (varianza relativa de la población de efectos-tratamiento con respecto a la varianza del error).

#### Intervalo de Confianza para $\sigma_T^2/\sigma^2$

De (6.12) se sigue que:

$$\frac{SCMT}{SCMR (c\sigma_T^2/\sigma^2 + 1)} \sim F_{I-1, n-I}. \quad (6.14)$$

A partir de la expresión anterior se deduce:

$$P \left( F_{I-1, n-I, 1-\alpha/2} \leq \frac{SCMT}{SCMR (c\sigma_T^2/\sigma^2 + 1)} \leq F_{I-1, n-I, \alpha/2} \right) = 1 - \alpha$$

Y, equivalentemente, un intervalo de confianza para el cociente  $\sigma_T^2/\sigma^2$  al nivel de confianza  $100(1 - \alpha)\%$  viene dado por:

$$\boxed{\frac{1}{c} \left( \frac{SCMT}{SCMR F_{I-1, n-I, \alpha/2}} - 1 \right) \leq \frac{\sigma_T^2}{\sigma^2} \leq \frac{1}{c} \left( \frac{SCMT}{SCMR F_{I-1, n-I, 1-\alpha/2}} - 1 \right)} \quad (6.15)$$

con  $c = \frac{n^2 - \sum_{i=1}^I n_i^2}{n(I-1)}$ , y  $c = r$  cuando todos los  $n_i$  son iguales e iguales a  $r$ .

### Intervalo de Confianza para $\sigma_T^2$

Existen varios métodos para obtener un intervalo de confianza para la varianza de la población de efectos-tratamiento y en esta sección se presenta uno que es especialmente útil cuando  $\sigma_T^2$  no es próximo a cero y que puede adaptarse fácilmente a modelos más complicados.

En primer lugar recuérdese que se obtuvo un estimador insesgado de  $\sigma_T^2$  en (6.9):

$$U = c^{-1} (SCMT - SCMR) \quad (6.16)$$

Luego, si es posible determinar la distribución del estadístico  $U$ , se deduce fácilmente un intervalo de confianza para  $\sigma^2$ . En realidad, bajo las hipótesis del modelo de efectos aleatorios, la distribución exacta de  $U$  está basada en la diferencia de dos distribuciones  $\chi^2$  independientes, multiplicadas cada una de ellas por un valor constante desconocido. El resultado es una distribución compleja y difícil de tabular. Sin embargo se prueba que una aproximación razonable a la verdadera distribución de  $U/\sigma_T^2$  es una distribución  $\chi^2$  dividida por sus grados de libertad  $x$ , donde  $x$  se estima mediante la expresión:

$$x = \frac{(scmT - scmR)^2}{\frac{scmT^2}{I-1} + \frac{scmR^2}{n-I}} \quad (6.17)$$

En otras palabras, la distribución de  $\frac{xU}{E(U)}$  es aproximadamente  $\chi_x^2$ . Usando esta aproximación se tiene:

$$P \left( \chi_{x, 1-\alpha/2}^2 \leq \frac{xU}{\sigma_T^2} \leq \chi_{x, \alpha/2}^2 \right) \approx 1 - \alpha$$

Y, por tanto, un intervalo de confianza al nivel aproximado del  $100(1 - \alpha)\%$  para  $\sigma_T^2$  es:

$$\boxed{\frac{xu}{\chi_{x,\alpha/2}^2} \leq \sigma_T^2 \leq \frac{xu}{\chi_{x,1-\alpha/2}^2}} \quad (6.18)$$

donde  $u$  es el valor observado de  $U$ , es decir,  $u = c^{-1}(scmT - scmR)$ .

## 6.4 Tamaños Muestrales para un Experimento con Un Efecto Aleatorio

Con el modelo de efectos fijos se siguieron dos caminos para determinar el tamaño muestral apropiado en cada nivel del factor tratamiento: uno basado en requerir de antemano una potencia en el test Anova para detectar que dos tratamientos difieren en más de una cantidad prefijada  $\Delta$  y otro basado en obtener longitudes concretas para los intervalos de confianza simultáneos de uno o más contrastes.

Para el modelo de efectos aleatorios es necesario determinar tanto el tamaño muestral de cada nivel  $r_i$  como el número de niveles  $I$ . El problema es por tanto bastante más complejo y no tiene una solución tan razonablemente sencilla como en el caso de efectos fijos.

Una primera aproximación para el cálculo de  $I$  y de  $r$  (supondremos que  $r_i = r \forall i$ ) pasa por investigar la forma de la varianza de estimadores insesgados de  $\sigma_T^2$  o de  $\sigma_T^2/\sigma^2$  y tratar de elegir  $r$  e  $I$  en orden a hacer dicha varianza pequeña.

Empecemos por  $\sigma_T^2$ . Según (6.9), un estimador insesgado de  $\sigma_T^2$  está dado por

$$U = r^{-1}(SCMT - SCMR),$$

que tiene como varianza:

$$Var(U) = \left(\frac{2n^2}{I^2}\right) \left(\frac{(n\sigma_T^2/I + \sigma^2)^2}{I-1} + \frac{\sigma^4}{n-I}\right) \quad (6.19)$$

para  $n > I$ . Tal conclusión se deriva de que  $SCMT$  y  $SCMR$  son independientes,  $SCT/(r\sigma_T^2 + \sigma^2) \sim \chi_{I-1}^2$ ,  $SCR/\sigma^2 \sim \chi_{n-I}^2$  y la varianza de una  $\chi^2$  es igual a dos veces sus grados de libertad.

