Análisis de varianza con dos criterios de clasificación (diseños factoriales)

©Pedro Morales Vallejo
Universidad Pontificia Comillas
Facultad de Ciencias Humanas y Sociales
(última revisión: 2 de Diciembre, 2009)

índice

1. Planteamiento general y conceptos previos 3
   1.1. *Clasificación de los sujetos en dos categorías* 3
   1.2. *Qué pretendemos comprobar* 3
   1.3. *Importancia de la interacción* 4
   1.4. *Características de la muestra* 5
   1.5. *Número de sujetos en cada clasificación* 5
      1.5.1. *Número idéntico de sujetos en cada clasificación* 6
      1.5.2. *El número de sujetos en términos absolutos* 7
   1.6. *Importancia del tipo de categorías de clasificación* 7

2. Método 8
   2.1. Disposición de los datos 8
   2.2. Visión global del proceso 9
   2.3. Cálculos previos 10
   2.4. Cálculos propios del análisis de varianza 11
      2.4.1. Sumas de Cuadrados 11
      2.4.2. Grados de libertad 12
      2.4.3. Varianzas o Cuadrados Medios 12
      2.4.4. Comparación o contraste de varianzas (razón F) 13
      2.4.5. Resultados finales: tabla de resultados 14
      2.4.6. Interpretación 14
      2.4.7. Resumen del procedimiento 15

3. Análisis adicionales 15
   3.1. Coeficientes de asociación 16
      3.1.1. Coeficiente ω^2 16
      3.1.2. Coeficiente η^2 17
   3.2. Contrastes posteriores 18
      3.2.1. Contraste entre las medias de niveles del mismo factor 19
      3.2.2. Contraste entre las medias *dentro del mismo nivel* de un factor (cuando la interacción es significativa) 22
   3.3. El *tamaño del efecto* en los diseños factoriales 24
      3.3.1. Cuando el factor secundario es una variable manipulada 24
      3.3.2. Cuando el factor secundario es una característica natural de la población 24

4. La representación gráfica de los resultados como ayuda a la interpretación en los diseños factoriales 26

5. Análisis de varianza para diseños factoriales en EXCEL y SPSS 28

6. Referencias bibliográficas 29

Anexo. Análisis de Varianza (diseños factoriales) en Internet 29

1. Planteamiento general y conceptos previos

1. 1. *Clasificación de los sujetos en dos categorías*

La disposición de los datos al clasificar a los sujetos es importante, no sólo como ilustración, sino porque contribuye a entender lo que estamos haciendo.

Se clasifican los sujetos según dos variables o *categorías de clasificación* (que suelen denominarse *factores)* en un cuadro de doble entrada. En el ejemplo sugerido en la tabla 1 los factores son a) actividades didácticas y b) la duración de estas actividades. La variable dependiente (la que medimos a los sujetos) sería en este caso nivel de aprendizaje. Los factores o criterios de clasificación pueden ser más de dos, pero en la presentación del método nos limitamos a dos nada más[1]. Cada uno de los dos factores está dividido en dos o más *niveles* o *subcategorías.* Dada la disposición de los datos, también suele hablarse de *filas* y *columnas* (en este caso hay tantas filas y columnas cuantos niveles tengan los factores).

Un ejemplo típico podría ser el puesto en la tabla 1; clasificamos a los sujetos según dos criterios:

a) Factor A: según hayan participado en una determinada *actividad.* Este factor tiene dos niveles; suponemos que se trata de dos actividades distintas, o dos variantes de la misma actividad.

b) Factor B: según el *tiempo* que haya durado esta actividad; en este factor de *duración* suponemos tres niveles o duraciones distintas.

			Factor A *(actividades)*	
			A_1	A_2
Factor B	20 minutos	B_1		
(duración)	40 minutos	B_2		
	una hora	B_3		

Tabla 1

En cada clasificación tenemos *el mismo número de sujetos*. Con esta misma disposición de los datos se pueden visualizar muchos posibles diseños experimentales utilizando otros criterios de clasificación.

1.2. *Qué pretendemos comprobar*

Lo que pretendemos comprobar es en qué grado las diferencias que encontremos en la *variable dependiente* (la que hemos *medido,* en este caso puede ser *rendimiento escolar)* se explican por las diferencias entre las *actividades,* o por las diferencias entre los *tiempos,* o por alguna combinación *actividad-duración*.

Analizamos por lo tanto tres posibles fuentes de diferencias:

[1] Con sólo dos criterios de clasificación, el planteamiento más frecuente, el método que proponemos es muy sencillo, sin más ayuda que una calculadora con la media y desviación típica programadas; además en procesos de aprendizaje ayuda a entender lo que estamos haciendo. Con más de dos criterios de clasificación se puede adaptar y utilizar el mismo procedimiento pero es preferible acudir directamente a programas de ordenador; los resultados son fáciles de interpretar si se ha aprendido antes a resolver e interpretar el planteamiento con dos criterios de clasificación. El análisis de varianza factorial se puede también hacer con EXCEL, SPSS y programas disponibles en Internet (mencionamos algunos en el Anexo).

a) Una actividad puede ser más eficaz que otra, independientemente de su duración,

b) Una duración puede ser más eficaz que otra en cualquier actividad,

c) Una actividad puede ser más eficaz que otra, pero solamente si dura en determinado tiempo.

Podemos formular por lo tanto tres *Hipótesis Nulas* (formuladas en referencia al caso planteado en la tabla 1):

1ª Las *dos* muestras que han seguido actividades distintas proceden de la misma población (no hay diferencia significativa entre las *actividades;* no se aprende más con una que con otra).

2ª Las *tres* muestras que han trabajado durante tiempos de distinta duración proceden de la misma población (no hay diferencia significativa entre las *duraciones).*

3ª Las *seis* muestras que han trabajado en actividades distintas y con duraciones distintas proceden de la misma población (no hay diferencias significativas entre los diversos subgrupos en los que se combinan *actividad-duración).*

Tendremos que calcular tres varianzas que corresponden a las *tres fuentes de diferencias,* y una cuarta varianza que expresa la variabilidad *normal* o *aleatoria,* y que nos servirá como término de comparación de las otras varianzas (denominador de la razón F). Consecuentemente al final del proceso tendremos tres razones F que nos permitirán aceptar o no aceptar las tres Hipótesis Nulas.

Como en el análisis de varianza en general, lo que verificamos es la relación que puede haber entre la *variable dependiente* (la característica de los sujetos que hemos *medido)* y los *criterios* que nos han servido para clasificar a los sujetos.

Una manera posible de analizar estos datos sería considerar que tenemos 6 grupos de muestras independientes y utilizar un análisis de varianza de una clasificación simple para varias muestras independientes. Pero en este caso, si obtuviésemos una F significativa que nos indicara que hay diferencias entre los grupos, no sabríamos si atribuirla a que las actividades son distintas, o a que las distintas duraciones producen resultados distintos, o a que determinadas combinaciones actividad-duración son más eficaces que otras. Los resultados podrían ser ambiguos o difíciles de interpretar con precisión.

1.3. *Importancia de la interacción*

Otra manera de enfocar la solución es hacer dos análisis de varianza: uno para comparar las dos *actividades* (nos bastaría en este ejemplo un simple contraste de medias ya que se trata solamente de dos actividades) y otro para comparar las tres *duraciones.* Pero nos quedaríamos sin saber la importancia de la *relación actividad-duración*. Esta relación es la que denominamos genéricamente *interacción*. En nuestro ejemplo puede ser que una actividad sea preferible, produzca mejores resultados, pero solamente si tiene una duración óptima. En general en la investigación educacional o psicológica la *interacción* puede tener mucha importancia.

Con frecuencia lo que tenemos es:

1. Un factor denominado *principal* que es el objeto principal de nuestro estudio (en nuestro ejemplo distintas actividades);

2. Otro factor cuyos niveles suelen ser *circunstancias* o *condiciones* que pueden afectar a los niveles del factor principal; por ejemplo duraciones distintas, con o sin alguna

circunstancia, distintas variantes, tipos de sujetos, *dosis*…; también puede tratarse de agrupaciones naturales de lo sujetos (por ejemplo género, lugar de procedencia, grupo étnico, etc.).

Un método (experiencia, actividad, terapia, etc.) puede ser mejor en unas circunstancias, pero en otras puede ser tan bueno o malo como los demás (circunstancias posibles que pueden modificar los resultados en la evaluación de unas actividades: medios auxiliares, tiempo dedicado, hora del día o época del año, etc.). Podemos pensar que de manera análoga algunas medicinas son mejores para unos pacientes que para otros, o pueden estar contraindicadas en determinadas personas, o pueden ser incompatibles con otros medicamentos… se trata también de *interacciones:* los efectos de un *tratamiento* pueden depender del nivel de una segunda variable. Podemos pensar en semillas de una planta: una variedad puede producir una mayor producción pero *solamente* en un determinado tipo de suelo, o con un determinado fertilizante, o una determinada cantidad de riego o de exposición solar, etc.[2]

Es fácil ver que este planteamiento tiene una especial utilidad *precisamente* porque permite analizar las interacciones. También es verdad que cuando la interacción es significativa la interpretación puede ser menos simple; hay que matizar más las conclusiones.

1.4. *Características de la muestra*

a) En este planteamiento todos los sujetos son *distintos* Se trata por lo tanto de *muestras independientes;* no hay sujetos *repetidos* en más de un grupo (otras variantes de este mismo modelo, con los sujetos clasificados en dos o más factores, son válidas para muestras relacionadas, pero no es éste el caso que presentamos ahora).

b) En un *diseño experimental en sentido propio* los sujetos deben ser asignados *aleatoriamente* a las distintas condiciones experimentales para poder controlar otras variables.

