Análisis de varianza para varias muestras independientes

© Pedro Morales Vallejo
Universidad Pontificia Comillas, Madrid,
Facultad de Ciencias Humanas y Sociales
(Última revisión 15 de Septiembre de 2011).

Índice

1. Introducción: cuándo debemos utilizar el análisis de varianza.

El *análisis de varianza para muestras independientes* es el modelo de análisis de varianza más frecuente: *un factor o criterio de clasificación, dividido en dos o más niveles*;; también se denomina análisis de varianza *unifactorial*[1].

Repetimos brevemente lo ya dicho en la introducción. El análisis de varianza se utiliza cuando tenemos en el mismo planteamiento más de dos muestras independientes (de sujetos físicamente distintos en cada muestra). También se puede utilizar cuando solamente tenemos dos muestras como alternativa a la t de Student (de hecho en este caso $t^2 = F$), aunque la práctica habitual en este caso es utilizar directamente el contraste de medias (t de Student). Más adelante (apartado 7) presentamos un ejemplo metodológico del análisis de varianza aplicado a dos muestras como método alternativo al contraste de medias.

Ya hemos indicado la razón principal para utilizar el análisis de varianza en vez de la t de Student *cuando tenemos más de dos grupos en el mismo planteamiento general:* aumentan mucho las probabilidades de no aceptar (rechazar) la Hipótesis Nula de *no diferencia* cuando es verdadera (lo que denominamos error tipo I; explicado con más detalle en el anexo I). Dicho de otra manera: cuando tenemos más de dos muestras y las contrastamos entre sí con la t de Student, tenemos el riesgo de aceptar la diferencia (no aceptamos la Hipótesis Nula) cuando realmente se trata de una diferencia *normal* o, lo que es lo mismo, la probabilidad de encontrar esa diferencia es mayor de lo que especificamos en nuestro nivel de confianza[2].

2. Cuándo podemos utilizar la t de Student cuando disponemos de más de dos grupos.

Aun así cuando comparamos más de dos grupos podemos utilizar la t de Student en dos situaciones que exponemos a continuación: a) siendo más estrictos al determinar la probabilidad mínima de error para aceptar una diferencia y b) manteniendo el nivel de probabilidad habitual pero cumpliendo las tres condiciones que veremos enseguida.

2.1. Modificando los niveles de confianza: los contrastes de de Bonferroni

Podemos utilizar unos niveles de confianza más estrictos; no es ésta la práctica más común, pero de hecho hay al menos un procedimiento, el *test de Bonferroni* (o de *Dunn-Bonferroni)* en el que se utiliza la t de Student convencional pero con unos niveles de confianza más exigentes en función del número de contrastes que se van a hacer.[3]

Cuando se utiliza el test de Bonferroni se utiliza la probabilidad (p) que expresa nuestro nivel de confianza dividida por el número de comparaciones previstas, así si nuestro nivel de confianza es .05 y tenemos tres comparaciones previstas utilizaremos como nivel de confianza .05/3 = .0167; en este caso .0167 *equivale* un nivel de confianza de .05. También si conocemos la probabilidad exacta (p) podemos multiplicarla por el número de contrastes para ver si llega a .05 (así si tenemos tres contrastes y p = .0167 tendremos p = (.0167)(3) = .05).

---

[1] En inglés también suele denominarse *one-way-ANOVA;* expresión que a veces se ve traducida al español como análisis de varianza de *una vía* (en español es muy frecuente el uso de anglicismos en la terminología estadística). En EXCEL se denomina *análisis de varianza de un factor.*

[2]Esta *mayor* probabilidad de error al aceptar la diferencia entre dos muestras cuando tenemos en el *mismo planteamiento* más de dos muestras y consiguientemente más de dos contrastes entre medias, suele denominarse *error de familia (family error* en inglés); la *familia* es en este caso el conjunto de contrastes posibles entre medias.

[3] Los contrastes de Bonferroni están más explicados en el anexo II (Carlo Emilio Bonferroni elaboró su teoría sobre la probabilidad en 1936).

La crítica hecha a este contraste es que es muy *conservador;* tiene poca *potencia* para rechazar la Hipótesis Nula cuando realmente es falsa (Hancock y Klockars, 1996; Jaccard, 1998, Perneger, 1998) (diríamos que da muchos *falsos negativos*) y en definitiva la interpretación de *un* resultado depende de que el análisis se haga *en solitario* o junto con otros análisis[4].