Supóngase que por razones presupuestarias el tamaño global de observaciones  $n = Ir$  es fijo y se desean seleccionar  $I$  y  $r$  tales que la  $Var(U)$  sea tan pequeña como sea posible. Si  $\sigma_T^2$  es mucho mayor que  $\sigma^2$ , el primer sumando en la derecha de la ecuación (6.19) es quién gobierna el tamaño de la  $Var(U)$ . Se observa que se hace pequeña si se toma  $I$  lo más grande posible. Incluso si  $\sigma_T^2$  es mucho menor que  $\sigma^2$ , la  $Var(U)$  se sigue minimizando para valores grandes de  $I$ . Tomando el caso más extremo en que  $I = n$  y  $r = 1$ , la  $Var(U) = 2\sigma_T^4/(I-1)$ , que es el valor más pequeño posible que puede tomar.

Si se atiende a minimizar la varianza de un estimador del cociente  $\sigma_T^2/\sigma^2$ , la elección óptima de  $r$  e  $I$  es diferente. Del hecho de que la media de una distribución  $F_{I-1, n-I}$  es  $(n-I)/(n-I-2)$  y de (6.14) se concluye que un estimador insesgado de  $\sigma_T^2/\sigma^2$  está dado por:

$$V = \frac{1}{r} \left( \frac{(n-I-2)}{(n-I)} \frac{SCMT}{SCMR} - 1 \right)$$

y, dada su distribución  $F$ , se deduce que su varianza es:

$$Var(V) = \left( \frac{\sigma_T^2}{\sigma^2} + \frac{1}{r} \right)^2 \left( \frac{2(n-3)}{(I-1)(n-I-4)} \right)$$

Así, si el número  $n$  de observaciones es fijo y esperamos que  $\sigma_T^2 \geq \sigma^2$ , entonces el término  $(\sigma_T^2/\sigma^2)^2$  del primer paréntesis de la expresión anterior es el que domina el tamaño de la varianza de  $V$ . Para minimizarla habrá que minimizar su coeficiente, lo que requiere tomar  $I = (n-3)/2$ . Entonces una posible solución sería tomar  $I = n/2$  y  $r = 2$ . Por otro lado, en el caso más extraño en que se espera que  $\sigma_T^2$  sea mucho más pequeño que  $\sigma^2$ , entonces el término  $(1/r)^2$  es el de mayor peso para determinar la varianza de  $V$ . En tal caso el correspondiente coeficiente se minimiza tomando  $I$  lo más pequeño posible.

Un último camino para tener un idea de que valores de  $I$  y  $r$  son apropiados sería examinar la potencia de la estadística de prueba propuesta para resolver la prueba de hipótesis:

$$\begin{cases} H_0^{\gamma T} : \sigma_T^2 \leq \gamma\sigma^2 \\ H_1^{\gamma T} : \sigma_T^2 > \gamma\sigma^2 \end{cases}$$

para alguna constante  $\gamma \geq 0$ .

En (6.13) se concluye:

$$\text{Rechazar } H_0^{\gamma T} \text{ a un nivel de significación } \alpha \text{ si } \frac{scmT}{scmR} > (c\gamma + 1) F_{I-1, n-I, \alpha} = k$$

¿Cuál es la probabilidad de rechazar  $H_0^{\gamma T}$  si el verdadero valor de  $\sigma_T^2/\sigma^2$  es igual a  $\Delta$ ? En otros términos, ¿cuál es la probabilidad de que  $SCMT/SCMR > k$  cuando  $\sigma_T^2/\sigma^2$  es igual a  $\Delta$ ?

Dado que  $\frac{SCMT}{SCMR(c\sigma_T^2/\sigma^2 + 1)} \sim F_{I-1, n-I}$ , cuando  $\frac{\sigma_T^2}{\sigma^2} = \Delta$  se tiene

$$P \left( \frac{SCMT}{SCMR} > k \right) = P \left( \frac{SCMT}{SCMR(c\Delta + 1)} > \frac{k}{c\Delta + 1} \right)$$

Si se desea que la potencia sea  $\pi$  cuando  $\frac{\sigma_T^2}{\sigma^2} = \Delta$ , entonces debe ocurrir que:

$$\frac{k}{c\Delta + 1} = F_{I-1, n-I, \pi}$$

Como  $k = (c\gamma + 1) F_{I-1, n-I, \alpha}$ ,  $n - I = I(r - 1)$  y  $c = r$ , se obtiene la igualdad:

$$\frac{F_{I-1, I(r-1), \alpha}}{F_{I-1, I(r-1), \pi}} = \frac{r\Delta + 1}{r\gamma + 1}$$

Por consiguiente, además de  $\Delta$  y  $\pi$ , es preciso seleccionar  $\alpha$  y  $\gamma$  para contrastar  $H_0^{\gamma T}$ . A partir de ahí se podrían determinar los valores  $i$  y  $r$  mediante pruebas de ensayo y error hasta aproximar lo máximo posible la desigualdad:

$$F_{I-1, I(r-1), \alpha} F_{I(r-1), I-1, 1-\pi} \leq \frac{r\Delta + 1}{r\gamma + 1} \tag{6.20}$$

## 6.5 Dos o Más Efectos Aleatorios

El modelo de una vía de efectos aleatorios puede extenderse al caso de más factores tratamiento con efectos aleatorios en un camino semejante al caso de efectos fijos. Veamos a continuación algunos modelos.

