

19

Análisis de Anova encajado

En este capítulo se tratan los modelos de ANOVA de un factor que tienen más de un nivel de análisis. No se trata de dos factores por separado, sino que dentro del factor principal de análisis, se consideran otros “factores” subordinados. Para evitar la confusión con los términos, se prefiere decir que hay dentro del factor un *nivel* principal y dentro de éste, uno o más subniveles de análisis. Es como si estos últimos estuviesen encajados en el primero. Esto permite un poco más de sofisticación en el estudio de las variabilidades que afectan a los experimentos. En la primera parte se explican los diferentes modelos de ANOVA encajado, de acuerdo a como son los modelos empleados en cada nivel. Luego se presentan las formas cortas de cálculo y ejemplos de aplicación.

19.1 Introducción

El modelo de Anova de un factor visto en los capítulos precedentes, es con mucha frecuencia insuficiente para abarcar toda la complejidad de un experimento dado y para extraer del mismo toda la información importante. En el presente capítulo se verá el caso en que cada grupo de muestras está subdividida en clases escogidas aleatoriamente. Por ejemplo, en un experimento donde se estudiarán 5 caldos de cultivo, se pueden preparar 3 cajas de Petri para cada caldo, y en cada una de estas, incubar la muestra extraída a un paciente. Así, los caldos de cultivo serán los grupos principales y las cajas los subgrupos dentro de estos. La ventaja respecto al Anova de un factor, es que ahora se pueden observar las diferencias entre los 3 tipos de siembra para estar seguros que las diferencias observadas entre las cajas se deben a la diferente composición de los caldos. Podría darse que diferencias morfológicas o ambientales y otras accidentales, entre los diferentes microbios cultivados en las diferentes cajas, ocasionen diferencias entre el resultado de los cultivos al final de la experiencia. La única manera de separar estos dos efectos es tener dos o más cajas de Petri para cada caldo de cultivo, y con el Anova, separar las variabilidades de estos efectos. Si no se encuentran diferencias significativas entre las cajas, se pueden atribuir las diferencias observadas a los distintos tipos de caldos de cultivo. Todavía, si se encuentran diferencias significativas entre las cajas, con la componente añadida de varianza entre las cajas, se puede comparar la variación entre los caldos para ver si son significativas, sobre la base de la variabilidad entre cajas tratadas de la misma manera.

A este tipo de análisis, se lo conoce como *Anova encajado o jerárquico*, pues una clasificación subordinada está encajada dentro del nivel más alto de clasificación. En el nivel subordinado los grupos siempre deben ser escogidos en forma aleatoria. Por lo tanto, siempre se tratará

de un Modelo tipo II de Anova en los subniveles. En cambio, en el nivel principal los grupos se pueden escoger en forma aleatoria (*Anova encajado puro*), o bien, en forma especial como en un Modelo tipo I de Anova (*Anova encajado mixto*). Además, se pueden seguir subdividiendo las clases en otros subgrupos y tener así más de dos niveles. Por ejemplo, para el estudio de una droga aplicada a seres humanos se pueden preparar 5 dosis y un placebo (clásico Modelo I) para el primer nivel. Dentro de cada grupo se pueden escoger 10 pacientes en forma aleatoria para conformar el segundo nivel de análisis. De cada paciente se pueden tomar tres muestras para evaluar la variabilidad de lecturas repetidas y conformar así el tercer nivel subordinado, para cuantificar la parte de la varianza debida al “error” de medición. Pero en el segundo nivel se puede determinar la variabilidad entre pacientes y en el primero la debida a las dosis. Otros usos importantes del modelo puro se ven frecuentemente en Genética cuantitativa, Sistemática, etc.