Los sujetos, por ejemplo alumnos, han sido asignados *aleatoriamente* a cada uno de los subgrupos para poder generalizar los resultados (controlamos de esta manera otras variables extrañas que pueden influir en la variable dependiente que hemos medido). Si no hay asignación aleatoria no se trata de un diseño experimental en sentido propio. La *aleatoriedad* permitirá generalizar las conclusiones *(validez externa)* a otros sujetos de la misma población (*representados* por *esta* muestra) con una mayor seguridad.

Si no hay asignación aleatoria de los sujetos a los grupos, caben otros modos de controlar variables con otros diseños (con *sujetos igualados* en variables importantes, o utilizando diseños más complejos). En muchas investigaciones (o simplemente en estudios experimentales) la *no aleatoriedad* es casi la norma (por su facilidad: se trabaja con grupos *hechos* o disponibles). Estos análisis (que no responden a un *diseño experimental* en sentido propio) son también útiles, pero hay que preguntarse siempre *qué otras cosas* pueden estar influyendo en la variable dependiente, y tener más *cautela* en la interpretación y sobre todo en la extrapolación de los resultados.

1.5. *Número de sujetos en cada clasificación*

Con respecto al número de sujetos hay que distinguir dos cuestiones distintas, a) la conveniencia de disponer de un número idéntico de sujetos en cada clasificación, y b) el número de sujetos en términos absolutos en cada clasificación (en cada celda).

---

[2] Precisamente este tipo de diseños tuvieron su origen en la investigación en el campo de experimentación agrícola y de la biología en general.

1.5.1. *Número idéntico de sujetos en cada clasificación*

En este modelo de análisis de varianza debe de haber el *mismo número de sujetos* en cada clasificación; el método es así más potente aunque no se cumplan los presupuestos de *normalidad* en la población y de *homogeneidad de varianzas.* El método resulta además mucho más sencillo y tal como lo explicamos aquí suponemos siempre que el número de sujetos es el mismo.

El distinto número de sujetos en cada clasificación deja de ser problemático cuando las frecuencias en cada clasificación son *proporcionales.* Números proporcionales quiere decir lo mismo que las *frecuencias teóricas* o *esperadas* en los planteamientos de *ji cuadrado* (aun así advertimos que el procedimiento que expondremos supone un idéntico número de sujetos en cada clasificación; con grupos de idéntico tamaño el procedimiento es además más sencillo).

Cuando el número de sujetos es desigual, los procedimientos sugeridos para igualar el número de sujetos son varios:

1º *Descartar* sujetos aleatoriamente.

Con muestras razonablemente grandes la recomendación habitual es *descartar aleatoriamente datos* para obtener frecuencias iguales[3] (también se pueden descartar observaciones para conseguir frecuencias proporcionales). Esta recomendación es aplicable en principio a cualquier planteamiento de análisis de varianza porque, como ya hemos indicado, con un número idéntico de sujetos en cada grupo pierde importancia la violación de normalidad y de homogeneidad de varianzas.

El desechar sujetos debe hacerse de manera *estrictamente aleatoria;* la condición que indican los autores mencionados es que el número mínimo de sujetos por celda no sea inferior a 10, más o menos.

2º *Estimar* las puntuaciones que faltan.

Si en algún caso falta algún sujeto en una celda (es normal la pérdida de sujetos en planteamientos experimentales) se pueden estimar las puntuaciones que faltan *(missing scores)* poniendo en su lugar la *media* de la celda; esta práctica no afecta apenas a los resultados si los sujetos por celda son al menos 10. Con este procedimiento se pueden estimar entre un 10 y un 25% de los datos.[4]

3º Utilizar las *medias*

Si utilizamos las medias de cada subgrupo en vez de las puntuaciones individuales, tendremos en cada clasificación n = 1, y también se puede llevar a cabo este análisis de varianza (diseños factoriales) con un solo sujeto en cada clasificación. En principio este tipo de solución no es deseable porque supone mucha pérdida de información, pero tampoco conviene descartar esta posibilidad porque puede tener su interés en sí misma. Al tratar de las variantes de los diseños factoriales trataremos este caso específico en el que podemos utilizar medias (y consecuentemente, n = 1 en cada celda) en vez de puntuaciones individuales. En este caso el procedimiento viene a ser el mismo que el análisis de vainas para muestrs relacionadas.

---

[3] Por ejemplo Glass y Stanley (1974:439); Escotet (1980:84-85); Linton, Gallo y Logan (1975:131).

[4] Linton, Gallo y Logan (1975:131). Cuando *los dos factores tienen idéntico número de niveles* (como en el diseño denominado *cuadrado latino* que no explicamos aquí) hay una fórmula específica para estimar el valor de las puntuaciones que nos faltan; puede verse en Tejedor (1984:236).

1.5.2. *Número de sujetos en términos absolutos*

El número de sujetos necesario (o conveniente) *en cada clasificación* depende de varias variables. Nos referimos a tablas nxn (dos criterios de clasificación). En términos generales necesitaremos más sujetos a) cuando los niveles de cada factor son menos (necesitamos más sujetos en tablas 2x2 que en tablas 4x4) y b) la magnitud de las diferencias en la que estamos interesados; si queremos detectar diferencias pequeñas nos harán falta muchos más sujetos que si sólo estamos interesados en diferencias grandes.

Como *criterio orientador,* y si se trata de detectar *diferencias entre moderadas y grandes* (ciertamente no pequeñas) podemos pensar en el número de sujetos en cada clasificación que figura en la tabla 2.

Magnitud de la tabla	*Número de sujetos*
2x2	entre 13 y 32
2x3	entre 11 y 26
2x4	entre 9 y 22
3x3	entre 7 y 18
3x4	entre 6 y 15
4x4	entre 5 y 12

Tabla 2

Para detectar con seguridad diferencias pequeñas (y por lo general de menor interés) son necesarios muchos más sujetos[5].

1.6. *Importancia del tipo de categorías de clasificación*

Tratando de diseños factoriales es importante recordar que las *categorías* de clasificación (variables independientes) pueden ser:

a) *Fijas* (escogidas según criterio del investigador),
b) *Aleatorias* (escogidas aleatoriamente de una población mayor);
c) *Mixtas* (una categoría fija y otra aleatoria).

Del hecho de que las categorías sean fijas o aleatorias:

1. No depende el método que vamos a seguir, que es el mismo;
2. Sí depende la *generalización* de las conclusiones, que será mayor con categorías aleatorias. Con categorías fijas, las conclusiones son aplicables en principio a las categorías empleadas.
3. Sí va a depender el *denominador* de la razón F empleado en cada caso, como veremos después (en la tabla 4).

En el caso de *categorías fijas* (el más frecuente) el denominador va a ser siempre el mismo que ya hemos visto en el análisis de varianza para varias muestras independientes (es decir, los *cuadrados medios dentro de los grupos),* por lo que esta distinción entre tipos de categorías no es en principio especialmente importante. Se trata sin embargo de conceptos que conviene recordar y que a la vez sugieren posibilidades de investigación. En investigación educacional (y de otro tipo) podemos tener categorías aleatorias si tenemos, como factor o

[5] El número de sujetos en cada clasificación lo tratamos en el anexo VI y también en el documento *Tamaño de la muestra: ¿Cuántos sujetos necesitamos?* http://www.upco.es/personal/peter/investigacion/Tama%F1oMuestra.pdf. Estos criterios orientadores están adaptados de las extensas tablas de Kirk (1995:401 y tabla E.15).

categoría de clasificación, centros, aulas, profesores, etc., seleccionados aleatoriamente de una población mayor.

Es importante caer en la cuenta que son las *categorías de clasificación,* y no los sujetos, las que pueden ser o no ser escogidas aleatoriamente. Los sujetos, cuyos resultados (variable dependiente) se van a analizar, deben asignarse aleatoriamente a los diversos subgrupos si se pretende que el diseño se aproxime más a lo experimental. Con muestras aleatorias es más legítimo generalizar las conclusiones a la población porque así se controlan otras variables desconocidas que podrían contaminar o explicar los resultados.

2. Método

2.1. Disposición de los datos

En el ejemplo que nos va a servir para exponer el método[6] tenemos dos factores (tabla 3):

1. Factor A (métodos) dividido en tres niveles (tres variantes metodológicas).
2. Factor B (profesores) dividido en cuatro niveles (cuatro profesores).

En la tabla 3 aparecen los datos de los sujetos según pertenezcan a un profesor (B_1, B_2, B_3 y B_4) y a un método determinado (A_1, A_2 y A_3).

En este hipotético caso consideramos que ambas categorías son fijas: los métodos han sido escogidos con criterios lógicos y los profesores suponemos que son simplemente los que estaban disponibles para participar en una investigación; no han sido escogidos aleatoriamente de una población mayor de profesores.

Tenemos por lo tanto 12 combinaciones profesor-método. En cada combinación hay cinco alumnos

Es conveniente disponer los datos con claridad, en un cuadro de doble entrada según los dos criterios de clasificación, en el que aparezcan las puntuaciones directas de todos los sujetos, tal como aparece en la tabla 3. Incluso haciendo todos los cálculos con un programa de ordenador, es útil tener *in mente,* en éste y otros modelos de análisis de varianza, cómo se disponen los datos.