### 6.5.1 Algunos Modelos

**Modelo de dos vías de efectos principales.**

$$\underbrace{Y_{ijt}}_{\text{aleatorio}} = \underbrace{\mu}_{\text{determinista}} + \underbrace{A_i + B_j + \varepsilon_{ijt}}_{\text{aleatorio}} \tag{6.21}$$

de modo que para todo  $t = 1, \dots, n_{ij}$ ,  $i = 1, \dots, a$  y  $j = 1, \dots, b$  se verifica:

$$A_i \sim N(0, \sigma_A), \quad B_j \sim N(0, \sigma_B), \quad \varepsilon_{ijt} \sim N(0, \sigma)$$

las variables  $A_i$ 's,  $B_j$ 's y  $\varepsilon_{ijt}$ 's son mutuamente independientes

**Modelo completo de dos vías.**

$$\underbrace{Y_{ijt}}_{\text{aleatorio}} = \underbrace{\mu}_{\text{determinista}} + \underbrace{A_i + B_j + (AB)_{ij} + \varepsilon_{ijt}}_{\text{aleatorio}} \tag{6.22}$$

de modo que para todo  $t = 1, \dots, n_{ijk}$ ,  $i = 1, \dots, a$  y  $j = 1, \dots, b$  se verifica:

$$A_i \sim N(0, \sigma_A), \quad B_j \sim N(0, \sigma_B), \quad (AB)_{ij} \sim N(0, \sigma_{AB}) \quad \varepsilon_{ijt} \sim N(0, \sigma)$$

las variables  $A_i$ 's,  $B_j$ 's,  $(AB)_{ij}$ 's y  $\varepsilon_{ijt}$ 's son mutuamente independientes

**Modelo completo de tres vías.**

$$\underbrace{Y_{ijkt}}_{\text{aleatorio}} = \underbrace{\mu}_{\text{determinista}} + \underbrace{A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + \varepsilon_{ijkt}}_{\text{aleatorio}} \quad (6.23)$$

de modo que para todo  $t = 1, \dots, n_{ijk}$ ,  $i = 1, \dots, a$ ,  $j = 1, \dots, b$  y  $k = 1, \dots, c$  se verifica:

$$A_i \sim N(0, \sigma_A), \quad B_j \sim N(0, \sigma_B), \quad C_k \sim N(0, \sigma_C)$$

$$(AB)_{ij} \sim N(0, \sigma_{AB}), \quad (AC)_{ik} \sim N(0, \sigma_{AC}), \quad (BC)_{jk} \sim N(0, \sigma_{BC}) \\ (ABC)_{ijk} \sim N(0, \sigma_{ABC}) \text{ y } \varepsilon_{ijkt} \sim N(0, \sigma)$$

las variables  $A_i$ 's,  $B_j$ 's,  $C_k$ 's,  $(AB)_{ij}$ 's,  $(AC)_{ik}$ 's,  $(BC)_{jk}$ 's,  $(ABC)_{ijk}$ 's y  $\varepsilon_{ijt}$ 's son mutuamente independientes

### 6.5.2 Estimación de $\sigma^2$

Igual que en el caso de un solo factor, la  $scmR$ , obtenida de igual forma que en el correspondiente modelo de efectos fijos, es un estimador insesgado de  $\sigma^2$  en cualquier modelo de factores cruzados de efectos aleatorios. Obsérvese que este resultado no debe sorprender en absoluto ya que, al fin y al cabo,  $scmR$  mide la variabilidad en los datos que no es achacable a aquellas fuentes de variación incluidas en el experimento.

### 6.5.3 Estimación de las Componentes de la Varianza

En la sección 6.3.3 se demostró que, para un modelo de una vía de efectos aleatorios,

$$E(SCMT) = c\sigma_T^2 + \sigma^2, \quad \text{con} \quad c = \frac{n^2 - \sum_{i=1}^a n_i^2}{n(a-1)}$$

donde  $SCMT$  es la varianza explicada por diferencias entre los efectos-tratamiento en el modelo de una vía de efectos fijos. A partir de este resultado derivabamos un estimador insesgado para  $\sigma_T^2$ , concretamente  $(SCMT - SCMR) / c$ .

Para modelos más complejos también es posible encontrar estimadores insesgados de las componentes de la varianza en base a las sumas de cuadrados esperadas en el correspondiente modelo de efectos fijos; sin embargo, cada estimador debe ser calculado individualmente.

**Ejemplo 6.3** *Estimador insesgado de  $\sigma_B^2$ .* Supóngase un experimento con tres factores tratamiento de efectos aleatorios  $A$ ,  $B$  y  $C$  con  $a$ ,  $b$  y  $c$  niveles respectivamente. Supóngase además que se toman  $r$  observaciones en cada una de las  $I = abc$  posibles combinaciones de niveles de los tres factores. Por último, supóngase que solo es razonable esperar como interacciones significativas las que se dan entre los factores  $AB$  y  $BC$ . El modelo es:

$$Y_{ijkl} = \mu + A_i + B_j + C_k + (AB)_{ij} + (BC)_{jk} + \varepsilon_{ijkl}$$

para  $t = 1, \dots, r$ ,  $i = 1, \dots, a$ ,  $j = 1, \dots, b$  y  $k = 1, \dots, c$  y con las hipótesis correspondientes acerca de las distribuciones de las variables efectos tratamiento y error.

