Tema 4

Diseño de experimentos

- Introducción. ¿Qué es el diseño experimental?
- Diseño completamente aleatorizado: análisis de la varianza con un solo factor.
- Diseño en bloques aleatorizados.
- Diseño factorial con dos factores.

Este tema está basado en el capítulo 6 de libro de Domingo de la bibliografía (reprografía), donde los conceptos IMPORTANTES aparecen remarcados (subrayados).

4.1. Introducción. ¿Qué es el diseño experimental?

Este apartado se corresponde con el punto 6.1 del libro. Se introducen algunos conceptos básicos en experimentación: qué es y cuál es el objetivo del diseño estadístico de experimentos, qué son los factores, los niveles o tratamientos.

Un **experimento** es un conjunto de pruebas o medidas, cuyo objetivo es obtener información que permita tomar decisiones sobre el producto o proceso bajo estudio.

Los experimentos diseñados estadísticamente permiten eficiencia y economía en el proceso experimental, y el empleo de los métodos estadísticos para el análisis de datos brinda **objetividad científica** al obtener conclusiones.

Los **factores controlados** en un experimento son las características para las que se prueban diferentes **niveles** o valores con el fin de ver su influencia sobre los resultados. Puede tratarse de factores cuantitativos (temperatura, velocidad, etc.) o cualitativos (proveedor, tipo de disolvente, etc.). Los diversos valores de un factor se llaman niveles del factor. En caso de controlar un único factor, sus niveles también se llaman **tratamientos**.

Además, decir que el diseño experimental utiliza tres principios básicos: obtención de réplicas, aleatorización y análisis por bloques. Veamos por orden el significado de estos tres principios, usando el siguiente ejemplo.

✍ **Ejemplo 4.1.:** Se trata de comparar unos métodos de evaluación.

Repetición: se asignan los mismos tratamientos a las diferentes unidades experimentales (elementos de un diseño). Sin la repetición, resulta imposible establecer la variabilidad natural y el error de medida.

Aleatorización: ¡Es el paso primordial de todas las estadísticas! Los tratamientos deben ser asignados de forma aleatoria a las unidades experimentales.

Control local: hace referencia a cualquier método que represente y reduzca la variabilidad natural. Una de sus formas es la agrupación de las unidades experimentales en bloques (cuando tratamos muestras apareadas ya estábamos utilizando este principio).

Asimismo, se muestran algunas directrices para el diseño experimental: comprensión y planteamiento del problema, elección de factores y niveles, selección de la variable respuesta, elección del diseño experimental, realización del experimento, análisis de datos, conclusiones y recomendaciones.

4.2. Diseño completamente aleatorizado: análisis de la varianza con un solo factor.

Se corresponde con el punto 6.2.

Ejemplo 4.2.: [Desempeño de robots] La capacidad de un robot para completar una tarea no depende sólo de la complejidad de la tarea, sino también de la complejidad del entorno en el que se ha desarrollar dicha tarea. Además, si la tarea supone localizar un objeto en el entorno, antes de llevar a cabo algún otro procedimiento secundario, la posición inicial del robot puede hacer variar el tiempo de finalización de la tarea. Supongamos que un ingeniero desea probar las capacidades de tres robots, numerados 1, 2 y 3, para localizar una caja con diversas herramientas en una habitación en la que también hay diversos objetos, y después tomar una herramienta determinada de la caja. Cada robot se ha probado 5 veces, seleccionando las posiciones iniciales aleatoriamente. Los 15 tiempos de finalización de la tarea (en segundos) se encuentran en la siguiente tabla.

Robot	Observaciones				
1	96	79	90	112	86
2	69	77	109	88	100
3	27	41	52	71	58

En un diseño experimental completamente aleatorizado, describiremos las observaciones con el modelo estadístico lineal (se considera que el factor de interés tiene a niveles y que inicialmente el número de observaciones, n, es igual para cada tratamiento):

$$Y_{ij} = \mu_i + \epsilon_{ij} \quad i = 1, \ldots, a \quad j = 1, ..., n,$$

donde Y_{ij} es una variable aleatoria que denota la observación ij-ésima, μ_i sería la media del tratamiento i y para los errores $\{\epsilon_{ij}\}$ haremos las siguientes suposiciones:

- Tienen esperanza nula.