[6] Los datos los tomamos del ejemplo que presentan Guilford y Fruchter (1973) para explicar este modelo de análisis de varianza (el procedimiento que seguimos aquí es distinto, semejante a los ya vistos); modificamos la designación de las variables para utilizar términos que nos resultan más familiares (métodos y profesores).

profesores (factor B)	métodos (factor A)			Medias de B
	A_1	A_2	A_3	
Profesor B_1	6	4	4	$M_{B1} = 3$
	4	1	2	
	2	5	2	
	6	2	1	
	2	3	1	
Media	$M_{A1B1} = 4$	$M_{A2B1} = 3$	$M_{A3B1} = 2$	
desviación	1.789	1.414	1.095	
Profesor B_2	8	6	3	$M_{B2}= 4$
	3	6	1	
	7	2	1	
	5	3	2	
	2	8	3	
Media	$M_{A1B2} = 5$	$M_{A2B2} = 5$	$M_{A3B2} = 2$	
desviación	2.280	2.191	.894	
Profesor B_3	7	9	6	$M_{B3} = 6$
	6	4	4	
	9	8	3	
	8	4	8	
	5	5	4	
Media	$M_{A1B3} = 7$	$M_{A2B3} = 6$	$M_{A3B3} = 5$	
desviación	1.414	2.098	1.789	
Profesor B_4	9	7	6	$M_{B4} = 7$
	6	8	5	
	8	4	7	
	8	7	9	
	9	4	8	
Media	$M_{A1B4} = 8$	$M_{A2B4} = 6$	$M_{A3B4} = 7$	
desviación	1.095	1.673	1.414	
Medias de los métodos (A)	$M_{A1} = 6$	$M_{A2} = 5$	$M_{A3} = 4$	Media y σ total $M_t = 5$ $\sigma_t = 2.4966$

Tabla 3

2.2. Visión global del proceso

Los cálculos que vamos a hacer los dividimos en tres tipos: cálculos previos, análisis de varianza en sentido propio y análisis complementarios.

1º *Cálculos previos*

Van a ser los mismos que en otros procedimientos análogos *(medias, desviaciones, varianza de medias…)*

El cálculo de las *desviaciones típicas de las medias* puede parecer poco usual, pero ya hemos visto que simplifica mucho las operaciones y ahorra tiempo si no se utiliza directamente un programa de ordenador. Ya hemos indicado además en otra ocasión que propiamente no se trata de calcular desviaciones o varianzas de medias, sino de todos los sujetos, pero asignándoles como puntuación individual la media de su grupo; como los grupos son de idéntico tamaño basta calcular la desviación de las medias. Estos cálculos se hacen con rapidez con una calculadora corriente con programación estadística o utilizando una *hoja de cálculo.*

2º *Cálculos específicos del análisis de varianza*.

A partir de los cálculos precedentes *(desviaciones típicas de medias)* se calculan con facilidad las Sumas de Cuadrados, que suele ser lo más laborioso en el análisis de varianza, y el resto de las operaciones.

3º *Análisis complementarios.*

Frecuentemente van a ser *coeficientes de asociación* para ver la importancia relativa de las distintas fuentes de varianza; también podemos hacer los contrastes posteriores que resulten de interés (con sus correspondientes *tamaños del efecto).*

Desde el comienzo debemos tener a la vista la *tabla de resultados* (tabla 5), que es semejante a las que encontramos en otros modelos de análisis de varianza.

Más adelante presentamos un resumen de todo el procedimiento de análisis (tabla 6), y puede resultar útil verlo desde el principio, ya que es semejante a otros procedimientos de análisis de varianza con los que ya estamos familiarizados.

Aunque vamos a explicar el proceso de manera muy pormenorizada, se puede ver de un *golpe de vista* observando la tabla 4 (datos y cálculos previos ya hechos) y la tabla 6 (tabla de resultados con las fórmulas incluidas); estas dos tablas pueden ser suficientes, como referencia, para llevar a cabo todo el proceso.

El proceso completo, lo que vamos a hacer o podemos hacer, lo sintetizamos de esta manera:

1. *Análisis de varianza* propiamente dicho, que nos va decir qué fuentes de variación son estadísticamente significativas (razones F).
2. *Coeficientes* del tipo ω^2 o η^2 que nos van a permitir apreciar la relevancia e importancia práctica de los resultados.
3. *Representación gráfica* de los resultados como ayuda a la interpretación; ya veremos cómo hacerla.
4. *Contrastes posteriores* en el caso de que sean necesarios o nos parezcan convenientes, y el *tamaño del efecto* al comparar dos medias entre sí..
5. *Interpretación global* de los resultados, incluyendo valoraciones e interpretaciones cualitativas, conclusiones, etc.

2.3. Cálculos previos

Estos *cálculos previos* facilitan los cálculos de las *Sumas de Cuadrados*. Se trata simplemente del cálculo de medias y desviaciones típicas (suponemos que estos cálculos se hacen al menos con calculadora programada.

1º Cálculo de medias y desviaciones

En la tabla 3 tenemos ya calculadas estas medias y desviaciones:

1) Media y desviación de cada uno de los subgrupos (de las 12 combinaciones AB)
2) Media de cada columna (tres medias; métodos, factor A)
3) Media de cada fila (cuatro medias; profesores, factor B)
4) Media y desviación de los totales (de los 60 datos)

La *desviación típica de todos los datos* (de N) en realidad no nos es necesaria; es útil para calcular la *suma de cuadrados total,* que es igual a la *suma de las otras sumas de cuadrados.* Esta *suma de cuadrados total* nos permite comprobar que las otra sumas de cuadrados están bien hechas.

La media total (de N = 60; *media de las medias* de A o de B) tampoco nos es necesaria; es útil si vamos a calcular la desviación típica de los totales siguiendo el procedimiento explicado en el Anexo III.

2º Cálculo de *varianzas de medias*

Estas varianzas (calculadas dividiendo por N, no por N-1) nos van a simplificar el cálculo de las *sumas de cuadrados* (en realidad ya sabemos que se trata de varianzas calculadas en toda la muestra, pero asignando a cada sujeto no la puntuación obtenida sino la media de su grupo; al tratarse de grupos de idéntico tamaño nos basta calcular las varianzas de las medias).

1) Varianza de las medias de los niveles del factor A (métodos, tres medias): $\sigma^2_{MA} = .816^2 =$ 0.8858

2) Varianza de las medias de los niveles del factor B (profesores, cuatro medias): $\sigma^2_{MB} = 1.581^2 =$ 2.50

3) Varianza de las medias de todas las combinaciones AxB (doce medias) $\sigma^2_{MAxB} = 1.871^2 =$ 3.50

Si tenemos desde el principio una visión global del proceso, podemos ir directamente al cálculo de las Sumas de Cuadrados, que veremos enseguida. Los resultados se pueden ir poniendo directamente en la *tabla de resultados* (tabla 5). En casi todos los casos se trata de multiplicar las *varianzas de las series de medias* por el número total (N) de sujetos.

2.4. Cálculos propios del análisis de varianza

2.4.1. Sumas de Cuadrados (SC)

A partir de los datos que ya tenemos vamos a hacer todos los cálculos del análisis de varianza que aparecerán en la tabla 5 *(tabla de resultados)*, pero que iremos presentando por partes. En la práctica, y según vamos calculando las varianzas necesarias, podemos ir directamente a la tabla de resultados, pues las operaciones que vamos a hacer son muy sencillas (resumidas en la tabla 6).

Para calcular las diversas *varianzas* (o Cuadrados Medios, CM) necesitamos calcular el *numerador* (o Suma de Cuadrados) y el *denominador* (o Grados de Libertad, gl).

Las Sumas de Cuadrados las calculamos multiplicando las varianzas (las desviaciones ya calculadas elevadas al cuadrado) por el número total de sujetos, ya que todos ellos contribuyen a todas las varianzas (en este caso N = 60):

Sumas de Cuadrados (o *numerador)* correspondiente a la varianza de:

1. **El total** $SC_t = N\sigma^2_t$ $= (60)(2.4966)^2 =$ 374

2. **Factor A** (métodos) $SC_A = N\sigma^2_A$ $= (60)(.816)^2 =$ 40

3. **Factor B** (profesores) $SC_B = N\sigma^2_B$ $= (60)(1.581)^2 =$ 150

4. **Interacción** entre los dos factores, o variabilidad (diferencias en los resultados) debida a las diversas combinaciones entre profesores y métodos y que simbolizamos como AxB,

$$SC_{AxB} = N\sigma^2_{AxB} - (SC_A + SC_B) \qquad = (60)(1.871)^2 - (150 + 40) = \qquad 20$$

5. **Dentro de los grupos**, o *residual;* variabilidad *dentro* de los 12 subgrupos una vez eliminada la variabilidad debida a los métodos, a los profesores y a la interacción entre métodos y profesores.

   Esta varianza *residual* es la varianza *dentro* de los grupos (de todas las combinaciones AB).

   Podemos calcularla de dos maneras:

   1) A partir de la *suma de cuadrados del total*

   $$SC_{dentro} = SC_t - (SC_A + SC_B + SC_{AxB}) = 374 - (150 + 40 + 20) = \qquad 164$$

   2) Esta suma de cuadrados *dentro* de los grupos también se puede calcular directamente, y puede servir de comprobación, a partir de las *varianzas dentro de los grupos* (12 en este ejemplo) si las hemos calculado previamente[7]: se suman y se multiplica esta suma por el número de sujetos que hay en cada grupo:

   $$SC_{dentro} = n(\sigma_1^2 + \sigma_2^2 + \ldots \sigma_n^2)$$

Las fórmulas para calcular las Sumas de Cuadrados (o numeradores de las varianzas) están puestas también en la tabla 6.

Podemos comprobar que la *suma de las sumas de cuadrados parciales* es igual a la *suma de cuadrados total;* simplemente hemos descompuesto el numerador de la varianza total.

2.4.2. Grados de libertad (gl)

Los *grados de libertad* van a ser el *denominador* de cada varianza. El determinar los grados de libertad que corresponden a cada varianza es muy sencillo:

1. Del factor A (métodos)	A-1 =	(3-1) =	2
2. Del factor B (profesores)	B-1 =	(4-1) =	3
3. De la interacción (AxB)	(A-1)(B-1) =	(3-1)(4-1) =	6
4. Dentro de los grupos	N -k =	(60 – 12) =	48
5. Del total	N - 1 =	(60-1) =	59

Podemos verificar que los grados de libertad del total es igual a la suma de todos los demás.