¿Cómo obtener un estimador insesgado para la varianza de la población de efectos de los niveles del factor  $B$ ,  $\sigma_B^2$ ? En primer lugar examinemos  $E(SCMB)$ , siendo  $SCMB = SCB / (b - 1)$ . En el correspondiente modelo de efectos fijos:

$$scB = acr \sum_{j=1}^b \bar{y}_{.j..}^2 - abc r \bar{y}_{....}^2$$

Pero

$$E(\bar{Y}_{.j..}) = E(\bar{Y}_{....}) = \mu$$

y

$$\begin{aligned} Var(\bar{Y}_{.j..}) &= \frac{\sigma_A^2}{a} + \sigma_B^2 + \frac{\sigma_C^2}{c} + \frac{\sigma_{AB}^2}{a} + \frac{\sigma_{BC}^2}{c} + \frac{\sigma^2}{acr} \\ Var(\bar{Y}_{....}) &= \frac{\sigma_A^2}{a} + \frac{\sigma_B^2}{b} + \frac{\sigma_C^2}{c} + \frac{\sigma_{AB}^2}{ab} + \frac{\sigma_{BC}^2}{bc} + \frac{\sigma^2}{abc} \end{aligned}$$

De donde,

$$\begin{aligned} E(SCB) &= acr \sum_{j=1}^b \left[ Var(\bar{Y}_{.j..}) + (E(\bar{Y}_{.j..}))^2 \right] - abc r \left[ Var(\bar{Y}_{....}) + (E(\bar{Y}_{....}))^2 \right] \\ &= acr(b - 1)\sigma_B^2 + cr(b - 1)\sigma_{AB}^2 + ar(b - 1)\sigma_{BC}^2 + (b - 1)\sigma^2 \end{aligned}$$

De modo que

$$E(SCMB) = acr\sigma_B^2 + cr\sigma_{AB}^2 + ar\sigma_{BC}^2 + \sigma^2 \quad (6.24)$$

Se concluye que si se quiere encontrar un estimador insesgado para  $\sigma_B^2$ , es preciso previamente encontrar estimadores insesgados de  $\sigma_{AB}^2$  y de  $\sigma_{BC}^2$ . Naturalmente, el camino lógico para obtenerlos es por medio de  $E(SCM(AB))$  y de  $E(SCM(BC))$  respectivamente.

Para ello otra vez se parte de las correspondientes sumas de cuadrados en el modelo de efectos fijos análogo. Se tiene:

$$\begin{aligned} sc(AB) &= cr \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{ij..}^2 - bcr \sum_{i=1}^a \bar{y}_{i...}^2 - acr \sum_{j=1}^b \bar{y}_{.j..}^2 + abcr \bar{y}_{....}^2 \\ sc(BC) &= ar \sum_{j=1}^b \sum_{k=1}^c \bar{y}_{.jk.}^2 - acr \sum_{j=1}^b \bar{y}_{.j..}^2 - abr \sum_{k=1}^c \bar{y}_{..k.}^2 + abcr \bar{y}_{....}^2 \end{aligned}$$

y

$$E(\bar{Y}_{ij..}) = E(\bar{Y}_{.jk.}) = E(\bar{Y}_{i...}) = E(\bar{Y}_{.j..}) = E(\bar{Y}_{..k.}) = E(\bar{Y}_{....}) = \mu$$

$$Var(\bar{Y}_{ij..}) = \sigma_A^2 + \sigma_B^2 + \frac{\sigma_C^2}{c} + \sigma_{AB}^2 + \frac{\sigma_{BC}^2}{c} + \frac{\sigma^2}{cr}$$

$$Var(\bar{Y}_{i...}) = \sigma_A^2 + \frac{\sigma_B^2}{b} + \frac{\sigma_C^2}{c} + \frac{\sigma_{AB}^2}{b} + \frac{\sigma_{BC}^2}{bc} + \frac{\sigma^2}{bcr}$$

$$Var(\bar{Y}_{.jk.}) = \frac{\sigma_A^2}{a} + \sigma_B^2 + \sigma_C^2 + \frac{\sigma_{AB}^2}{a} + \sigma_{BC}^2 + \frac{\sigma^2}{ar}$$

$$Var(\bar{Y}_{..k.}) = \frac{\sigma_A^2}{a} + \frac{\sigma_B^2}{b} + \sigma_C^2 + \frac{\sigma_{AB}^2}{ab} + \frac{\sigma_{BC}^2}{b} + \frac{\sigma^2}{abr}$$

y también

$$\begin{aligned} E(SC(AB)) &= \left( cr \sum_{i=1}^a \sum_{j=1}^b \bar{Y}_{ij..}^2 - bcr \sum_{i=1}^a \bar{Y}_{i...}^2 \right) - E(SCB) \\ &= acr(b-1)\sigma_B^2 + acr(b-1)\sigma_{AB}^2 + ar(b-1)\sigma_{BC}^2 + a(b-1)\sigma^2 \\ &\quad - acr(b-1)\sigma_B^2 - cr(b-1)\sigma_{AB}^2 - ar(b-1)\sigma_{BC}^2 - (b-1)\sigma^2 \\ &= cr(a-1)(b-1)\sigma_{AB}^2 + (a-1)(b-1)\sigma^2 \end{aligned}$$