- Su varianza es siempre constante, σ^2.
- Tienen una distribución normal.
- Son independientes entre sí.

Una formulación alternativa de estas hipótesis es:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2),$$

con τ_i definida como desviaciones de la media global μ, por lo que $\sum_{i=1}^{a} \tau_i = 0$.

La hipótesis de interés en este tipo de problemas es la de que *no hay diferencias significativas entre los niveles del factor*, que queda formalmente expresada por $H_0 : \mu_1 = \ldots = \mu_a$ o equivalentemente $H_0 : \tau_i = 0 \; \forall i$.

La idea para deducir el estadístico de contraste se basa en descomponer la *variabilidad total* de los datos en dos términos: la *variabilidad entre* las medias de cada muestra y la media general, y la *variabilidad dentro* de cada grupo o *residual*:

$$\begin{array}{ccccc} \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2 & = & n \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2 & + & \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 \\ SC_T & = & SC_{Tratamientos} & + & SC_E \end{array}$$

En esta identidad, y_{ij} representa la observación j-ésima obtenida en el tratamiento i-ésimo, $\bar{y}_{..}$ es la media de todas las observaciones, $\bar{y}_{i.}$ es la media de las observaciones bajo el tratamiento i-ésimo. Además, SC_T denota la suma de cuadrados total, $SC_{Tratamientos}$ la suma de cuadrados de los tratamientos (*variabilidad entre*) y SC_E la suma de cuadrados del error (*variabilidad dentro*).

La $SC_{Tratamientos}$ mediría la variabilidad explicada por las diferencias entre las medias de tratamientos, mientras que SC_E mediría la variabilidad no explicada. Cuando haya diferencias reales entre las medias en cada nivel, la variabilidad entre será grande comparada con la variabilidad residual. Juzgar su tamaño relativo requiere conocer su distribución en el muestreo.

Se demuestra que cuando H_0 es cierta SC_E/σ^2 y $SC_{Tratamientos}/\sigma^2$ son variables independientes que tienen una distribución ji-cuadrado con $a(n-1)$ y $a-1$ grados de libertad respectivamente.

El estadístico de contraste será pues:

$$F_0 = \frac{SC_{Tratamientos}/(a-1)}{SC_E/a(n-1)}$$

que tendrá una distribución F con $a-1$ y $a(n-1)$ grados de libertad. El numerador es conocido como cuadrado medio de los tratamientos ($CM_{Tratamientos}$) y el denominador como cuadrado medio de error (CM_E). Puede demostrarse también, que CM_E es un estimador insesgado de σ^2. Por otra parte, si la hipótesis nula es cierta, $CM_{Tratamientos}$ es un estimador insesgado de σ^2. Sin embargo, si la hipótesis nula es falsa, el valor esperado de $CM_{Tratamientos}$ es mayor que σ^2. Por consiguiente, H_0 se rechazará al nivel α si $f_0 > F_{\alpha,a-1,a(n-1)}$, es decir, tendremos un contraste unilateral en el que la región crítica es la cola derecha de la distribución F.

Los términos de la descomposición en que se basa este contraste suelen disponerse en una tabla conocida como *tabla ANOVA* (según la arraigada terminología anglosajona: ANalysis Of VAriance). STOP En la tabla 4.1 podemos ver la tabla ANOVA para este primer modelo. En esta tabla aparece el análisis de varianza cuando contamos con un diseño desbalanceado o desequilibrado (el número de observaciones en cada tratamiento puede ser diferente), sólo deben realizarse ligeras modificaciones en las fórmulas anteriores de las sumas de cuadrados. Elegir un diseño balanceado tiene dos ventajas: 1) si los tamaños son iguales, el procedimiento es relativamente insensible a las pequeñas desviaciones del supuesto de la igualdad de varianzas y 2) la potencia de la prueba se maximiza si las muestras tienen igual tamaño. Denotaremos por n_i las observaciones en el tratamiento i-ésimo y N el total de observaciones.