19.2 Modelos teóricos y formas de cálculo

Los datos medidos en un Anova encajado, pueden explicarse en forma similar a la vista como Modelos I y II del Anova de un factor. Con las diferencias siguientes:

$$X_{ijk} = \mu + A_i + B_{ij} + \varepsilon_{ijk} \quad \text{Modelo de Anova encajado puro}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i
 μ : es la media Paramétrica de la población
 A_i : es la contribución aleatoria del efecto del grupo i
 B_{ij} : es la contribución aleatoria del efecto del subgrupo j, del grupo i
 ε_{ijk} : es el error de la medición k, en subgrupo j, del grupo i

Se supone que A_i se distribuye en forma gaussiana con media 0 y varianza σ_i^2 ; mientras que B_{ij} también se distribuye normalmente con media 0 y varianza σ_{BcA}^2 , donde se indica que la varianza es del subgrupo B dentro del grupo A. Por su parte ε_{ijk} tiene media 0 y varianza σ^2

Mientras que para el modelo mixto la relación anterior se transforma en :

$$X_{ijk} = \mu + \alpha_i + B_{ij} + \varepsilon_{ijk} \quad \text{Modelo de Anova encajado mixto}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i
 μ : es la media Paramétrica de la población
 α_i : es la contribución fija del efecto del grupo i
 B_{ij} : es la contribución aleatoria del efecto del subgrupo j, del grupo i
 ε_{ijk} : es el error de la medición k, en subgrupo j, del grupo i

Esta fórmula es la misma que la vista más arriba, solo que el efecto del factor principal es fijo en lugar de aleatorio, esto es que α_i es una constante para todo los datos del grupo i.

Para presentar las formas de cálculo en este modelo de Anova, se muestra el siguiente:

Ejemplo 1) Una empresa farmacéutica tiene sucursales en todo el país. Luego de lanzar una nueva línea de productos de cosmética decide analizar la variabilidad de sus ventas. Para ello escoge

3 sucursales al azar y en cada una de ellas se sortearon 4 días cualquiera del mes. Como se trabaja 24 horas en 3 turnos, para separar la varianza error se sortearon 2 turnos en cada caso. Su propósito al tomar 4 días al azar, diferentes en cada sucursal, es ver si las diferencias entre sucursales, añaden varianza a la variabilidad total de las ventas debidas a razones estacionales y de mercado. La razón es mejorar la planificación general de las compras para no quedarse con un stock excesivo de mercadería. Como las sucursales (nivel 1) se escogieron al azar, lo mismo que los días del mes (subnivel 2) y los turnos (mediciones repetidas dentro del subnivel 2). Se trata claramente de un *Modelo de Anova encajado puro*. Muy diferente hubiera sido si las sucursales las elige a propósito para ver si hay diferencia significativa entre ellas; en tal caso sería un modelo mixto. Los datos obtenidos son el monto total de venta diario, expresados en pesos, de la nueva línea de cosmética. Notar que se podría tomar las ventas mensuales de cada sucursal durante un semestre y diseñar un modelo simple de Anova. Pero el diseño escogido no necesita esperar varios meses para trabajar los datos, sino que tomando una sola quincena ya es suficiente para comenzar a seguir el proceso de venta de la nueva línea lanzada.

Cuadro 19.1: Ventas diarias en \$ de una nueva línea cosmética en 3 sucursales

	Suc. 1				Suc. 2				Suc. 3			
	1	2	3	4	1	2	3	4	1	2	3	4
1	175,5	233,4	252	210,3	209,4	168	152,1	191,4	169,8	233,1	209,7	187,2
2	178,5	242,7	250,8	204,9	209,4	163,5	147,9	197,4	172,5	237,0	207,6	193,5
T _{jk}	354	476,1	502,8	415,2	418,8	331,5	300	388,8	342,3	470,1	417,3	380,7
T _k	1748,1				1439,1				1610,4			

Paso 1) Se calculan los totales diarios T_{jk} de las mediciones repetidas (X_{ijk}) y el total por sucursal. Entonces hay N = a . b . n = 3 . 4 . 2 = 24 datos y es:

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^2 X_{ijkl} = \sum_{i=1}^a T_{ik} = 1748,1 + 1439,1 + 1610,4 = 4797,6 = T$$

Paso 2) Se calcula la suma del cuadrado de todos los datos:

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 = (175,5)^2 + (178,5)^2 + \dots + (193,5)^2 = 980.569,44 = T_x^2$$