De modo que

$$E(SCM(AB)) = cr\sigma_{AB}^2 + \sigma^2 \tag{6.25}$$

Y, análogamente,

$$E(SCM(BC)) = ar\sigma_{BC}^2 + \sigma^2 \tag{6.26}$$

Concluyendo que un estimador insesgado para  $\sigma_B^2$  es

$$U = (SCMB - SCM(AB) - SCM(BC) + SCMR) / acr$$

y una estimación insesgada para  $\sigma_B^2$  es por tanto:

$$u = (scmB - scm(AB) - scm(BC) + scmR) / acr$$

Como se deduce del ejemplo precedente, calcular una *SCM* es bastante engorroso y pesado. Sin embargo, cuando los tamaños muestrales son iguales se puede aplicar una regla bastante sencilla y mecánica. Nótese en el ejemplo anterior que todas las componentes de la varianza que aparecen en la expresión (6.24) para  $E(SCMB)$  son aquellas que incluyen algún subíndice igual al empleado para el factor *B* en el modelo. En concreto, hemos empleado el subíndice *j* para *B*, y *j* se emplea también como tal en  $(AB)_{ij}$ ,  $(BC)_{jk}$  y  $\varepsilon_{ijkt}$ . Por otro lado, el coeficiente que multiplica a cada componente de la varianza es el número de observaciones tomadas para cada combinación de subíndices; es decir, hay *acr* observaciones en cada uno de los *b* niveles de *B*, *cr* observaciones en cada uno de los *ab* niveles de *AB* y así sucesivamente.

### 6.5.4 Intervalos de Confianza para las Componentes de la Varianza

De la sección anterior se deduce que se puede calcular un estimador insesgado para cualquiera de las componentes de la varianza, digamos  $\sigma_*^2$ , a partir de la lista de sumas de cuadrados medios esperados. De hecho, tal estimador es siempre una combinación lineal de las sumas de cuadrados:  $U = \sum_i k_i (SCM)_i$ ,  $k_i$  constante.

Una aproximación a la distribución de  $xU/\sigma_*^2$  es una  $\chi^2$  con *x* grados de libertad, donde

$$x = \frac{\left( \sum_i k_i (scm)_i \right)^2}{\sum_i k_i^2 (scm)_i^2 / x_i} \tag{6.27}$$

con  $x_i$  son los grados de libertad de la *i*-ésima suma de cuadrados y  $(scm)_i$  es el valor observado de  $(SCM)_i$ . Veamos a continuación un ejemplo.

**Ejemplo 6.4** Volvamos al modelo descrito en el Ejemplo 6.3. Allí se determinó que un estimador insesgado para  $\sigma_B^2$  es:

$$U = \sum_i k_i (SCM)_i = \frac{SCMB}{acr} - \frac{SCM(AB)}{acr} - \frac{SC(BC)}{acr} + \frac{SCMR}{acr}$$

Una aproximación a la distribución de  $xU/\sigma_B^2$  es una  $\chi^2$  con  $x$  grados de libertad, donde

$$\begin{aligned} x &= \frac{[(scmB - scm(AB) - scm(BC) + scmR)/acr]^2}{\frac{scmB^2}{(acr)^2(b-1)} + \frac{scm(AB)^2}{(acr)^2(a-1)(b-1)} + \frac{scm(BC)^2}{(acr)^2(b-1)(c-1)} + \frac{scmR^2}{(acr)^2(g.l.(error))}} \\ &= \frac{[scmB - scm(AB) - scm(BC) + scmR]^2}{\frac{scmB^2}{(b-1)} + \frac{scm(AB)^2}{(a-1)(b-1)} + \frac{scm(BC)^2}{(b-1)(c-1)} + \frac{scmR^2}{g.l.(error)}} \end{aligned}$$

de modo que los grados de libertad del error,  $g.l.(error)$ , se calculan como siempre: restando a los grados de libertad totales los de las fuentes de variación explicadas.

En este caso:

$$\begin{aligned} g.l.(error) &= (abc - 1) - (a - 1) - (b - 1) - (c - 1) - (a - 1)(b - 1) - (b - 1)(c - 1) \\ &= ab(cr - 1) - b(c - 1) + 1 \end{aligned}$$

Por supuesto, a partir de la distribución aproximada para  $U$  se puede construir el correspondiente intervalo de confianza que tendrá la forma:

$$\frac{xu}{\chi_{x,\alpha/2}^2} \leq \sigma_*^2 \leq \frac{xu}{\chi_{x,1-\alpha/2}^2} \quad (6.28)$$

### 6.5.5 Pruebas de Hipótesis para las Componentes de la Varianza

Para centrar la exposición seguiremos trabajando con el modelo descrito en los Ejemplos 6.3 y 6.4.

Resolver la prueba de hipótesis  $H_0^{AB} : \{\sigma_{AB}^2 = 0\}$  frente a  $H_1^{AB} : \{\sigma_{AB}^2 > 0\}$  es muy sencillo dada su similitud con el modelo de una vía. Si  $H_0^{AB}$  es cierta, la esperanza del numerador del cociente  $scm(AB)/scmR$  es  $\sigma^2$ , igual que la esperanza del denominador. En otro caso la esperanza del numerador será mayor y, por tanto, la estadística de prueba es:

$$\text{Rechazar } H_0^{AB} \text{ a un nivel de significación } \alpha \text{ si } \frac{scm(AB)}{scmR} > F_{(a-1)(b-1),g.l.(error),\alpha}$$

con los grados de libertad calculados como siempre:

$$g.l.(error) = ab(cr - 1) - b(c - 1) + 1$$

La prueba también se podría modificar en el sentido de contrastar:

$$\begin{cases} H_0^{\gamma AB} : \sigma_{AB}^2 \leq \gamma\sigma^2 \\ H_1^{\gamma AB} : \sigma_{AB}^2 > \gamma\sigma^2 \end{cases}$$

para alguna constante  $\gamma > 0$ .