2.2. Condiciones para utilizar la t de Student cuando tenemos más de dos muestras

Cuando tenemos *más de dos grupos* podemos utilizar la *t de Student* para comparar grupos de dos en dos si se dan <u>todas</u> estas condiciones:

1. Cuando *antes de recoger los datos tenemos hipótesis explícitas* acerca de la diferencia entre dos grupos en particular. No podemos comparar *todo con todo,* pero sí podemos utilizar el contraste de medias *normal* de la *t de Student* cuando tenemos alguna hipótesis *a priori,* formulada y justificada de manera explícita. Esta posibilidad (una hipótesis sobre la diferencia entre dos grupos en particular cuando tenemos más de dos grupos) no es lo habitual, pero tampoco hay que excluirlo rutinariamente.

2. Cuando no vamos a *combinar medias* de varios grupos en una sola media para comparar esta nueva media con otras medias. Para este tipo de comparaciones (uniendo subgrupos) tenemos los contrastes de Scheffé, posteriores al análisis de varianza, que veremos más adelante.

3. Cuando *no vamos a hacer todas las comparaciones posibles,* ya que, suponiendo que se cumplen las dos condiciones anteriores, el número de comparaciones que podemos hacer con la *t de Student* cuando hay más de dos grupos es limitado: no puede ser superior a los grados de libertad, es decir, *no puede ser superior al número de grupos menos uno*. Por ejemplo, con tres grupos (A, B y C) se podrían hacer tres comparaciones (entre A y B, entre A y C y entre B y C) pero sólo podemos hacer dos con la *t de Student,* con cuatro grupos podemos hacer tres comparaciones, etc., si *además* se cumplen las dos condiciones anteriores.

3. Análisis de varianza:

3.1. Fases del proceso

Al explicar el procedimiento seguiremos el proceso que es útil tener claro desde el comienzo y que, más o menos, seguiremos en la presentación de otros modelos de análisis de varianza, porque responde a los *pasos* lógicos que debemos seguir:

1º Explicación del procedimiento con un ejemplo resuelto paso a paso. Añadiremos una serie de aclaraciones metodológicas que nos pueden ayudar a entender mejor lo que realmente estamos haciendo.

2º Con el análisis de varianza propiamente dicho (razón F) no termina nuestro análisis. Dos tipos de cálculos posteriores completan el proceso:

a) *Contrastes posteriores,* para poder comparar las medias de dos en dos en el caso de que la razón F sea estadísticamente significativa.

---

[4] Una crítica bien razonada a los ajustes de Bonferroni puede verse en Perneger (1998): *this paper advances the view, widely held by epidemiologists, that Bonferroni adjustments are, at best, unnecessary and, at worst, deleterious to sound statistical inference... The main weakness is that the interpretation of a finding depends on the number of other tests performed …The likelihood of type II errors is also increased, so that truly important differences are deemed non-significant... Bonferroni adjustments imply that a given comparison will be interpreted differently according to how many other tests were performed.*

b) Diversos *coeficientes de asociación* que nos ayudan a aclarar la *importancia* de los resultados, y el *tamaño del efecto* (semejante al que ya conocemos y aplicamos en el contraste de dos medias).

3º Otras variantes metodológicas para hacer este mismo análisis de varianza (que o simplifican el proceso, o son útiles en determinadas circunstancias).

4º Sugerencias sobre cómo presentar los resultados.

Vamos a presentar dos procedimientos muy semejantes; uno que se basa en el cálculo de desviaciones típicas (o varianzas) de la muestra (dividiendo por N) y otro utilizando las desviaciones típicas o varianzas de la población (dividiendo por N-1).

3.2. Procedimiento utilizando desviaciones típicas; justificación

Los procedimientos que vamos a exponer en primer lugar suponen el uso de calculadoras con programación estadística, sobre todo para calcular medias y desviaciones típicas; con esos datos ya hemos visto que los procedimientos son fáciles y rápidos, sobre todo si se trata de muestras pequeñas.