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	F
Tratamientos (entre grupos)	$\sum_{i=1}^{a} n_i(\bar{y}_{i.} - \bar{y}_{..})^2$	$a-1$	$SC_{Tratamientos}/(a-1)$	$CM_{Tratamientos}/CM_E$
Error (dentro grupos)	$\sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$	$N-a$	$SC_E/(N-a)$	
Total	$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$	$N-1$		

Tabla 4.1: Tabla ANOVA de un factor.

✍ **Ejemplo 4.2.:** Especifica el modelo y prueba la igualdad de los efectos.

```
ANOVA Table for Robot by Crobot

                            Analysis of Variance
-----------------------------------------------------------------------------
Source      Sum of Squares     Df  Mean Square F-Ratio P-Value
-----------------------------------------------------------------------------
Between groups    5588,8      2       2794,4    11,95 0,0014
Within groups     2807,2     12      233,933
-----------------------------------------------------------------------------
Total (Corr.)       8396,0    14
```

Los estimadores para los parámetros del modelo que después usaremos en prácticas para construir los residuos y comprobar la validez del modelo:

$\hat{\mu} = \bar{y}_{..} = \frac{\sum_i \sum_j y_{ij}}{N}$

$\hat{\mu}_i = \bar{y}_{i.} = \frac{\sum_j y_{ij}}{n_i}$

$\hat{\tau}_i = \hat{\mu}_i - \hat{\mu}$, donde se usa $\sum_{i=1}^{a} n_i\tau_i = 0$.

Los residuos ($e_{ij} = y_{ij} - \hat{y_{ij}} = y_{ij} - \bar{y}_{i.}$) son útiles para verificar las hipótesis básicas del modelo: comprobar su normalidad, representar los residuos frente a los valores ajustados o frente a los niveles del factor, la homogeneidad de varianzas también puede comprobarse mediante algún test como el de Bartlett, representar los residuos frente al tiempo, comprobar si existen valores atípicos. En prácticas se trabajará este punto.

Seguidamente, se verán algunos métodos para comparar las medias, cuando el efecto ha sido declarado significativo por la prueba F.

Podrían dividirse en *comparaciones a priori* (nosotros no las veremos en este curso), es decir, antes de llevar a cabo el experimento, ya se saben las comparaciones de interés: se considerarán los contrastes ortogonales. Un contraste lo podemos plantear como $H_0 : \sum_i c_i\mu_i = 0$ con $c_1, \ldots, c_a$ constantes conocidas con $\sum_i n_i c_i = 0$. Dos contrastes con coeficientes $\{c_i\}$ y $\{d_i\}$ serán ortogonales si $\sum_i n_i c_i d_i = 0$. Para probar un contraste se debe comparar su suma de cuadrados $SC_C = (\sum_{i=1}^a c_i y_{i.})^2 / \sum_{i=1}^a n_i c_i$ ($y_{i.}$ es el total de las observaciones del tratamiento i-ésimo) con la media de cuadrados del error. El estadístico resultantes tiene una distribución F con 1 y $N-a$ grados de libertad.

Por otro lado, si no se tiene planteada ninguna pregunta con respecto a las medias de los tratamientos (*comparaciones a posteriori*), se presentará el método de la mínima diferencia significativa o LSD (*Least Significant Difference*). De esta manera, se compararán todos los pares de medias con las hipótesis nulas $H_0 : \mu_i = \mu_j$ (para toda $i \neq j$) y el par de medias μ_i y μ_j se declarará significativamente diferente si $|\bar{y}_{i.} - \bar{y}_{j.}| >$ LSD, donde LSD al nivel α viene definida como $t_{\alpha/2,N-a}\sqrt{CM_E(1/n_i + 1/n_j)}$.

✍ **Ejemplo 4.2.:** Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados.