En tal caso, la estadística de prueba sería:

$$\text{Rechazar } H_0^{\gamma AB} \text{ a un nivel de significación } \alpha \text{ si } \frac{scm(AB)}{scmR} > (c\gamma + 1) F_{(a-1)(b-1), g.l.(error), \alpha}$$

Y, en modo similar se resolverían las pruebas de hipótesis relativas a  $H_0^{BC}$  frente a  $H_1^{BC}$  y  $H_0^{BC}$  frente a  $H_1^{BC}$ .

Contrastar  $H_0^A : \{\sigma_A^2 = 0\}$  frente a  $H_1^A : \{\sigma_A^2 > 0\}$  es más complicado. Si seguimos empleando los mismos estadísticos que con los modelos de efectos fijos y, así, utilizamos el cociente  $scmA/scmR$  para testar  $H_0^A$ , surge un problema. Si  $H_0^A$  es cierta, la esperanza del numerador del cociente  $scm(A)/scmR$  es  $c\sigma_{AB}^2 + \sigma^2$ , mientras que la del denominador es  $\sigma^2$ . Esto sugiere dos cosas:

1. usar  $scm(AB)$  como denominador y no  $scmR$ ,
2. cuestionarse si tiene sentido contrastar  $H_0^A$  cuando la interacción  $AB$  es significativa.

Desde luego el segundo punto reproduce la misma problemática que en el caso de efectos fijos y aquí, como allí, la respuesta es que, en efecto, no tiene sentido testar  $H_0^A$  cuando la interacción  $AB$  es significativa. Luego únicamente se plantearán contrastes de los efectos principales cuando interacciones de orden superior involucrando a esos factores no sean significativas. Ahora bien, a pesar de ello, se empleará la suma de cuadrados medios de la interacción en el denominador del cociente  $F$  para testar los efectos principales, aún cuando la inclusión de la interacción en el modelo ha sido una decisión incorrecta.

En resumen, la regla de decisión para resolver la prueba de hipótesis  $H_0^A : \{\sigma_A^2 = 0\}$  frente a  $H_1^A : \{\sigma_A^2 > 0\}$  es:

$$\text{Rechazar } H_0^A \text{ a un nivel de significación } \alpha \text{ si } \frac{scm(A)}{scm(AB)} > F_{a-1, (a-1)(b-1), \alpha}$$

Nótese que, por supuesto, los segundos grados de libertad de la  $F$  se corresponden con los grados de libertad del denominador. De forma análoga se resolvería el contraste relativo a  $H_0^C : \{\sigma_C^2 = 0\}$ .

Otra vez se complican algo las cosas para encontrar un denominador adecuado para el contraste  $F$  concerniente a la prueba de hipótesis  $H_0^B : \{\sigma_B^2 = 0\}$  frente a  $H_1^B : \{\sigma_B^2 > 0\}$ . Bajo  $H_0^B$  el valor esperado de  $SCM(B)$  es:

$$E [SCMB] = c\sigma_{AB}^2 + a\sigma_{BC}^2 + \sigma^2$$

Tal y como se ha mencionado anteriormente,  $H_0^B$  solo debería ser contrastada en la creencia de que las interacciones  $BC$  y  $AB$  son ambas no significativas pero, aún siendo significativas las interacciones anteriores, nosotros estaríamos interesados en un denominador con el mismo valor esperado. Pues bien, se prueba que si  $U = SCM(AB) + SCM(BC) - SCMR$ , entonces

$$E[U] = E[SCM(AB) + SCM(BC) - SCMR] = cr\sigma_{AB}^2 + ar\sigma_{BC}^2 + \sigma^2$$

Además,  $xU / (cr\sigma_{AB}^2 + ar\sigma_{BC}^2 + \sigma^2)$  se distribuye aproximadamente como una  $\chi^2$  con  $x$  grados de libertad, calculándose  $x$  como en la sección 6.5.4, es decir,  $x$  está dado por:

$$x = \frac{[scm(AB) + scm(BC) + scmR]^2}{\frac{scm(AB)^2}{(a-1)(b-1)} + \frac{scm(BC)^2}{(b-1)(c-1)} + \frac{scmR^2}{g.l.(error)}}$$

De modo que, si  $H_0^B$  es cierta,  $scmB/u$  es aproximadamente una  $F_{b-1,x}$  y la regla de decisión para resolver la prueba de hipótesis  $H_0^B : \{\sigma_B^2 = 0\}$  frente a  $H_1^B : \{\sigma_B^2 > 0\}$  es:

Rechazar  $H_0^B$  a un nivel de significación  $\alpha$  si  $\frac{scm(B)}{scm(AB) + scm(BC) - scmR} > F_{b-1,x,\alpha}$

En general resultará muy útil listar en una tabla las sumas de cuadrados medios esperadas para el modelo particular que se investiga. La Tabla 6.2 muestra una tabla de estas características para un modelo de tres efectos aleatorios con dos interacciones.