2.4.3. Varianzas o Cuadrados Medios (CM)

Dividiendo la *Suma de Cuadrados* por los *Grados de Libertad* tenemos las diversas varianzas en las que hemos descompuesto la varianza total, tal como están ya calculadas en la

7 Recordamos que la *suma de las varianzas* se calcula con rapidez introduciendo en la calculadora (con programación estadística) todas las desviaciones típicas; la suma de las varianzas la obtenemos en la función $\sum x^2$

tabla 5, y que muestra los resultados finales del análisis de varianza. Normalmente es la única tabla que se presenta pues allí están los datos que permitirán interpretar las conclusiones.

Como en el caso de varias muestras independientes, si hubiéramos calculado en los 12 subgrupos las desviaciones típicas dividiendo por N-1, *los cuadrados medios dentro de los grupos* (denominador de la razón F) es simplemente la *varianza media de los grupos* $(\Sigma\sigma^2_{n-1}/k)$.

2.4.4. Comparación o contraste de varianzas (razón F)

Finalmente comparamos las varianzas de interés (las de los factores y su interacción) con la varianza que podemos considerar como *aleatoria,* para comprobar qué *fuentes de varianza* podemos considerar significativas o superiores a lo que se puede esperar por azar.

En estos planteamientos tenemos tres fuentes de varianza que nos interesa examinar:

1) La varianza debida al factor A (métodos)
2) La varianza debida al factor B (profesores)
3) La varianza debida a la interacción entre los dos factores (A x B).

Es decir, deseamos calibrar cuál es el influjo de las dos variables que hemos utilizado para clasificar a los sujetos (y que genéricamente llamamos aquí factor A y factor B, en este caso métodos y profesores), y también el influjo de la relación entre las dos variables. En una palabra: entre los sujetos tenemos diferencias (varianza, expresada en la varianza total), y nos preguntamos ¿Cuáles de estas tres *fuentes de diferencias* son importantes, por encima de lo puramente casual?

Estas tres varianzas van a ser el *numerador* de otros tantos cálculos de la *razón F*. El denominador será el *término de comparación.* Compararemos cada una de las tres varianzas con la varianza que podamos considerar como *aleatoria,* la que hubiéramos encontrado de no existir lo específico de los *métodos, los profesores y su interacción* que es precisamente lo que estamos investigando. Ésta es la varianza que hemos denominado antes *dentro de los grupos,* y que quizás con más propiedad puede denominarse varianza *residual.* Sin embargo esto no es siempre así, el término apropiado de comparación va a depender de que las categorías de clasificación sean fijas, aleatorias o mixtas. En la tabla 4 exponemos cual debe ser el *denominador* de la razón F[8].

En nuestro ejemplo se trata de un modelo de categorías fijas, por lo que en todos los casos el denominador o término de comparación será CM_{dentro} o la varianza (Cuadrados Medios) *dentro de los grupos* o *residual.* Si en este ejemplo los profesores hubieran sido escogidos aleatoriamente de un conjunto mayor, el factor B sería una categoría aleatoria, pero suponemos que se trata de los profesores que estaban disponibles para hacer esta investigación.

---

[8] La justificación puede verse en Guilford y Fruchter (1973:257)

	Numerador, o varianza cuya significación deseamos comprobar	**Denominador** *(varianza aleatoria)*, o término de comparación del numerador
Modelo de categorías ***fijas***	CM_A .... CM_B .... CM_{AxB}	CM_{dentro} en todos los casos
Modelo de categorías ***aleatorias***	CM_A .... CM_B .... CM_{AxB}	CM_{AxB} CM_{AxB} CM_{dentro}
Modelo de categorías ***mixtas***	CM de la categoría aleatoria .... CM de la categoría fija .... CM_{AxB}	CM_{dentro} CM_{AxB} CM_{dentro}

Tabla 4, denominador de la razón F

2.4.5. Resultados finales: *tabla de resultados*

Los resultados finales del análisis de varianza se ponen en una tabla, tal como lo hacemos ahora en la tabla 5.

Origen de la variación	SC *numerador*	gl *denominador*	$CM = \sigma^2 = \frac{SC}{gl}$	$F = \frac{\sigma^2}{\sigma^2}$	p
Factor A (métodos)	40	2	$\frac{40}{2} = 20$	$\frac{20}{3.42} = 5.85$	< .01
factor B (profesores)	150	3	$\frac{150}{3} = 50$	$\frac{50}{3.42} = 14.50$	< .01
interacción A x B	20	6	$\frac{20}{6} = 3.32$	$\frac{3.32}{3.42} = .97$	no sign.
dentro de los grupos	164	48	$\frac{164}{48} = 3.42$		
total	374	59			

Tabla 5

2.4.6. Interpretación

1. La varianza debida a la combinación profesor-método es prácticamente nula; la variabilidad (diferencias) en los resultados no se debe a determinadas combinaciones profesor-método (no hay profesores que funcionen especialmente mejor o peor que los demás con un determinado método).

2. Las varianzas originadas tanto por los profesores como por los métodos son ambas significativas; si hay diferencias es porque los profesores son distintos y también porque los métodos son distintos. Hay profesores más eficaces (con cualquier método) y métodos más eficaces (con cualquier profesor).

3. La mayor variabilidad se debe con mucho a los profesores (factor B), independientemente del método que utilicen. Sin embargo si lo que se deseaba probar era que los métodos son distintos en eficacia, que no da lo mismo uno que otro, esto puede darse por

probado, porque la varianza de los métodos (factor A) es significativamente superior a cero (hemos obtenido una F= 5.85 y el valor necesario que vemos en las tablas de la F es 5.08). Aun así el *peso* de los profesores, en los resultados finales, parece mayor que el de los métodos. Para analizar e interpretar mejor los resultados haremos después algún cálculo adicional.

2.4.7. Resumen del procedimiento

El procedimiento puede parecer a primera vista un tanto complejo, pero si se examina *paso a paso* se advierte enseguida su simplicidad.

La tabla 6 puede ser una referencia de uso cómodo para hacer todos los cálculos con rapidez, si estamos familiarizados con los símbolos :

N = Número total de sujetos

n = Número de sujetos en cada grupo

σ^2_{MA} = *Varianza de las medias* del factor A (medias de las tres *columnas)*

σ^2_{MB} = *Varianza de las medias* del factor B (medias de las cuatro *filas)*

σ^2_{MAxB} = *Varianza de las medias* de todos los subgrupos (medias de los doce subgrupos)

$\Sigma\sigma^2_{dentro}$ = Suma de las varianzas de todos los subgrupos

σ^2_t = Varianza de los totales (de todos los sujetos como si se tratara de un solo grupo)

tabla de resultados:

Origen de la variabilidad	Suma de Cuadrados	Grados de libertad	Cuadrados Medios	Razón F	p
Factor A	$SC_A = N\sigma^2_{MA}$ =	A - 1 =			
Factor B	$SC_B = N\sigma^2_{MB}$ =	B - 1 =			
AxB (interacción AxB)	$SC_{AxB} = N\sigma^2_{MAB}$ - $(SC_A + SC_B)$ =	(A-1)(B-1) =			
error, dentro de los grupos	$SC_{dentro} = n\Sigma\sigma^2_{dentro}$ =	N - k =			
Variabilidad total	$SC_{total} = N\sigma^2_t$ =	N - 1 =			

Tabla 6

3. Análisis adicionales

Vamos a distinguir dos tipos de análisis complementarios:

1º Coeficientes de asociación,

2º Contrastes entre medias, con sus correspondientes *tamaños del efecto*

Veremos además cómo hacer una representación gráfica de los resultados, que es muy útil como ayuda a la interpretación y para comunicar los resultados. No hay que hacer siempre todo lo que es posible hacer, sino lo que aporte en cada planteamiento información útil.

3.1. Coeficientes de asociación

Una razón F significativa nos indica que una determinada *fuente de varianza* (que corresponde a la variable puesta en el numerador de la razón F) se relaciona con las diferencias en la variable dependiente (la que hemos medido en los sujetos) más de lo que cabría esperar por azar: la variabilidad observada en la variable dependiente se puede atribuir a las variables del numerador de la razón F. Sin embargo no es fácil comparar entre sí dos F significativas; aunque sus valores sean muy distintos también lo son los grados de libertad.

Con frecuencia es suficiente comprobar *la proporción de varianza* atribuible a cada variable (o a su interacción si resulta significativa). Esto nos lo indica el coeficiente ω^2 o el coeficiente η^2 (sus valores oscilan entre 0 y 1).

3.1.1. Coeficiente ω^2

Las fórmulas del coeficiente ω^2 están indicadas en la tabla 7, junto con los resultados correspondientes a este ejemplo (podemos observar que el denominador es siempre el mismo).

Estimación de la *proporción de varianza* atribuible a:	Valor del coeficiente ω^2	Resultados (en este ejemplo)
Factor A (métodos)	$\omega^2_A = \dfrac{SC_A - (A-1)(CM_{dentro})}{CM_{dentro} + SC_{total}}$ [1]	$\omega^2_A = \dfrac{40-(2)(3.42)}{3.42+374} = .088$
Factor B (profesores)	$\omega^2_B = \dfrac{SC_B - (B-1)(CM_{dentro})}{CM_{dentro} + SC_{total}}$ [2]	$\omega^2_B = \dfrac{150-(3)(3.42)}{3.42+374} = .370$
La interacción (AxB)	$\omega^2_{AB} = \dfrac{SC_{AxB} - (A-1)(B-1)(CM_{dentro})}{CM_{dentro} + SC_{total}}$ [3]	$\omega^2_{AB} = \dfrac{20-(2)(3)(3.42)}{3.42+374} = -.001$

Tabla 7

Si algún valor de ω^2 es negativo se interpreta como cero.