Paso 3) Se calcula la suma de los cuadrados de los totales de los subgrupos, dividido por el tamaño muestral respectivo (en este caso todos los tamaños son iguales a 2):

$$\sum_{i=1}^a \sum_{j=1}^b T_{jk}^2 / n_{jk} = (1/n) \sum_{i=1}^a \sum_{j=1}^b T_{jk}^2 = (1/2)[(354)^2 + (476,1)^2 + \dots + (380,7)^2] = 980.436,15 = T_b^2$$

Paso 4) Se calcula la suma de los cuadrados de los totales de los grupos, dividido por su tamaño muestral respectivo (en este caso todos los tamaños son iguales a n.b = 8)

$$\sum_{i=1}^a T_k^2 / n_k = (1/n.b) \sum_{i=1}^a T_k^2 = (1/8)[(1748,1)^2 + (1439,1)^2 + (1610,4)^2] = 965.031,32 = T_a^2$$

Paso 5) Se calcula el término de corrección T² / N = TC = (4797,6)² / 24 = 959.040,24

Paso 6) Se calculan las sumas de cuadrados necesarias

$$SS_{\text{total}} = T_x^2 - TC = (\text{Paso 2} - \text{Paso 5}) = (980.569,44 - 959.040,24) = 21529,2$$

$$SS_{\text{grupos}} = T_a^2 - TC = (\text{Paso 4} - \text{Paso 5}) = (965.031,32 - 959.040,24) = 5991,08$$

$$SS_{\text{subgrupos}} = T_b^2 - T_a^2 = (\text{Paso 3} - \text{Paso 4}) = (980.436,15 - 965.031,32) = 15404,83$$

$$SS_{\text{dentro}} = T_x^2 - T_b^2 = (\text{Paso 2} - \text{Paso 3}) = (980.569,44 - 980.436,15) = 133,29$$

Paso 7) Se calculan el cuadro de Anova encajado como sigue:

TABLA DE ANOVA

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre grupos	5991,08	(a-1) = 2	MSe /a=(5991,08/2)=2995,54	MSe / MSeg = 1,75 (ns)
Entre subgrupos	15404,83	a(b-1) = 9	MSeg /a(b-1)=(15404,83/9)=1711,7	MSeg / MSd =154,2***
Dentro subgrupos	133,29	ab(n-1) = 12	MSd/ab(n-1)=(133,29/12) =11,1	
Total	21529,2	23		

Si en lugar del modelo encajado, se hubiera hecho el modelo unifactorial simple, la suma de cuadrados entre las sucursales (entre grupos) sería la misma $SS_{\text{entre suc.}} = 5991,08$. Y la suma de cuadrados error, dentro de los grupos, sería la diferencia entre la total y la que hay entre las sucursales, $SS_{\text{error}} = (21529,2 - 5991,08) = 15538,12$. Esto significa que, el estadígrafo $F = (5991,08 / 2) / (15538,12 / 21) = 4,05^*$ sería significativo.

Cabe preguntarse la razón por la cual, en este modelo hay diferencias significativas entre las sucursales y en el encajado no. La respuesta es que en el ANOVA simple la variabilidad remanente o error contiene la debida a la aleatoriedad de ventas diarias entre los turnos. Y al no hacer esa discriminación el análisis se torna confuso, muestra una significación disfrazada porque el cuadrado medio remanente o error ha cambiado. Ahora se puede discriminar mejor y se puede ver que la mayor parte de la variabilidad se debe a la que hay entre los turnos. A medida que el investigador subdivida en más niveles, mejor podrá discriminar las cosas. *La idea básica en diseño de ANOVA es tratar de reducir lo más posible el MS remanente debido al error.*

La riqueza del modelo encajado reside, en este caso, en que la suma de cuadrados remanente o error (dentro de las sucursales) puede descomponerse en dos partes: una al considerar los diferentes turnos (entre turnos, dentro de las sucursales) y la nueva remanente más chica. Hay menor error (indeterminación) y se pueden encontrar mejores explicaciones a las cosas. Se debe continuar con un análisis de la componente añadida de varianza:

Paso 8) Se calculan las componentes añadidas con:

La variabilidad error (dentro de subgrupos) es: $\sigma^2 = 11,1$

La variabilidad entre subgrupos, es decir entre turnos dentro de las sucursales es:

$$\sigma_{BcA}^2 = (MS_{\text{subgrupos}} - MS_{\text{error}}) / n = (1711,7 - 11,1) / 2 = 850,3$$

La variabilidad entre grupos (sucursales) es:

$$\sigma_A^2 = (MS_{\text{grupos}} - MS_{\text{subgrupos}}) = (2995,54 - 1711,7) / 8 = 160,48. \quad \text{Luego,}$$

$$\text{Variabilidad total} = \sigma^2 + \sigma_{BcA}^2 + \sigma_A^2 = 11,1 + 850,3 + 160,48 = 1021,88. \quad \text{Entonces:}$$

$$\sigma^2 \% \text{ representa el } (11,1 / 1021,88) \cdot 100 = 1,08\% \text{ de la variabilidad total (el error).}$$

$$\sigma_{BcA}^2 \% \text{ representa el } (850,3 / 1021,88) \cdot 100 = 83,21\% \text{ del total (turnos dentro de sucursales).}$$

$$\sigma_A^2 \% \text{ representa el } (160,48 / 1021,88) \cdot 100 = 15,71\% \text{ del total (entre las sucursales).}$$

19.3 ANOVA encajado de tres niveles

Para este modelo aparece un nuevo subnivel y entonces, otra vez, quedan dos tipos de planteos o diseños experimentales posibles: el Modelo puro y el mixto.

$$X_{ijkl} = \mu + A_i + B_{ij} + C_{ijk} + \varepsilon_{ijkl} \quad \text{Modelo de Anova encajado puro}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i

μ : es la media Paramétrica de la población

A_i : es la contribución aleatoria del efecto del grupo i

B_{ij} : es la contribución aleatoria del efecto del subgrupo j, del grupo i

C_{ijk} : es la contribución aleatoria del efecto del sub-subgrupo k, del subgrupo j en grupo i

ε_{ijkl} : es el error de la medición l, en sub-subgrupo k, en subgrupo j, del grupo i

Se supone que A_i se distribuye en forma gaussiana con media 0 y varianza σ_i^2 ; mientras que B_{ij} también se distribuye normalmente con media 0 y varianza σ_{BcA}^2 , donde se indica que la varianza es del subgrupo B dentro del grupo A. Además, C_{ijk} se distribuye normalmente con media 0 y varianza σ_{CcB}^2 (se indica el sub-subgrupo C dentro del subgrupo B). Y por su parte ε_{ijkl} tiene media 0 y varianza σ^2 y se trata de la varianza remanente o error.

Mientras que para el modelo mixto la relación anterior se transforma en :

$$X_{ijkl} = \mu + \alpha_i + B_{ij} + C_{ijk} + \varepsilon_{ijkl} \quad \text{Modelo de Anova encajado mixto}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i

μ : es la media Paramétrica de la población

α_i : es la contribución fija del efecto del grupo i

Y los demás términos son similares al visto modelo puro.

19.3.1 Formas cortas de cálculo

Para ilustrar este modelo se irá desarrollando la forma de cálculo con el modo paso a paso, para facilitar su comprensión. Se ha tomado un experimento donde el primer nivel se escoge en forma deliberada, esto es, se trata de un modelo mixto.