A continuación, se presenta una alternativa no paramétrica de la prueba F, el test de Kruskal-Wallis, que sólo veremos en prácticas. Únicamente se requiere que las ϵ_{ij} tengan la misma distribución continua para todos los niveles del factor. Esta prueba se basa en los rangos (orden) de las observaciones y el estadístico de la prueba es:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{a} \frac{R_{i.}^2}{n_i} - 3(N+1),$$

donde $R_{i.}$ es el total de los rangos del tratamiento i-ésimo. Se rechazará H_0 si el valor observado $h \geq \chi_{\alpha,a-1}$, con un nivel de significación aproximado α.

4.3. Diseño en bloques aleatorizados.

Este punto se corresponde con el punto 6.3. del Domingo. El siguiente modelo que se presenta es el diseño en bloques aleatorizados. En el anterior modelo los factores no controlados por el experimentador y que podían influir en los resultados se asignaban al azar a las observaciones. En este modelo, las unidades experimentales han sido agrupadas según otra causa de variabilidad que puede influir en los resultados: es una variable, denominada *variable de bloqueo*, cuyo efecto sobre la respuesta no es directamente de interés, pero de esta manera se obtienen comparaciones homogéneas, de forma análoga al procedimiento de la prueba t apareada. En este diseño, se tomarán el mismo número de muestras por tratamiento dentro de cada bloque y el orden de las medidas dentro del bloque se decidirá aleatoriamente.

Las hipótesis básicas del modelo (sin repetición) son:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad i = 1, \ldots, a \quad j = 1, \ldots b,$$

donde ϵ_{ij} son variables $N(0, \sigma^2)$ independientes.

El modelo descompone la respuesta en:

- Una media global μ.
- El efecto incremental en la media debida al nivel del factor, τ_i ($\sum_i \tau_i = 0$).
- El efecto incremental en la media debida al bloque, β_j ($\sum_j \beta_j = 0$).
- El error experimental, ϵ_{ij}, que recoge el efecto de todas las restantes causas posibles de variabilidad del experimento.

Notar que este modelo supone que los efectos del factor y de la variable de bloqueo son aditivos, es decir, no existe interacción entre ambos.

Al igual que antes, resulta interesante ilustrar este diseño con un ejemplo.

Ejemplo 4.3.: [Eficiencia de algoritmos] Muchas veces, un procedimiento numérico puede realizarse mediante diversos algoritmos. Se considera comparar la eficiencia de 3 algoritmos, escritos en un lenguaje de alto nivel, como Pascal. Cada uno se ejecuta en 4 máquinas distintas. Como los tiempos de ejecución se ven afectados por la elección del hardware, en el experimento se consideró como bloques, las cuatro máquinas. Los datos (en segundos) aparecen en la tabla siguiente.

	Máquinas			
Algoritmo	1	2	3	4
1	6.42	10.51	4	6.2
2	6.85	11.45	5.12	6.72
3	5.6	9.5	3.94	5.66

La hipótesis de interés será $H_0 : \tau_i = 0\ \forall i$ pero en algunas ocasiones también puede ser de interés contrastar $H_0 : \beta_j = 0\ \forall j$.

La deducción de los estadísticos de contraste y sus distribuciones muestrales se obtiene de una manera completamente análoga al anterior. Así, la *variabilidad total* puede descomponerse en *variabilidad entre tratamientos*, *variabilidad entre bloques* y el *error* o *variabilidad dentro de los tratamientos y bloques.*

$$\begin{array}{ccccccc} \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2 & = & b\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 & + & a\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 & + & \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j}-\bar{y}_{i.}+\bar{y}_{..})^2 \\ SC_T & = & SC_{Tratamientos} & + & SS_{Bloques} & + & SC_E \end{array}$$

Las fórmulas para el caso de un diseño por bloques aleatorizados con repetición pueden consultarse en el Domingo, pero no se tratarán.