Estas fórmulas son apropiadas solamente cuando:

1º El número de sujetos es el mismo en cada clasificación
2º Se trata de muestras independientes
3º En diseños o análisis de categorías fijas (como suele ser frecuente).

En este ejemplo podemos ver que una proporción apreciable de la varianza del factor B (los profesores) está asociada a la variable dependiente (las puntuaciones analizadas). La proporción de varianza correspondiente al factor A (métodos) es muy pequeña. Estas proporciones se comparan entre sí y se interpretan mejor que si sólo disponemos de los valores de la razón F. Si este ejemplo fuera real concluiríamos que la fuente importante de varianza (de las diferencias en lo que hayamos medido) está sobre todo en los profesores, no tanto en los diversos métodos. Ahora podríamos simplemente fijarnos en las medias de los distintos profesores (o de los métodos si fuera ése el caso).

La proporción de varianza que se puede atribuir a la interacción puede considerarse igual a cero, ya que su F correspondiente no llega a 1.

Este coeficiente ω^2 no es extrapolable; solamente se refiere a los datos analizados, y sólo puede utilizarse con las categorías fijas. Si se trata de categorías aleatorias (y si una de ellas es aleatoria también lo es la interacción a estos efectos) las fórmulas que se utilizan son las mismas del análisis de varianza para varias muestras relacionadas.

3.1.2. Coeficiente η^2

También puede utilizarse el coeficiente η^2 $$\eta^2 = \frac{SC_x}{SC_{total}} \qquad [4]$$

SC_x simboliza cualquier suma de cuadrados (de los factores y de la interacción). El valor de η^2 tiene un sesgo positivo y suele ser bastante mayor que el de ω^2 por lo que no son comparables entre sí.

En nuestro ejemplo tenemos:

$$\eta^2\,(A) = \frac{40}{374} = .106$$

$$\eta^2\,(B) = \frac{150}{374} = .40$$

$$\eta^2\,(AxB) = \frac{20}{374} = .053$$

Una variante de interés de este coeficiente es *eta al cuadrado parcial,* que nos indica la proporción de varianza asociada a un factor (o a la interacción) *neutralizando la varianza debida a otras fuentes* (otros factores o su interacción)[9]:

$$\eta^2\,parcial = \frac{SC_x}{SC_x + SC_{dentro}} \qquad [5]$$

Si deseamos conocer la proporción de varianza asociada al factor B (profesores) prescindiendo del influjo del método, tenemos:

$$\eta^2\,(B)\,parcial = \frac{150}{150+164} = .477$$

Observamos que ahora η^2 correspondiente a B (profesores) es ligeramente mayor si neutralizamos el efecto del método (como si todos fueran igualmente eficaces).

Estos coeficientes (ω^2 o η^2) se deberían utilizar rutinariamente como complemento al análisis de varianza. Resultados (valores de F) *estadísticamente significativos* son compatibles con una relación muy baja entre la variable independiente y la variable dependiente. Si la variable independiente explica menos del 5% de la varianza de la variable dependiente, se trata de una relación muy baja aunque sea *real* (no casual…).

Para valorar las magnitudes de estos coeficientes ya se dieron algunas indicaciones a propósito del análisis de varianza para varias muestras independientes; en cualquier caso siempre se pueden comparar entre sí los coeficientes dentro de un mismo planteamiento. Resultados significativos pero que explican una proporción de varianza muy baja, pueden indicar que quizás habría que refinar la metodología (si se trata de un diseño experimental); o

---

[9] Jaccard (1998:38). En el mismo lugar pueden verse otros coeficientes.

quizás haya que redefinir los conceptos o ajustarlos más. La proporción de varianza explicada en un estudio dado está afectada por la variabilidad (varianza) no controlada, debida a variables extrañas. Por esta razón más que buscar valores absolutos altos, es más útil comparar unos coeficientes con otros y comprobar cuál es la importancia relativa de estas fuentes de varianza.

3.2. Contrastes posteriores

Es frecuente limitarse a calcular los distintos coeficientes de asociación (η^2, ω^2), pero también podemos hacer los contrastes posteriores que sean de interés[10].

Vamos a distinguir dos situaciones:

1) En tablas 2x2, cuando los criterios de clasificación son solamente dos y cada uno está dividido en dos niveles;

2) En tablas nxn, cuando al menos uno de los dos factores está dividido en más de dos categorías.

1º *En tablas 2x2,* con sólo dos niveles en cada categoría.

En estos casos los contrastes posteriores son innecesarios. Si por ejemplo la razón F del factor A es estadísticamente, habrá una diferencia clara entre sus dos niveles A_1 y A_2. Lo mismo sucede con el otro factor. Si la interacción es significativa, esto quiere decir que uno de los dos niveles de un factor es mayor que el otro. En todos estos casos sólo interesa comparar dos medias entre sí, y si la razón F es significativa, ya sabemos que una media es significativamente mayor que la otra sin hacer más contrastes (sí nos puede interesar añadir el *tamaño del efecto).*

2º *En tablas nxn,* cuando tenemos al menos un criterio de clasificación (o factor) dividido en más de dos niveles.

En estos casos sí pueden hacernos falta los contrastes posteriores apropiados. En algunos casos son estos contrastes los que permiten llegar a conclusiones más interpretables: cuando tenemos más de dos niveles en alguno de los factores y la varianza correspondiente a alguno de ellos o la de la interacción es significativa, tendremos más de dos medias y será importante ver dónde está la diferencia.

En este cuadro tenemos el planteamiento clásico de un análisis de varianza, con dos factores principales (A y B); el factor A está dividido en tres niveles y el factor B en dos (tabla 8).

		Factor A			
		A_1	A_2	A_3	*medias* de B
Factor B	B_1	A_1B_1	A_2B_1	A_3B_1	$\overline{B}_1$
	B_2	A_1B_2	A_2B_2	A_3B_2	$\overline{B}_2$
	medias de A	$\overline{A}_1$	$\overline{A}_2$	$\overline{A}_3$	

Tabla 8

Con este planteamiento obtenemos cuatro varianzas o *cuadrados medios* (de los dos factores principales, A y B, de la interacción, AxB, y del término del *error* o varianza *dentro*

10 Pueden verse en monografías más especializadas, como Toothaker (1993).

de los grupos) y tres razones F al dividir las varianzas de A, B y AB por la varianza del término del error (dentro de los grupos, aleatoria). Estas razones F nos dirán si son *significativas* las diferencias debidas a A, a B y a AxB (o a una combinación entre niveles de factores).

Los contrastes posteriores que nos pueden interesar son entre dos tipos de medias:

a) Entre distintos niveles del mismo factor
b) Entre las medias del mismo nivel de un factor.

1º. Diferencias entre los diversos niveles de A o de B; es decir, diferencias *entre niveles del mismo factor*.

En el esquema anterior podríamos comparar $\overline{A}_1$ con $\overline{A}_2$, etc.

A veces estas comparaciones pueden ser innecesarias cuando sólo tenemos dos medias que contrastar. Si en el esquema anterior la razón F correspondiente al factor B es significativa, esto quiere decir, sin necesidad de más contrastes, que entre las dos medias correspondientes a B, $\overline{B}_1$ y $\overline{B}_2$, existe una diferencia estadísticamente significativa, porque sólo tenemos dos medias. En cambio si la razón F correspondiente al factor A es significativa, podemos encontrar diferencias entre $\overline{A}_1$ y $\overline{A}_2$, entre $\overline{A}_1$ y $\overline{A}_3$, y entre $\overline{A}_2$ y $\overline{A}_3$. En este caso sí necesitaremos hacer los contrastes posteriores si nos interesa saber entre qué niveles de A hay una diferencia estadísticamente significativa.

2º Si la varianza *(cuadrados medios)* correspondiente a la *interacción* es *estadísticamente significativa,* nos puede interesar comparar entre sí diversas combinaciones de AB.

Tenemos 6 combinaciones distintas de AB, de donde nos salen 15 posibles comparaciones [(6 x 5)/2], aunque ya veremos que en este caso nos interesa hacer menos comparaciones de las que es posible hacer.

3.2.1. Contraste entre las medias de niveles del mismo factor

Cuando la razón F correspondiente a uno o a los dos factores es estadísticamente significativa nos puede interesar comparar los diversos niveles de A entre sí ($\overline{A}_1$, $\overline{A}_2$, etc.). El procedimiento que exponemos es el adecuado para hacer estas comparaciones, pero no lo es para comparar una media de un nivel de A con otra de otro nivel B (y tampoco suele ser de especial interés).

Hay que advertir que *si la interacción es estadísticamente significativa,* estas comparaciones pueden no interesar. Concluir con un *este método es mejor que otro* sin añadir *pero sólo en esta condición o con este grupo,* no es decir mucho, e incluso puede inducir a error. Y es eso lo que quiere decir que la interacción es significativa: que entre los niveles de un factor hay diferencias (por ejemplo entre A_1 y A_2), pero *no en general,* sino en un nivel del otro factor (por ejemplo, A_1 y A_2 difieren en B_2 pero no en B_1).

Para comparar entre sí las medias de los distintos niveles de un factor tenemos los mismos procedimientos ya vistos para varias muestras independientes (Scheffé, Tukey, etc.), pues estamos en ese caso. Posiblemente el contraste más común es el de Tukey, y es el que recomiendan algunos autores en esta situación (Toothaker, 1990):

$$q = \frac{M_1 - M_2}{\sqrt{\frac{CM_{dentro}}{n}}} \qquad [6]$$

Podemos también calcular la *Diferencia Estadísticamente Significativa* (DHS), despejando la diferencia entre las medias en la fórmula [6].

a) n es el *número de datos (o de sujetos) que sumamos para calcular cada media* y no el número de sujetos que hay en *cada* clasificación. Naturalmente en las dos medias que contrastamos el número de sujetos es el mismo. En estos planteamientos de análisis de varianza (diseños factoriales, dos o más criterios de clasificación) el número de sujetos en cada clasificación es el mismo. Si en este ejemplo tuviéramos *cuatro* sujetos en cada clasificación (cuatro en A_1B_1 y cuatro en A_1B_2), el *n* de la fórmula [6] sería *ocho* (si estamos comparando $\overline{A}_1$ con $\overline{A}_2$).

b) El valor de q resultante lo consultamos en las *tablas* de q *(rango estudentizado)* teniendo en cuenta cuál es el *número de medias* (k) y cuántos son los *grados de libertad* aquí.