Ejemplo 2) Se ha medido el contenido glucosa en sangre humana. Se tiene un grupo control y dos preparados con dos drogas diferentes A y B. Para cada uno de los tres grupos ($a = 3$), se eligieron dos pacientes al azar ($b = 2$). A cada paciente se le tomó una muestra en tres horarios diferentes del día escogidos al azar ($c = 3$). Luego se hicieron tres repeticiones de la determinación para disminuir el error remanente ($n = 3$). Los datos medidos se muestran en el Cuadro 19.2 siguiente, junto con los cálculos preliminares:

Cuadro 19.2: Contenido de glucosa en sangre

Tratamientos:	Control			Droga A			Droga B											
Pacientes	1			2			1			2								
Horario	1	2	3	1	2	3	1	2	3	1	2	3						
Datos	99	97	95	96	98	94	88	88	86	80	83	80	78	80	78	79	80	78
	98	96	96	94	99	94	87	89	84	81	82	81	79	81	76	80	81	76
	99	98	94	95	98	92	86	86	85	80	80	82	78	80	74	78	82	74
Suma:	296	291	285	285	295	280	261	263	255	241	245	243	235	241	228	237	243	228
Total preparados:	872			860			779			729			704			708		
Total Grupos :	1732			1508			1412											

Paso 1) Se calcula el total de las observaciones realizadas:

$$T = \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X_{ijkl} = 99 + 98 + 99 + 97 + \dots + 76 + 74 = 4.652$$

Paso 2) Se calcula el total del cuadrado de las observaciones realizadas:

$$T_x^2 = \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X_{ijkl}^2 = (99)^2 + (98)^2 + (99)^2 + \dots + (74)^2 = 404.092$$

Paso 3) Se calcula la suma de los cuadrados de los totales de los sub-subgrupos (horarios) dividido por su tamaño muestral ($n = 3$) respectivo:

$$T_c = (1/n) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \left(\sum_{l=1}^{n=3} X \right)^2 = (1/3) [296^2 + 291^2 + 285^2 + \dots + 228^2] = 404.044,667$$

Paso 4) Se calcula la suma de los cuadrados de los totales de los subgrupos (pacientes) dividido por su tamaño muestral respectivo. ($n \cdot c = 9$)

$$T_b = (1/n \cdot c) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \left(\sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X \right)^2 = (1/9) [872^2 + 860^2 + 779^2 + 729^2 + 704^2 + 708^2] = 403.905,111$$

Paso 5) Se calcula la suma de los cuadrados de los totales de los grupos (tratamientos) dividido por su tamaño muestral respectivo. (b. n . c = 18)

$$T_a = (1/b.n.c) \sum_{i=1}^{a=3} \left(\sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X \right)^2 = (1/18) [1732^2 + 1508^2 + 1412^2] = 403.757,333$$

Paso 6) Se calcula el término de corrección:

$$T^2 / N = (1/a.b.c.n) \left(\sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X \right)^2 = (1/54) [4.652^2] = 400.761,185$$

Paso 7) Se calcula la suma de cuadrados total:

$$SS_T = \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X^2_{ijkl} - T^2 / N = T X^2 - T^2 / N = \text{Paso 2} - \text{Paso 6}$$

$$SS_T = 404.092 - 400.761,185 = 3.330,815$$

Paso 8) Se calcula la suma de cuadrados entre los grupos (entre tratamientos):

$$SS_E = (1/b.n.c) \sum_{i=1}^{a=3} \left(\sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X \right)^2 - T^2 / N = T_a - T^2 / N = \text{Paso 5} - \text{Paso 6}$$

$$SS_E = 403.757,333 - 400.761,185 = 2.996,148$$

Paso 9) Se calcula la suma de cuadrados dentro de grupos (pacientes dentro de tratamientos):

$$SS_{\text{subgrupos}} = (1/n.c) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \left(\sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X \right)^2 - (1/b.n.c) \sum_{i=1}^{a=3} \left(\sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X \right)^2 = T_b - T_a$$

$$SS_{\text{subgrupos}} = \text{Paso 4} - \text{Paso 5} = 403.905,111 - 403.757,333 = 147,778$$

Paso 10) Se calcula la suma de cuadrados dentro de subgrupos (horarios dentro de pacientes):

$$SS_{\text{sub-subgrupos}} = (1/n) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \left(\sum_{l=1}^{n=2} X \right)^2 - (1/n.c) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \left(\sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X \right)^2 = T_c - T_b$$

$$SS_{\text{sub-subgrupos}} = \text{Paso 3} - \text{Paso 4} = 404.044,667 - 403.905,111 = 139,556$$