| Fuente de Variación | Grados de Libertad | Esperanza de SCM   |
|---------------------|--------------------|--|
| A                   | $a - 1$            | $bcr\sigma_A^2 + cr\sigma_{AB}^2 + \sigma^2$                   |
| B                   | $b - 1$            | $acr\sigma_B^2 + cr\sigma_{AB}^2 + ar\sigma_{BC}^2 + \sigma^2$ |
| C                   | $c - 1$            | $abr\sigma_C^2 + ar\sigma_{BC}^2 + \sigma^2$                   |
| AB                  | $(a - 1)(b - 1)$   | $cr\sigma_{AB}^2 + \sigma^2$                                   |
| BC                  | $(b - 1)(d - 1)$   | $ar\sigma_{BC}^2 + \sigma^2$                                   |
| Error               | $g.l.(error)$      | $\sigma^2$   |

Tabla 6.2: Esperanzas de las sumas de cuadrados promediadas y grados de libertad para un modelo de tres vías de efectos aleatorios con dos interacciones de segundo orden.

## 6.6 Problemas Propuestos

**Problema 1** Para calibrar alcoholímetros se emplean soluciones de alcohol. De un enorme cargamento de botellas conteniendo soluciones de alcohol se toman al azar seis botellas.

A continuación se toma, de cada botella, un total de cuatro muestras de la solución y se determinan las concentraciones de alcohol. Los resultados del experimento figuran en el archivo ALCOHOL.SAV.

1. Chequear las hipótesis del modelo de una vía de efectos aleatorios para estos datos.
2. Calcular un intervalo de confianza al 95% para la varianza del error.
3. Calcular un intervalo de confianza al 95% para la varianza de las concentraciones de alcohol en el cargamento de botellas investigado.
4. Contrastar al 5% de significación la hipótesis nula de que la varianza de las concentraciones de alcohol en el cargamento es menor que el doble de la varianza del error frente a la alternativa de que no es así.

**Problema 2** En un estudio para determinar cuáles son las fuentes de variación importantes en un proceso industrial, tres operadores seleccionados al azar toman tres mediciones del rendimiento cuando se emplean cuatro lotes concretos de materia prima también seleccionados aleatoriamente. Los datos figuran en el archivo INDUSTRY.SAV. Determinar las componentes de la varianza en un hipotético modelo de dos vías (lotes y operadores) con interacción de efectos aleatorios para este experimento.

**Problema 3** Para estimar las diferentes componentes de variabilidad en un proceso de filtración, se midió el porcentaje de la pérdida de material en un licor madre bajo doce condiciones experimentales, tres pruebas de cada condición. Se seleccionaron aleatoriamente tres filtros y cuatro operadores para el experimento (ver archivo FILTRADO.SAV).

1. Contrastar la hipótesis de no interacción entre los filtros y los operadores al 5% de significación.
2. Contrastar la hipótesis de que los filtros y los operadores no tienen efecto sobre la variabilidad del proceso de filtración al 5% de significación.
3. Estimar las componentes de la varianza asociadas al modelo

**Problema 4** (Stacie Taylor, 1995) Se ejecutó este experimento para determinar que marcas de tostadas de pan congelado de desayuno propocionaba un pan más esponjoso al hornearse. Se examinaron tres marcas concretas (factor  $A$  con tres niveles de efectos fijos). Las rebanadas se situaron en la bandeja central de un horno y se calentaron durante siete minutos a la temperatura de 425 grados Fahrenheit. Por la capacidad de las bandejas tan solo era posible incluir un total de seis rebanadas cada vez. Se optó entonces por un diseño general en bloques completos con bloques de tamaño  $k = 6$ . Los datos figuran en el archivo BISCUIT.SAV.

1. Usar un modelo mixto con interacción para representar los datos donde: cada hornada representa un bloque (efectos aleatorios) y cada marca un tratamiento (efectos fijos). Escribir el modelo matemático incluyendo las hipótesis estructurales que se formulan sobre él.
2. Chequear las hipótesis estructurales del modelo.
3. Escribir las sumas de cuadrados medios esperadas para cada componente del modelo.
4. Dibujar un gráfico de interacciones.
5. Realizar un análisis de la varianza y extraer conclusiones.
6. Realizar las comparaciones dos a dos.

# Apéndice: Tablas

En esta sección se presentan las siguientes tablas:

**TABLA 1** POTENCIA DEL CONTRASTE  $F$ :  $\pi(\phi) = P(F_{\nu_1, \nu_2, \phi} > F_{\nu_1, \nu_2, \alpha})$

**TABLA 2** MÉTODO DE TUKEY:  $\alpha$ -CUANTILES  $q_{v, df, \alpha}$  DE LA DISTRIBUCIÓN DEL RANGO ESTUDENTIZADO

**TABLA 3** MÉTODO DE DUNNET (CONTRASTES UNILATERALES): VALORES CRÍTICOS  $\omega_{D1}$  QUE SE CORRESPONDEN CON LOS  $\alpha$ -CUANTILES  $t_{I-1, df, \alpha}^{0.5}$  DEL MÁXIMO DE LA DISTRIBUCIÓN  $t$ -MULTIVARIANTE DE DIMENSIÓN  $(I - 1)$  CON CORRELACIÓN COMÚN  $\rho = 0.5$  Y GRADOS DE LIBERTAD  $df$