1º El *número de medias* (k) es el número de niveles del factor correspondiente. Si estamos en este ejemplo comparando entre sí las medias de A, k = 3, porque tenemos tres medias en el factor A.

2º CM_{dentro} es, como siempre en estos casos, el valor de los *cuadrados medios del error* o *dentro* de los grupos; lo que es distinto son *los grados de libertad* para consultar las tablas de q.

Los *grados de libertad* para consultar las tablas de q son igual a *N-ab*

N: es el número *total* de sujetos. Si en cada clasificación tenemos cuatro sujetos, como tenemos seis clasificaciones, aquí tendríamos N = 4 x 6 = 24 sujetos

a y b son el *número de niveles* (o subdivisiones) de los factores A y B. En este caso a = 3 y b = 2. Si tuviéramos N = 24, los *grados de libertad* serían 24 - (2x3) = 18.

En las tablas buscaremos 18 grados de libertad (si no figuran en las tablas estos grados de libertad, buscamos el número inmediatamente inferior que encontremos). En realidad, y como en otros casos semejantes, los grados de libertad son $\Sigma(n-1)$ o N - k (k número de grupos; *ab* es el número de grupos).

Vamos a verlo con un ejemplo.

En el ejemplo que nos ha servido para explicar este modelo de análisis de varianza (tabla 3) hemos visto que la F correspondiente a los métodos (factor A) es estadísticamente significativa. Tenemos tres medias de A y entre estas tres medias habrá diferencias estadísticamente significativas. Vamos a comparar la media de A_1 (= 6) con la media de A_3 (= 4).

$$q = \frac{M_{A1} - M_{A2}}{\sqrt{\frac{CM_{dentro}}{n}}} = \frac{6-4}{\sqrt{\frac{3.42}{20}}} = 4.83$$

3.42 son los *cuadrados medios dentro* (tabla 5) y 20 es el número de sujetos utilizados para calcular estas medias.

En las tablas buscaremos el valor que corresponde a k = 3 (porque tenemos tres medias de A) y grados de libertad = 48 [N-ab = 60- (3x4)]. En los tablas no aparecen los 48 grados de libertad; nos quedamos con el número inmediatamente inferior, 40.

Para k = 3 y 40 grados de libertad vemos estos valores de q: 3.44 (.05) y 4.37 (.01); en nuestro caso con q = 4.83 tenemos que p <.01, por lo que podemos afirmar la diferencia entre estas medias con *mucha seguridad;* el que la diferencia sea *grande* es una cuestión distinta (más adelante tratamos del *tamaño del efecto).* Si comparamos A_1 con A_2 o A_2 con A_3 tendremos que q = 2.42 y p > .05.

Una observación sobre procedimientos alternativos.

En textos y manuales de estadística aparecen a veces fórmulas aparentemente distintas pero que son equivalentes y nos pueden confundir.

Por ejemplo, para este mismo contraste de medias podemos encontrar esta fórmula (Toothaker, 1993):

$$t' = \frac{M_1 - M_2}{\sqrt{\frac{CM_{\text{dentro}}}{n} \text{ x } 2}} \qquad [7]$$

Aquí se calcula un valor de t'; la diferencia con la fórmula de q [6] es ese "x 2" que aparece en el denominador de [7].

En este caso tendríamos que calcular los valores críticos de t', como en ocasiones semejantes, y que son:

$$\text{Valores críticos de t'} = \frac{q}{\sqrt{2}} \qquad [8]$$

El valor de q es el que venga en las tablas (para α = .05 y gl = N - ab, como antes).

Aplicando los datos anteriores tenemos que $t' = \frac{6 - 4}{\sqrt{\frac{3.42}{20} \text{x}2}} = 3.42$

Ahora tenemos que hallar los valores críticos de t’: para p = .05, $t' = \frac{3.44}{\sqrt{2}} = 2.43$

para p = .01, $t' = \frac{4.37}{\sqrt{2}} = 3.09$

Con t’= 3.42, nuestra conclusión es la misma (p<.01). Posiblemente es más cómodo calcular directamente el valor de q y consultar las tablas.

También podemos calcular directamente el valor de la diferencia necesaria *(diferencia crítica)* para rechazar la Hipótesis Nula; la que Tukey denomina *diferencia honestamente significativa*. Para esto, en la fórmula de q, nos basta despejar la diferencia del numerador y sustituir q por el valor que venga en las tablas:

$$\textit{Diferencia crítica} = q\sqrt{\frac{CM_{\text{dentro}}}{n}} \qquad [9]$$

En este caso, para localizar el valor de q en las tablas, tanto el valor de k (número de grupos o de medias) como el de los grados de libertad son los indicados anteriormente.

3.2.2. Contraste entre las medias *dentro del mismo nivel* de un factor (cuando la interacción es significativa)

Lo expuesto hasta ahora es el procedimiento para verificar si hay diferencias significativas entre las medias de los distintos niveles (medias de columnas entre sí y medias de filas entre sí). Pero también nos pueden interesar otras comparaciones.

El comparar entre sí las distintas medias que encontramos en los diversos niveles de cualquier factor nos interesa *cuando la interacción es significativa:* un método puede ser superior a otro, pero solamente con una determinada duración, o un medicamento puede ser superior a otro, pero solamente en una determinada dosis o una determinada etapa de la enfermedad, etc.

Vamos a suponer que A_1, A_2 y A_3 son tres grupos de alumnos divididos según la edad (tres edades o cursos distintos) y B_1 y B_2 son dos actividades didácticas (tabla 9, idéntica a la tabla 8):

		Factor A (cursos)			
		1º (A_1)	2º (A_2)	3º (A_3)	*medias* de los métodos (B)
Factor B	método 1 (B_1)	A_1B_1	A_2B_1	A_3B_1	método 1 ($\overline{B}_1$)
(métodos)	método 2 (B_2)	A_1B_2	A_2B_2	A_3B_2	método 2 ($\overline{B}_2$)
	medias de los cursos (A)	$\overline{A}_1$	$\overline{A}_2$	$\overline{A}_3$	

Tabla 9

Supongamos que llegamos a esta conclusión: la razón F correspondiente al factor B (método o actividad) es significativa. Esto puede querer decir que un método, por ejemplo B_2, es mejor que el otro, B_1. *Si la interacción no es significativa,* la interpretación es que B_2 es mejor independientemente del curso (edad) de los alumnos.

Suponemos ahora que además del factor métodos (B) *también la interacción es significativa:* en este caso el concluir simplemente que B_2 es mejor que B_1 no es decir mucho: es mejor pero según con qué alumnos. En este caso nos interesará contrastar las tres medias de B_2 entre sí para ver en qué grupo o grupos es mejor el método B_2 pues B_2 está implicado en los tres niveles de A: A_1, A_2 y A_3.

Qué comparaciones interesa hacer cuando la interacción es significativa.

Como primer paso podemos inspeccionar la *representación gráfica* de los resultados (explicada en un apartado posterior) que nos aclarará la interpretación de manera intuitiva y nos dirá dónde pueden estar las comparaciones de interés. Cuando la interacción es significativa, los textos de nivel básico o intermedio suelen limitarse a recomendar la inspección estas representaciones gráficas (Oshima y McCarty, 2000)[11].

En el esquema que nos sirve de ejemplo (A con tres niveles y B con dos), si queremos comparar entre sí todas las posibles combinaciones AB, como tenemos 6 grupos distintos tendremos (6x5)/2 = 15 comparaciones posibles. Con cuatro niveles en un factor y tres en el otro tendríamos 12 grupos distintos y (12x11)/2 = 66 contrastes posibles. Es claro que son muchas comparaciones y que posiblemente muchas no tienen especial interés.

[11] Oshima y McCarty (2000) reconocen la falta de consenso entre autores sobre cómo hacer estos contrastes cuando la interacción es significativa y exponen diversos enfoques.

Por lo general no nos interesan todas las comparaciones posibles; lo que sí puede tener interés, y es lo que tratamos aquí, es *comparar solamente las medias dentro de un mismo nivel:* por ejemplo las comparaciones entre las tres B_2. Dicho de otra manera, las comparaciones que interesan son entre medias que pertenecen a la misma *fila* o a la misma *columna* (*filas* y *columnas* son aquí los distintos niveles o subclasificaciones de cada factor).

Estas medias se denominan *no confundidas (unconfounded)* porque las diferencias entre ellas se deben solamente a un factor. Si por ejemplo hay diferencias entre las medias de A_1B_1, A_2B_1, y A_3B_1, estas diferencias se deberán solamente a diferencias en A, porque mantenemos constante B_1.

En el método expuesto aquí para hacer estas comparaciones *suponemos* que:

a) Hay un idéntico número de sujetos en cada posible clasificación;
b) Sólo se van a comparar entre sí medias que pertenecen al mismo nivel de un factor.

El procedimiento va ser el mismo visto antes, pero van a ser distintos el valor de *n* que entra en la fórmula y el *número de medias* (k) para consultar las tablas (propiamente lo que cambia es el valor de k, que, como veremos, no va ser igual al número de medias, aunque *número de medias* suele ser la expresión que viene en las tablas).

Fórmulas:

Como antes, podemos hallar el valor de q y consultar las tablas de q, o podemos calcular el valor de t', y ya sabemos que los valores críticos de t' son igual a $q/\sqrt{2}$, donde q es el valor que venga en las tablas según el número de grupos (k) y los grados de libertad correspondientes.