Puede demostrarse que: $\mathrm{E}(CM_{Tratamientos}) = \mathrm{E}(SC_{Tratamientos}/a-1) = \sigma^2 + b\sum_{i=1}^{a}\tau_i^2/a-1$, $\mathrm{E}(CM_{Bloques}) = \mathrm{E}(SC_{Bloques}/b-1) = \sigma^2 + a\sum_{j=1}^{b}\beta_j^2/b-1$ y $\mathrm{E}(CM_E) = \mathrm{E}(SC_E/(a-1)(b-1))$ $= \sigma^2$. Por tanto, la hipótesis nula de que todos los efectos de los tratamientos son cero, se rechazará con el nivel de significación α si el valor calculado del estadístico $F_0 = CM_{Tratamientos}/CM_E$ $> F_{\alpha,a-1,(a-1)(b-1)}$. En la tabla 4.2 podemos ver la tabla de ANOVA para este modelo.

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	F
Tratamientos	$SC_{Tratamientos}$	$a-1$	$SC_{Tratamientos}/(a-1)$	$CM_{Tratamientos}/CM_E$
Bloques	$SC_{Bloques}$	$b-1$	$SC_{Bloques}/(b-1)$	
Error	SC_E	$(a-1)(b-1)$	$SC_E/(a-1)(b-1)$	
Total	SC_{Ts}	$ab-1$		

Tabla 4.2: Tabla ANOVA de un factor diseño en bloques aleatorizados.

✍ **Ejemplo 4.3.:** Especifica el modelo y prueba la igualdad de los efectos.

```
 Analysis of Variance for Eficiencia - Type III Sums of Squares
--------------------------------------------------------------------------------
Source          Sum of Squares     Df     Mean Square F-Ratio P-Value
--------------------------------------------------------------------------------
MAIN EFFECTS
 A:Algoritmo          3,71322       2         1,85661  25,27     0,0012
 B:Computadora        60,6061       3          20,202  274,99    0,0000

RESIDUAL            0,440783        6       0,0734639
--------------------------------------------------------------------------------
TOTAL (CORRECTED)   64,7601        11
--------------------------------------------------------------------------------
All F-ratios are based on the residual mean square error.
```

Es interesante también observar los resultados que se hubiesen obtenido si no se hubiesen considerado los bloques, comparar el diseño en bloques con el diseño aleatorizado para mostrar los beneficios de analizar por bloques, poniendo por caso el ejemplo 4.3.

Ejemplo 4.3.: Supongamos ahora que no se han considerado las 4 máquinas distintas como bloques, entonces:

```
ANOVA Table for Eficiencia by Algoritmo

                            Analysis of Variance
-----------------------------------------------------------------------------
Source       Sum of Squares   Df  Mean Square    F-Ratio P-Value
-----------------------------------------------------------------------------
Between groups    3,71322      2      1,85661      0,27   0,7667
Within groups     61,0469      9      6,78299
-----------------------------------------------------------------------------
Total (Corr.)     64,7601     11
```

De igual forma que antes, es importante examinar los residuos (lo haremos en prácticas). Ahora los valores valores ajustados o estimados de y_{ij} serían: $\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) = \bar{y}_{i.} + \bar{y}_{j.} - \bar{y}_{..}$. Una gráfica de los residuos frente a los valores ajustados con forma de curva, podría sugerir una interacción entre los tratamientos y los bloques.

También, siguiendo el esquema anterior, si el análisis de varianzas hubiera indicado la existencia de diferencias entre las medias de los tratamientos, podremos utilizar el método LSD, ahora calculada como LSD $= t_{\alpha/2,(a-1)(b-1)}\sqrt{2CM_E/b}$.