**TABLA 4** MÉTODO DE DUNNET (CONTRASTE BILATERAL): VALORES CRÍTICOS  $\omega_{D2}$  QUE SE CORRESPONDEN CON LOS  $\alpha$ -CUANTILES  $|t|_{I-1, df, \alpha}^{0.5}$  DEL VALOR ABSOLUTO DEL MÁXIMO DE LA DISTRIBUCIÓN  $t$ -MULTIVARIANTE DE DIMENSIÓN  $(I - 1)$  CON CORRELACIÓN COMÚN  $\rho = 0.5$  Y GRADOS DE LIBERTAD  $df$

**TABLA 1** POTENCIA DEL CONTRASTE  $F$ :  $\pi(\phi) = P(F_{\nu_1, \nu_2, \phi} > F_{\nu_1, \nu_2, \alpha})$

**TABLA 1** POTENCIA DEL CONTRASTE  $F$ :  $\pi(\phi) = P(F_{\nu_1, \nu_2, \phi} > F_{\nu_1, \nu_2, \alpha})$

**TABLA 1** POTENCIA DEL CONTRASTE  $F$ :  $\pi(\phi) = P(F_{\nu_1, \nu_2, \phi} > F_{\nu_1, \nu_2, \alpha})$

**TABLA 1** POTENCIA DEL CONTRASTE  $F$ :  $\pi(\phi) = P(F_{\nu_1, \nu_2, \phi} > F_{\nu_1, \nu_2, \alpha})$

**TABLA 1** POTENCIA DEL CONTRASTE  $F$ :  $\pi(\phi) = P(F_{\nu_1, \nu_2, \phi} > F_{\nu_1, \nu_2, \alpha})$

**TABLA 1** POTENCIA DEL CONTRASTE  $F$ :  $\pi(\phi) = P(F_{\nu_1, \nu_2, \phi} > F_{\nu_1, \nu_2, \alpha})$

**TABLA 2** MÉTODO DE TUKEY: CUANTILES  $q_{v,df,\alpha}$  DE LA DISTRIBUCIÓN DEL RANGO ESTUDENTIZADO.

**TABLA 2** MÉTODO DE TUKEY: CUANTILES  $q_{v,df,\alpha}$  DE LA DISTRIBUCIÓN DEL RANGO ESTUDENTIZADO.

**TABLA 3** MÉTODO DE DUNNET (CONTRASTE UNILATERAL): VALORES CRÍTICOS  $\omega_{D1}$

**TABLA 3** MÉTODO DE DUNNET (CONTRASTE UNILATERAL): VALORES CRÍTICOS  $\omega_{D1}$

**TABLA 4** MÉTODO DE DUNNET (CONTRASTE BILATERAL): VALORES CRÍTICOS  $\omega_{D2}$

**TABLA 4** MÉTODO DE DUNNET (CONTRASTE BILATERAL): VALORES CRÍTICOS  $\omega_{D2}$



# Referencias Bibliográficas

- BOX, G.E.P.; HUNTER, W.G. & HUNTER, J.S. *Estadística para Investigadores. Introducción al diseño de experimentos, análisis de datos y construcción de modelos*. Ed. Reverté, Barcelona (1989).
- CLARKE, G.M., & KEMPSON, R.E. *Introduction to the design and analysis of experiments*. Arnold Publishers (1996).
- COBB, G.W. *Introduction to Design and Analysis for Experiments*. Springer-Verlag, New York (1998).
- CONOVER, W.J. *Practical Nonparametric Statistics*. 2nd ed. J. Wiley (1980).
- COX, D. & REID, N. *The Theory of the Design of Experiments*. Monographs on Statistics and Applied Probability. Chapman & Hall/ CRC Press (2000).
- DEAN, A. & VOSS, D. *Design and Analysis of Experiments*. Springer-Verlag, New York (1999).
- GIBBONS, J.D. & CHAKRABORTI, S. *Nonparametric Statistical Inference*. 3rd ed. Marcel Dekker, New York (1992).
- HICKS, C.R. *Fundamental Concepts in the Design of Experiments*. 4th ed. Oxford University Press Inc, New York (1993).
- HSU, J.C. *Multiple comparisons: Theory and methods*. Chapman and Hall, London (1996).
- KEMPTHORNE, O. *Design and Analysis of Experiments*. J. Wiley and Sons, New York (1976).
- KUEHL, R.O. *Statistical Principles of Research Design and Analysis*. Duxbury Press, Belmont, California (1994).
- MONTGOMERY, C. *Diseño y Análisis de Experimentos*. Grupo Editorial Iberoamerica (1991).

- PEÑA SÁNCHEZ DE RIVERA, D. *Estadística, modelos y métodos. 2: Modelos lineales y series temporales*. 2nd. ed. Ed. Alianza Universidad Textos (1989).
- PETER, J. *Statistical Design and Analysis of Experiments*. The MacMillan Company, New York (1971).
- PETER, J. *Incomplete Blocks designs*. Lecture Notes in Statistics, Vol. I, Marcel Dekker, New York (1980).
- PRAT, A.; TORT-MARTORELL, X.; GRIMA, P. & POZUETA, L. *Métodos Estadísticos*. Edicions Univ. Politècnica Catalunya (1997).
- RATKOWSKI, D.A., EVANS, M.A. & ALLDREDGE, J.R. *Cross-over experiments. Design, analysis and application*. Statistics Textbooks and Monographs, 135, Marcel Dekker, New York (1993).