Las fórmulas para comparar los subgrupos *de dos en dos* ya las hemos visto:

$$q = \frac{M_1 - M_2}{\sqrt{\frac{CM_{dentro}}{n}}} \quad [6]$$

los valores críticos de q los consultamos en las tablas de la manera indicada más adelante (tabla 10)

$$t' = \frac{M_1 - M_2}{\sqrt{\frac{CM_{dentro}}{n} x2}} \quad [7]$$

los valores críticos de t' son = $\frac{q}{\sqrt{2}}$

Una *observación importante* para estas dos fórmulas: *n* es el número de sujetos o de observaciones que sumamos para calcular cada media. Si en cada subclasificación hay cuatro sujetos, entonces n = 4. El principio es el mismo visto antes para contrastar las medias de los niveles entre sí (allí *n* era igual al número de sujetos que entran en la media de cada nivel).

Para consultar las tablas de q, el número de medias, como hemos indicado, no es exactamente el número de medias; el valor de k va a estar en función del tamaño de la tabla. En la práctica basta consultar la tabla 10 que hemos confeccionado para estos casos[12].

[12] La justificación de esta tabla la incluimos en el anexo V, que nos puede servir para ampliar la tabla si es necesario.

Tamaño de la tabla de dos factores	Valor de k para consultar las tablas de q
2 x 2	3
2 x 3	5
2 x 4	6
2 x 5	8
3 x 3	7
3 x 4	8
3 x 5	10
4 x 4	10
4 x 5	12
5 x 5	15

Tabla 10

3.3. El *tamaño del efecto* en los diseños factoriales

Además de verificar entre qué medias tenemos una diferencia estadísticamente significativa, podemos calcular el *tamaño del efecto* para apreciar mejor la *magnitud* de la diferencia, como hacemos en otros planteamientos. Cuando solamente tenemos dos niveles en un factor y éste es significativo, ya sabemos que entre los dos niveles existe una diferencia estadísticamente significativa, y sin hacer ningún contraste adicional podemos pasar directamente al tamaño del efecto.

Nos vamos a fijar en el tamaño del efecto cuando comparamos dos niveles del mismo factor; nos referimos fundamentalmente al factor que consideramos principal o de mayor interés[13]. La cuestión es qué desviación típica podemos poner en el denominador.

Podemos distinguir dos situaciones.

3.3.1. Cuando el factor secundario es una variable manipulada

Cuando el factor secundario o *no principal* es una variable *manipulada, introducida,* etc. por el investigador (como pueden ser *modalidades, duraciones, circunstancias,* etc.).

En estos casos podemos utilizar en el denominador del tamaño del efecto la *raíz cuadrada de los cuadrados medios dentro de los grupos* de la tabla de resultados (tabla 5). Esa desviación típica (recordemos que los *cuadrados medios* son varianzas y su raíz cuadrada una desviación típica) no contiene la variabilidad producida por ese factor secundario cuya influencia en la variable dependiente mantenemos así constante. Este factor, *en sí mismo,* no suele tener un interés teórico especial.

3.3.2. Cuando el factor secundario es una característica natural de la población

Otras veces el factor secundario o *no principal* es una característica natural de la población a la cual queremos generalizar los resultados; quizás el caso más frecuente sea el *género* (o grupo étnico, o *tipo* de muestra, como clase social, etc.).

En estos casos la desviación típica del denominador del tamaño del efecto sí debe incluir la variabilidad que se da naturalmente en ese factor. No debemos utilizar los *cuadrados medios dentro de los grupos* de la tabla de resultados (tabla 5) porque ésa es la

[13] Un tratamiento más amplio del tamaño del efecto en el contexto del análisis de varianza puede verse en Cortina y Nouri (2000).

varianza residual, la que queda una vez eliminado el influjo de los dos factores y de su interacción y nosotros queremos incluir el influjo del factor secundario.

Siempre que no queramos prescindir de la variabilidad producida por el factor secundario debemos utilizar la *raíz cuadrada de los cuadrados medios dentro de los grupos* pero la que hubiéremos obtenido haciendo un *análisis de varianza unifactorial solamente con el factor principal*. Esta desviación típica será mayor que en el caso anterior y el tamaño del efecto no quedará sobrestimado (en cualquier caso este procedimiento es más conservador y en caso de duda puede ser preferible).

Para obtener estos *cuadrados medios dentro de los grupos* (luego utilizaremos la raíz cuadrada) tenemos dos caminos.

1º Hacer un *análisis de varianza unifactorial* con sólo el factor principal.

Como solamente nos interesan los *cuadrados medios dentro,* realmente no necesitamos hacer el análisis de varianza completo, podemos calcularlos directamente. Con muestras de idéntico tamaño la fórmula es ésta:

$$\textit{Cuadrados Medios dentro} = \frac{\mathrm{n}\Sigma\sigma^2}{\mathrm{N}-\mathrm{k}}$$

n es el número de sujetos en cada nivel del factor
σ es la desviación típica de cada nivel del factor
k es el número de grupos (de niveles)

Lo vamos a ver utilizando los datos de la tabla 3; nos interesa calcular el tamaño del efecto al comparar las medias de los tres métodos. Los datos los tenemos en la tabla 11.

	A_1	A_2	A_3
n	20	20	20
Media	6	5	4
σ	2.3237	2.236	2.5099

Tabla 11

Si comparamos estos datos con los de la tabla 3, lo único que hemos tenido que hacer es calcular las desviaciones típicas de los tres niveles de A (de las columnas), y además caer en la cuenta de que n = 20 (número de sujetos en cada método).

Aplicamos la fórmula de los *cuadrados medios dentro:*

$$\mathrm{CM}_{\text{dentro}} = \frac{20(2.3237^2 + 2.236^2 + 5.099^2)}{60-3} = 5.859\text{, y } \sigma = \sqrt{5.859} = 2.42$$

Ésta es la desviación típica que utilizaremos en el denominador del *tamaño del efecto;* entre A_1 y A_3 tendremos d = (6-4)/2.42 = .826 (diferencia que podemos considerar *grande).*

2º Cálculo a partir de *la tabla de resultados* del análisis de varianza

Los *cuadrados medios dentro* de un análisis unifactorial hecho con los niveles del factor principal (que es lo que nos interesa) podemos *rescatarlos* de la *tabla de resultados* del diseño factorial que ya tenemos (tabla 5). Nos basta calcular los *cuadrados medios* sumando todas las *sumas de cuadrados* y todos los *grados de libertad* excepto los correspondientes al factor principal (A); la fórmula es por lo tanto:

$$CM_{dentro} = \frac{SC_B + SC_{AxB} + SC_{dentro}}{gl_B + gl_{AxB} + gl_{dentro}} = \frac{150 + 20 + 164}{3 + 6 + 48} = 5.859$$

Llegamos al mismo resultado con los dos procedimientos; éste segundo puede resultar más cómodo, pero es útil verificar que los dos son equivalentes.

4. La representación gráfica de los resultados como ayuda a la interpretación en los diseños factoriales

En un diseño factorial con *dos factores principales* tenemos tres posibles fuentes de variación sistemática: los dos factores principales, A y B, y su interacción.

Además de interpretar directamente los valores de F, un sencillo gráfico nos puede ayudar en la interpretación de lo que *está sucediendo* y a *comunicarlo* con más claridad.

Vamos verlo con unos ejemplos hipotéticos. Suponemos que los dos factores principales son:

Factor A: método: trabajo en grupo con dos niveles o variantes:
A_1 *con trabajo individual* antes de la discusión en grupo
A_2 *sin tarea individual* antes de la tarea grupal

Factor B: tiempo, duración: B_1 media hora
B_2 una hora

	A_1	A_2
B_1	A_1B_1	A_2B_1
B_2	A_1B_2	A_2B_2

La variable *dependiente* (la que hemos *medido* después de las actividades) puede ser satisfacción por la tarea (o nivel de participación o de aprendizaje, etc.).

En los gráficos puestos a continuación el eje de las *abcisas* (X) corresponde a uno de los dos factores principales; en este caso situamos en este eje los dos niveles del factor B (suficientemente distanciados para que el gráfico quede más claro).

El eje vertical de las *ordenadas* (Y) corresponde a la *variable dependiente* (la que hemos medido). En este eje podemos situar algunos valores representativos (en este ejemplo ficticio ponemos de 1 a 8).

En este *espacio* situamos los valores del otro factor principal, el factor A en este caso: situamos las *medias* de A_1 y A_2 correspondientes a B_1 y B_2 y unimos con una recta las medias de cada nivel (una recta representa a A_1 y la otra a A_2).

Por razones de claridad sólo ponemos dos niveles en cada factor; naturalmente podrían ser más. Si hubiera tres valores de A (tres niveles), tendríamos tres rectas. Si tuviéramos además tres niveles de B las rectas correspondientes a los niveles de A serían probablemente quebradas. En cualquier caso estos gráficos añaden claridad informativa y ayudan a la interpretación, sobre todo cuando la interacción es significativa.

En cada extremo de las líneas que representan los dos niveles de A podríamos poner sus medias exactas, pero suele ser suficiente situar los valores en el eje vertical. También se puede prescindir de los números, que se supone que además figuran en otro lugar.