✍ **Ejemplo 4.3.:**

```
Multiple Range Tests for Eficiencia by Algoritmo

--------------------------------------------------------------------------------
Method: 95,0 percent LSD

 Algoritmo  Count  LS Mean          Homogeneous Groups
--------------------------------------------------------------------------------
3             4        6,175             X

1             4        6,7825             X

2             4        7,535                X
--------------------------------------------------------------------------------
Contrast                       Difference          +/- Limits
--------------------------------------------------------------------------------
1 - 2                         *-0,7525            0,468966

1 - 3                         *0,6075             0,468966
```

```
2 - 3                           *1,36                    0,468966
--------------------------------------------------------------------------------
* denotes a statistically significant difference.
```

4.4. Diseño factorial con dos factores.

Este punto no se encuentra en el libro de domingo y sólo se trabajará en prácticas.

El último modelo de ANOVA que se estudiará es el de dos factores, suponemos ahora que la observación de la variable Y está influida por dos factores de interés.

La idea básica de los diseños factoriales es cruzar los niveles de los factores a todas las combinaciones posibles. Resultará conveniente resaltar las ventajas de estos diseños frente a experimentos en los que se varía un factor, dejando constantes los demás.

Ejemplo 4.4.: Efecto medicamento y alcohol.

Ejemplo 4.5.: [Algoritmos de reconocimiento y filtros] Una ingeniera desea determinar si existe o no diferencia significativa entre los efectos de dos algoritmos de reconocimiento de objetivos, A y B. Antes de la aplicación de A o B, la imagen debe filtrarse para reducir el ruido. La ingeniera puede elegir entre el filtro F o G. La tabla siguiente da las clasificaciones del sistema de reconocimiento para cada combinación algoritmo-filtro.

	Filtro	
Algoritmo	1	2
A	72	63
	68	59
	74	57
B	53	52
	55	56
	49	57

Para este modelo consideraremos que tomamos una muestra de tamaño n en cada combinación de factores, uno con a niveles (factor A) y el otro con b (factor B), y las abn observaciones se realizan en orden aleatorio.

Las hipótesis básicas del modelo son ahora:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, a \quad j = 1, \ldots b \quad k = 1, \ldots, n,$$

donde las variables ϵ_{ijk} son $N(0, \sigma^2)$ independientes.

El modelo descompone la respuesta en:

- Una media global μ.
- El efecto incremental en la media debida al nivel i-ésimo del factor A, τ_i ($\sum_i \tau_i = 0$).
- El efecto incremental en la media debido al nivel j-ésimo del factor B, β_j ($\sum_j \beta_j = 0$).
- $(\tau\beta)_{ij}$ representa el efecto de la interacción entre ambos factores, así que $\sum_i (\tau\beta)_{ij} = 0$ y $\sum_j (\tau\beta)_{ij} = 0$.
- El error experimental, ϵ_{ijk}, que recoge el efecto de todas la restantes causas posibles de variabilidad del experimento.

Tres son las hipótesis de interés: $H_0 : \tau_i = 0\ \forall i$, $H_0 : \beta_j = 0\ \forall j$ y $H_0 : (\tau\beta)_{ij} = 0\ \forall i, j$.

Análogamente, el análisis de varianza prueba estas hipótesis descomponiendo la variabilidad total de los datos en partes componentes y comparando después los diferentes elementos de esta descomposición. En la tabla 4.3 podemos ver la tabla de ANOVA para este modelo.

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	F
Factor A	$bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$	$a-1$	$SC_A/(a-1)$	CM_A/CM_E
Factor B	$an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$	$b-1$	$SC_B/(b-1)$	CM_B/CM_E
Interacción	$n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$	$(a-1)(b-1)$	$SC_{AB}/(a-1)(b-1)$	CM_{AB}/CM_E
Error	$\sum_i \sum_l \sum_k (y_{ijk} - \bar{y}_{ij.})^2$	$ab(n-1)$	$SC_E/ab(n-1)$	
Total	$\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2$	$abn-1$		

Tabla 4.3: Tabla ANOVA de dos factores con interacción.