Paso 11) Se calcula la suma de cuadrados remanente o error (datos dentro de tratamientos):

$$SS_T = \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X^2_{ijkl} - (1/n) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \left(\sum_{l=1}^{n=2} X \right)^2 = T X^2 - T_c = \text{Paso 2} - \text{Paso 3}$$

$$SS_T = 404.092 - 404.044,667 = 47,333$$

Paso 12) Se prepara la Tabla de ANOVA:

Fuente de Variación	v	SS	MS	F
Entre Tratamientos	(a-1) = 2	2996,148	1498.074	30,41*
Pacientes dentro de Tratamientos	a(b-1)=3	147,778	49,259	4,24*
Horarios dentro de pacientes	ab(c-1)=12	139,556	11,629	8,85***
Datos dentro de Tratamientos (error)	abc(n-1)=36	47,333	1,315	

Total 53 3330,815

En la primer columna se indican las diferentes fuentes de variación analizadas. En la segunda columna los grados de libertad respectivos. Las sumas de cuadrados y los cuadrados medios se colocan en la tercer y cuarta columna. Finalmente se calculan los estadígrafos F como sigue:

$$F_a = MS_a / MS_b = 1.498,074 / 49,259 = 30,41* > F_{0,95; 2; 3} = 9,55$$

$$\text{Pero } 30,41 < F_{0,99; 2; 3} = 30,81$$

Por lo tanto, se encontraron diferencias significativas entre los 3 grupos analizados.

El segundo test se hace para comprobar si hay diferencias significativas entre los pacientes dentro de los tratamientos que recibieron. El estadígrafo es:

$$F_b = MS_b / MS_c = 49,259 / 11,629 = 4,24* > F_{0,95; 3; 12} = 3,49$$

$$\text{Pero } 4,24 < F_{0,99; 3; 12} = 5,95$$

Se encontraron diferencias significativas entre los pacientes, dentro de los tratamientos hechos.

El tercer test, se efectúa para comprobar si hay diferencias significativas entre los horarios dentro de los pacientes. El estadígrafo es:

$$F_b = MS_c / MS_{\text{error}} = 11,629 / 1,315 = 8,85*** > F_{0,999; 12; 36} = 3,76$$

Se encontraron diferencias altamente significativas entre los horarios dentro de pacientes.

Como el primer test resultó significativo es necesario realizar las comparaciones múltiples entre los grupos investigados. Para los demás subniveles, como hay significación, se debe realizar la cuantificación de las componentes añadidas de varianza. En cuyo caso es:

Paso 13) Estimación de las componentes de la varianza:

$$DS^2_{BCA} = (49,259 - 11,629) / 6 = 6,27 \text{ (pacientes dentro de tratamientos: } n = 6)$$

$$DS^2_{CCB} = (11,629 - 1,315) / 3 = 3,438 \text{ (horarios dentro de pacientes: } n = 2)$$

$$DS^2 = 1,315 \text{ (remanente o error)}$$

La variación total de los niveles analizados, se calcula como la suma de las tres anteriores:

$$VAR = DS^2_{BCA} + DS^2_{CCB} + DS^2 = 11,02$$

Paso 14) Se calculan las componentes añadidas de varianza como un porcentaje:

$$(DS^2_{BCA} / VAR) 100 = (6,27 / 11,02) 100 = 56,9\% \text{ (pacientes dentro de tratamientos)}$$

$$(DS^2_{CCB} / VAR) 100 = (3,438 / 11,02) 100 = 31,2\% \text{ (horarios dentro de pacientes)}$$

$$(DS^2 / VAR) 100 = (1,315 / 11,02) 100 = 11,9\% \text{ (error)}$$

La notación con los paréntesis en los primeros pasos anteriores se hizo para denotar la manera de generalizar este tipo de ANOVA a más de tres niveles. Por cada subnivel que se agregue habrá un nuevo subíndice, o sea, una sumatoria más de los valores medidos. Luego el cálculo se hace así: Los dos primeros pasos son análogos, esto es, la suma de todos los valores encontrados para obtener el gran total y la suma de los cuadrados de los datos.