En estos ejemplos vemos:

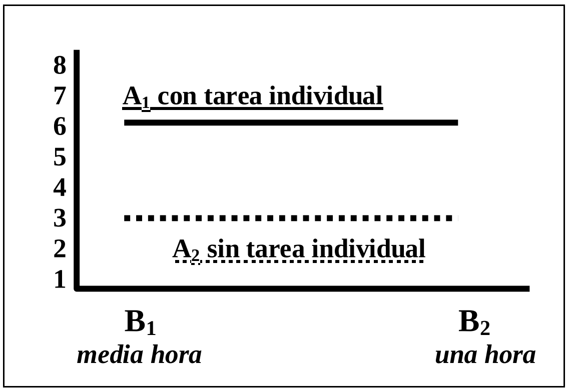

Figura 1

Figura 1. Claramente el método A_1 es superior al método A_2, cualquiera que sea la duración (factor B). Cuando las líneas son *paralelas* (o *casi* paralelas) es claro que la *interacción* no es significativa. Va a ser significativa la razón F correspondiente al factor A (A_1 es claramente mayor que A_2). En cambio la razón F correspondiente al factor B no es estadísticamente significativo; B_1 y B_2 no difieren entre sí; sus medias van a ser muy parecidas.

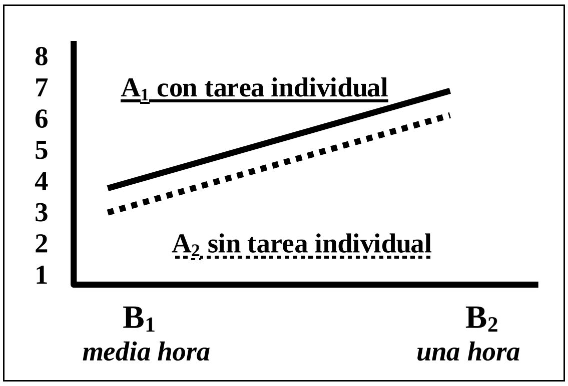

Figura 2

Figura 2. En este caso tendríamos que apenas hay diferencia entre los métodos (factor A), pero sí las hay en el factor B, duración de la actividad. B_2 es superior a B_1 independientemente de la actividad: una hora produce mejores resultados que media hora.

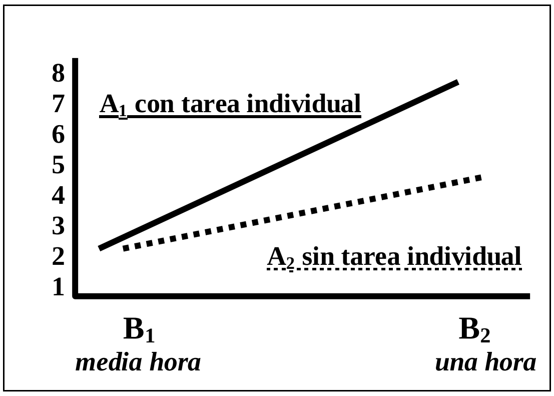

Figura 3

Figura 3. Los dos métodos son parecidos en B_1, pero con una mayor duración (B_2) uno de los métodos, A_1, es claramente superior. Aquí la *interacción* es significativa: hay una combinación AxB (A_1B_2) claramente superior a las otras ($A_1 > A_2$ en B_2). Cuando las líneas no son, más o menos, *paralelas,* la interacción está presente (poco o mucho).

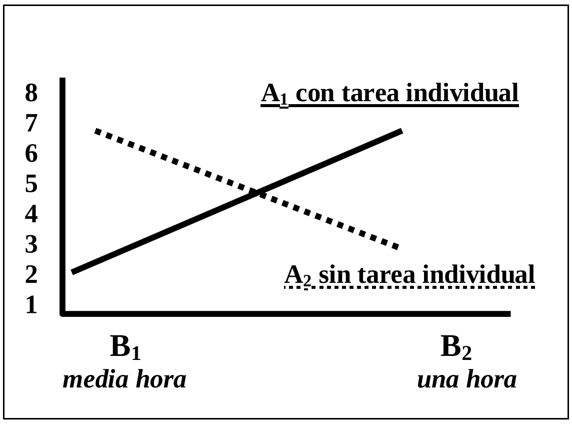

Figura 4

Figura 4. Las rectas no solamente no son paralelas, sino que se cruzan: un método es mejor con una duración (A_1 en B_2) y el otro con la otra duración (A_2 en B_1). La interacción va ser significativa e importante; la eficacia de los niveles un factor (A, los dos métodos) *está en función* de los niveles del otro (B). En un ejemplo como el sugerido por esta figura, la interacción sería la única fuente de varianza importante.

En general cuando la *interacción es significativa* tienen menor interés (o muy poco interés) los factores tomados aisladamente aunque sus valores de F sean significativos. La eficacia de un nivel de un factor (el que un método sea mejor que otro en este caso) está *mediatizada* por su relación con el otro factor. Cuando demostramos que la *interacción* es significativa (y además *grande*…) es aquí donde tenemos que centrar la interpretación. Por ejemplo en la situación hipotética de la figura 3. Vamos a suponer que la razón F correspondiente a los métodos (A) es significativa y concluimos que A_1 es mejor que A_2… esa conclusión no tiene mayor interés y además puede inducir a error si no tenemos en cuenta y dejamos claro en la interpretación que A_1 es preferible a A_2 solamente en la condición B_2

Como ya hemos advertido antes, aunque *todo esto* podemos interpretarlo a partir de los *números* (valores de F, de η^2, inspección de las medias, etc.), estos gráficos contribuyen a una mejor *comprensión* y c*omunicación* de los resultados obtenidos.

5. Análisis de varianza para diseños factoriales en EXCEL y SPSS

EXCEL. Este modelo de análisis de varianza lo tenemos en *Herramientas - Análisis de datos Análisis de Varianza de dos factores con varias muestras por grupo.*

Los datos se disponen en EXCEL como en una tabla nxn convencional, *poniendo también* los nombres o rótulos de filas y columnas. Se señala la tabla completa, incluidos los rótulos de filas y columnas.

En el cuadro de diálogo *en número de filas por muestra* se pone el número de sujetos en cada subgrupo o clasificación. El resultado son los datos descriptivos de las muestras y la tabla de resultados convencional, sin contrastes posteriores ni otros cálculos complementarios.

SPSS. Este modelo de análisis de varianza lo tenemos en *Analizar, Modelo lineal general, Univariante.* Hay que indicar al programa qué variable va actuar como dependiente y en “factores fijos” se integran las dos variables independientes que queremos estudiar.

Este análisis de varianza es complejo y conviene consultar manuales específicos (puede verse Pardo Merino y Ruíz Díaz, 2005). No presenta la *tabla de resultados* convencional por lo que, si interesa presentar esta tabla, se puede hacer fácilmente con EXCEL. El SPSS calcula los contrastes posteriores y los coeficientes *eta cuadrado* de cada efecto.

6. Referencias bibliográficas

CORTINA, JOSE M. and NOURI, HOSSSEIN (2000). *Effect Size for ANOVA Designs*. Quantitative Applications in the Social Sciences. Thousand Oaks: Sage.

ESCOTET, MIGUEL A., (1980). *Diseño multivariado en psicología y educación.* Barcelona: Ceac.

GLASS, GENE V. y STANLEY, JULIAN C., (1974). *Métodos Estadísticos Aplicados a las Ciencias Sociales.* Madrid, Prentice-Hall Internacional.

GUILFORD, J. P. y FRUCHTER, B., (1984). *Estadística aplicada a la psicología y la educación,* México: McGraw-Hill. [En Inglés: *Fundamental Statistics in Psychology and Education,* 1973. New York: McGraw-Hill].

JACCARD, JAMES (1998). *Interaction Effecs in Factorial Analysis of Variance,* Sage University Paper Series on Quantitative Applications in the Social Sciences. Thousand Oaks: Sage.

KIRK, ROGER E., (1995). *Experimental Design, Procedures for the Behavioral Sciences.* Boston: Brooks/Cole.

LINTON, MARIGOLD, GALLO JR., PHILLIP S. and LOGAN, CHERYL A., (1975). *The Practical Statistician, Simplified Handbook of Statistics*. Monterey: Brooks/Cole.

OSHIMA, T. C. and MCCARTY, FRANCES (2000). *How Should We Teach Follow-Up Tests After Significant Interaction in Factorial Analysis of Variance?* Paper presented at American Educational Research Association, New Orleans, April 2000 http://www2.gsu.edu/~epstco/aeraMain.pdf (consultado 22, Nov., 2009).

TEJEDOR, FRANCISCO JAVIER, (1984). *Análisis de varianza aplicada a la investigación en pedagogía y psicología.* Madrid: Anaya

PARDO MERINO, A. y RUÍZ DÍAZ, M.A. (2005). *Análisis de datos con SPSS 13 Base*. Madrid: McGraw Hill

TOOTHAKER, LARRY E., (1993). *Multiple Comparison Procedures,* Sage University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park: Sage.

Anexo. Análisis de Varianza (diseños factoriales) en Internet

Además de programas de ordenador como el SPSS y hojas de cálculo como EXCEL disponemos de programas en Internet.

LOWRY, RICHARD, VASSARSTATS: Web Site for Statistical Computation,Vassar College, Poughkeepsie, NY, USA; http://faculty.vassar.edu/lowry/VassarStats.html

En el menu de la izquierda en ANOVA: **Two-Way Analysis of Variance for Independent Samples.**

Se trata del análisis de varianza con *dos criterios de clasificación* cada uno dividido entre *dos* y *cuatro* niveles. Se pueden introducir los datos o se pueden copiar de una tabla. También calcula el test de Tukey para los contrastes posteriores.

En la misma dirección también están programadas algunas variantes de este análisis factorial, como **2x2x2 ANOVA for Independent Samples** (tres criterios de clasificación con dos niveles cada uno) y otros.

VADUM RANKIN Statistical Applets **2X2 Analysis of Variance** for the statistical tests for a 2x2 factorial design. http://www.assumption.edu/users/avadum/applets/applets.html ; basta introducir la media, desviación típica (de la población, dividiendo por n-1) y número de sujetos de los cuatro grupos.

Estas direcciones se pueden encontrar en JOHN C. PEZZULLO'S HOME PAGE, http://statpages.org/JCPhome.html (en Interactive Statistics Pages).