✍ **Ejemplo 4.5.:** Tabla ANOVA.

```
Analysis of Variance for Clasficacion - Type III Sums of Squares
--------------------------------------------------------------------------------
Source    Sum of Squares     Df    Mean Square F-Ratio P-Value
--------------------------------------------------------------------------------
MAIN EFFECTS
 A:Algtarget    420,083      1    420,083      48,01     0,0001
 B:Filtro        60,75       1     60,75       6,94      0,0299

INTERACTIONS
 AB             154,083      1    154,083      17,61     0,0030

RESIDUAL         70,0        8      8,75
--------------------------------------------------------------------------------
TOTAL (CORRECTED)704,917     11
--------------------------------------------------------------------------------
All F-ratios are based on the residual mean square error.
```

Nuevamente, se estudiarán los residuos: $e_{ijk} = y_{ijk}$ - $\bar{y}_{ij.}$ y también pruebas para medias individuales, usando el método de la LSD. Si la interacción es significativa, podría aplicarse a las medias de un factor, fijando el otro a un nivel particular.

Problemas del tema 4

Los datos de los ejercicios se encuentran en el fichero probt4.txt que puedes bajar de mi página web (también están en papel en reprografía). Para los siguientes ejercicios (del 1 al 6), se deben realizar las siguientes tareas (es posible ayudarse del programa Statgraphics): especificar el modelo considerado, construir la tabla ANOVA según el modelo pertinente, probar la igualdad de los efectos o la interacción (según el caso) con $\alpha = 0.05$. Utiliza el método de la LSD si la hipótesis nula correspondiente es rechazada con ANOVA de una vía (usa el Statgraphics para más de un factor), comentando los resultados.

Las soluciones están en reprografía.

1. Cuando se compila un lenguaje de alto nivel, la eficiencia del tiempo de ejecución depende del compilador. Desean compararse tres compiladores que se han desarrollado para un nuevo lenguaje de alto nivel para una determinada máquina. Se seleccionan aleatoriamente 15 programas de aquellos considerados apropiados para el nuevo lenguaje, cada uno escrito por el mismo programador en el nuevo lenguaje y 5 son asignados aleatoriamente a cada compilador. Los datos se encuentran en las variables: TiempoCPU (para las observaciones) y Compilador (para identificar cada compilador).

2. Tres marcas de pilas (1.5 V nominales) quieren evaluarse, seleccionando aleatoriamente 4 de cada marca. Los datos se encuentran en las variables: Voltaje (para las observaciones) y Marca (para las tres marcas).

3. Se desea evaluar la capacidad de tres marcas de baterías para mantener sus estados de carga durante el tiempo (una lectura de 1260 correspondería con carga completa). Tres baterías de cada marca se probará en tres vehículos distintos que actuarán como bloques, pues el desgaste depende del tipo de vehículo. Los datos aparecen en las variables: Lectura (para las observaciones), Batería (para las tres marcas de baterías) y Vehículo (para indicar cada vehículo).

4. Un pequeño motor eléctrico puede ensamblarse de tres formas distintas por un individuo. Para decidir cuál es más eficiente, se considerarán 5 observaciones para cada técnica de ensamblado. Como el montaje depende de las habilidades de los operadores, 5 operadores se tratarán como bloques. Para cada operador, tres motores sin ensamblar se asignan aleatoriamente a las técnicas. De igual forma, el orden de montaje se determina aleatoriamente. Los datos obtenidos se encuentran en las variables: Tiempoens (minutos de ensamblado), Técnica (para cada técnica) y Operador (para indicar cada operador).

5. Un diseñador de sistemas operativos desea comparar los efectos de cuatro disciplinas de colas para trabajos en espera. Éstas dependen de la intensidad computacional y los requer-

imientos de Entrada/Salida, por ello el diseñador decide bloquear el experimento observando cada disciplina en tres entornos de programación distintos, que servirán de bloques. Los datos se recogen en las variables: Longitudcola (para la media de longitudes de cola observadas), Disciplina (para cada disciplina) y Entorno (para cada entorno).