En el tercer paso habrá tantas sumatorias como subíndices (subniveles) haya, luego se coloca el paréntesis en la última de ellas para elevar al cuadrado los subtotaes comprendidos. En el cuarto paso se corre el paréntesis una posición a la izquierda para que abarque las dos últimas sumatorias y así elevar al cuadrado los subtotaes respectivos. Y así sucesivamente hasta llegar al término de corrección, que es cuando el paréntesis abarca a todas las sumatorias.

Para obtener las Sumas de Cuadrados se comienza restando las cantidades obtenidas en el Paso 2 y el término de corrección, para obtener el SS entre los grupos. A continuación, para el último subnivel, se resta el término de corrección al Paso inmediato anterior a éste. Y así sucesivamente. Por ejemplo, si el Paso 8 es el cálculo del término de corrección, la primera diferencia se hace entre: Paso 8 – Paso 7, la siguiente Paso 7 – Paso 6, luego Paso 6 – Paso 5 y así hasta llegar al Paso 2 – Paso 3 que da la remanente o error.

El Cuadro de ANOVA es enteramente similar al visto, y debe verificarse el total de los grados de libertad volcados, lo mismo que el total de la suma de cuadrados. Los MS se obtienen, como siempre, dividiendo los SS por sus respectivos grados de libertad. Y los estadígrafos F se obtienen dividiendo el MS del primer renglón por el MS del segundo, luego el MS del segundo por el MS del tercero y así sucesivamente. Los demás pasos a seguir dependen de sí se encuentra significación, en la misma forma que la del ejemplo visto.

19.4 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|-------|-------|
| 1) El ANOVA encajado tienen ventajas respecto al ANOVA simple. | V | F |
| 2) El encajado puede separar la variabilidad debida a los distintos subniveles estudiados. | V | F |
| 3) Mencionar las ventajas de este modelo:..... | | |
| 4) El modelo puro del encajado es de Modelo I en todos los niveles. | V | F |
| 5) Explicar como se diseña un modelo mixto de ANOVA encajado: | | |
| 6) El encajado se llama también jerárquico porque toma en cuenta los distintos niveles. | V | F |
| 7) Explicar teóricamente los modelos puros y mixtos. | | |
| 8) Al volcar los datos conviene realizar los totales preliminares. | V | F |
| 9) El primer paso siempre es encontrar el total de las mediciones efectuadas. | V | F |
| 10) En el segundo paso hay que calcular el total de los cuadrados de las observaciones. | V | F |
| 11) En el tercer paso se calcula la SS entre los grupos analizados (segundo nivel). | V | F |
| 12) Explicar como se generalizan los demás pasos:..... | | |
| 13) Siempre hay que calcular la componente añadida de varianza. | V | F |
| 14) Los estadígrafos F para cada nivel se calculan dividiendo su MS por el MSError. | V | F |
| 15) Hay que comparar F con un valor de tablas, con los grados de libertad respectivos. | V | F |
| 16) Si hay significación en el primer nivel, hay que hacer comparaciones múltiples. | V | F |
| 17) Cuando en los subniveles haya significación se debe calcular la componente añadida. | V | F |
| 18) Explicar como se generaliza este modelo a más de tres niveles:..... | | |

2) Resolver numéricamente el siguiente problema:

Tratamientos:	Control			Droga 1			Droga 2											
Pacientes	1			2			1			2								
Dieta	1	2	3	1	2	3	1	2	3	1	2	3						
Datos	450	431	446	445	440	449	478	470	468	480	477	483	510	518	501	499	512	502
	440	425	442	448	443	450	476	465	463	475	479	485	520	519	509	505	515	505

3) Ídem anterior

Pacientes	1			2			1			2			1			2		
Horario	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Datos	131	131	136	150	140	160	157	154	147	151	147	162	134	138	135	138	139	134
	130	125	142	148	143	150	145	142	153	155	147	152	125	138	136	140	138	127