6. Se realizó un estudio de simulación para investigar el desempeño de máquina de varios algoritmos nuevos para funciones de la biblioteca de programas de computadora en FORTRAN (*IBM Journal of Research and Development, marzo 1986*). Se calculó el tiempo por llamada (en microsegundos) para varias funciones escalares (promediando para 10000 argumentos aleatorios) en cada una de tres máquinas IBM System/370 distintas. Se trataron las funciones como bloques. Los datos se encuentran en las variables: TiempoIBM (para los tiempos observados), IBM (para las tres máquinas IBM) y Función (para cada función).

Autoevaluación del tema 4

1. Se realizó un estudio encaminado a evaluar la valía del reconocimiento del habla en interacciones entre personas y sistemas de computadora (*Special Interest on Computer-Human Interaction Bulletin, julio 1993*). Una muestra de 45 sujetos se dividió aleatoriamente en tres grupos, y se les pidió a cada sujeto realizar tareas con un sistema básico de correo de voz. Se utilizó un interfaz diferente en cada grupo: (1) digitación de tonos, (2) operador humano o (3) reconocimiento simulado del habla. Una de las variables medidas fue el tiempo global (en segundos) para realizar las tareas asignadas. Se realizó un análisis para comparar los tiempos de desempeño global medios de los tres grupos.

   a) ¿Cuál es el diseño de experimento empleado en el estudio? ¿Cuál es la hipótesis nula a probar?

   b) Las medias observadas de los tiempos de desempeño para cada grupo fueron: (1) 1.4, (2) 1.03 y (3) 1.04. A pesar de las diferencias entre las medias anteriores, no fue posible rechazar la hipótesis nula del apartado anterior con $\alpha = 0.05$, ¿cómo es posible esto?

2. Se efectúa un experimento para comparar la acción de limpieza de dos detergentes, el A y el B. Veinte muestras de tela se ensucian con mugre y grasa, cada una se lava con uno de los detergentes en una lavadora y luego se mide la "blancura" en ellas. Critica los siguientes aspectos del experimento:

   a) El detergente A se usó con agua blanda, y el B con agua dura.

   b) Quince muestras se lavaron con el detergente A y 5 con el B.

   c) Las lecturas de "blancura" de todas las muestras lavadas con el detergente A se tomaron primero.

3. Un asunto que interesa mucho a los diseñadores y usuarios de monitores de computadora, es la facilidad de lectura del texto que aparece en la pantalla. Se cree que dos factores que afectan la facilidad de lectura de los textos que se desplazan verticalmente es el tamaño de la ventana (es decir, el número de caracteres exhibidos por línea) y la longitud del salto (es decir, el número de caracteres que se avanza en cada salto). A fin de investigar este fenómeno, se realizó un experimento factorial 2 x 3, con el tamaño de ventana en dos niveles (20 y 40 caracteres) y la longitud del salto en tres niveles (1, 5 y 9 caracteres). La variable de respuesta de interés, fue la velocidad de lectura (medida en palabras por minuto) de estudiantes universitarios chinos que participaron en el estudio (*Human Factors, junio de 1988*).

a) La prueba F de ANOVA para la interacción entre los dos factores arrojó un p-valor mayor que 0.1. Interpreta este resultado.

b) La prueba F de ANOVA para el efecto principal del tamaño de ventana produjo un p-valor mayor que 0.1. Interpreta este resultado.

c) La prueba F de ANOVA para el efecto principal de la longitud del salto arrojó un p-valor menor que 0.05. Interpreta este resultado.

4. Se ha de seleccionar un componente electrónico para utilizarlo en la fabricación de un equipo informático. Se dispone de muestras de cuatro proveedores. Se ha evaluado la fiabilidad de cada componente de cada muestra y con los datos (de cada proveedor tenemos 5 componentes, excepto del cuarto proveedor que contamos con 4 componentes), se ha obtenido la tabla ANOVA. Sin embargo, hay algunos huecos sin rellenar, completalos. Especifica el modelo y realiza la prueba (usa α=0.05).

```
ANOVA Table for Fiabilidad by Proveedor

                            Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio

-----------------------------------------------------------------------------
Between groups
 Within groups                            0,00716667
-----------------------------------------------------------------------------
Total (Corr.)            0,587368
```