Aunque disponemos de programas informáticos y de Internet (comentados en otros apartados) el proceso explicado en primer lugar calculando medias y desviaciones típicas con una simple calculadora, son útiles por estas razones.

a) Para llegar a una comprensión más cabal de lo que estamos haciendo, *sobre todo en procesos de aprendizaje*. Hay interpretaciones que no se pueden hacer correctamente sin entender bien el proceso que se capta mejor siguiéndolo paso a paso que explicando los resultados que encontramos ya hechos en cualquier programa disponible. Hay que insistir que este método puede ser muy apropiado (al menos es una buena alternativa didáctica) cuando se trata de *entender y aprender;* en otro momento lo normal será utilizar programas informáticos.

b) Cuando de diversas muestras ya disponemos del tamaño de cada muestra y de las medias y las desviaciones típicas (o es información que vemos publicada); para estos casos tenemos *también* los programas de Internet que veremos en otro apartado, pero vamos a ver que si llega el caso se pueden hacer con toda facilidad con una simple calculadora.

c) Cuando en cualquier caso se trata de muestras pequeñas y no resulta especialmente laborioso el cálculo de medias y desviaciones con una simple calculadora.

3.2.1. Procedimiento utilizando desviaciones típicas de las muestras (σ_n)

En todos los modelos de análisis de varianza hay que calcular varianzas parciales (porque descomponemos la varianza total). En el cálculo de estas varianzas, el denominador no tiene ninguna dificultad *(los grados de libertad),* pero el numerador de las varianzas, la *suma de cuadrados,* sí resulta mucho más laborioso.

Cuando no se utiliza directamente un programa de ordenador, lo más cómodo es seguir el procedimiento que utilizamos aquí para calcular las *sumas de cuadrados* y que no es el tradicional que es normal ver en los libros de texto[5].

---

[5] Adaptamos y simplificamos aquí el procedimiento propuesto por Gordon (1973). Siguiendo la misma intuición para el cálculo de las *sumas de cuadrados* (*suma de cuadrados* o numerador de la varianza = $N\sigma^2$) hemos diseñado procedimientos análogos para otros modelos de análisis de varianza que simplifican notablemente los cálculos si, como hemos indicado, se dispone de una simple calculadora con programación estadística para calcular medias y desviaciones

Aunque dispongamos de programas de ordenador y hojas de cálculo (EXCEL, SPSS), este procedimiento, sobre todo en *procesos de aprendizaje:*

1) Facilita la *comprensión* del análisis de varianza, se visualiza mejor *cómo* se descompone la varianza total y se evitan automatismos que no suponen comprender lo que se está haciendo,

2) Es *muy útil* cuando de hecho no disponemos de los datos de cada sujeto y solamente tenemos las *medias, desviaciones típicas y número de sujetos* de cada grupo. Esta situación es frecuente; son datos que podemos encontrar publicados o que podemos tener ya anotados. Los programas de ordenador (como el SPSS o EXCEL) no suelen tener previsto cómo llevar a cabo un análisis de varianza a partir *solamente* de estos datos (N, media y σ de cada grupo), aunque sí es posible hacerlo en programas disponibles en Internet (mencionamos algunos en el apartado 9).

Para entender el procedimiento hay que recordar en primer lugar la fórmula de la varianza (de la muestra):

$$\sigma^2 = \frac{\Sigma(X - M)^2}{N} \qquad [1]$$

donde podemos despejar la *suma de cuadrados:* $\Sigma(X- M)^2 = N\sigma^2$

Es decir, *la suma de cuadrados es igual a la varianza multiplicada por el número de sujetos*. Se trata de la varianza de la muestra (dividiendo por N y no por N-1). En cada planteamiento del análisis de varianza hay qué saber qué varianza hay que calcular y multiplicar luego por el número de sujetos para obtener las *sumas de cuadrados*. Esto es lo que iremos viendo en los diversos ejemplos al ir proponiendo los distintos modelos de análisis de varianza. Al dividir estas sumas de cuadrados por los *grados de libertad* obtendremos los *cuadrados medios* o varianzas parciales que van resultando al descomponer la varianza total.

Si utilizamos este sencillo y rápido procedimiento, habrá que calcular en primer lugar *(cálculos previos)* una serie de *desviaciones típicas* (que irán variando en los distintos modelos de análisis de varianza).

El procedimiento lo presentamos con un ejemplo (tabla 1)[6]. Tenemos tres grupos (tres muestras independientes, de sujetos distintos); En la terminología del análisis de varianza cada grupo es un *nivel*. Cada grupo puede representar una edad distinta, o haber seguido un método de aprendizaje distinto, o tener una procedencia distinta, etc. (es la *variable independiente*) La *variable dependiente,* en la que hemos *medido* a cada sujeto y cuyos datos hemos tabulado, puede ser un examen de conocimientos, una escala de actitudes, etc.

típicas. Huck y Malgady (1978) exponen un procedimiento similar para el análisis de varianza de los diseños factoriales. Las ventajas de estos procedimientos es que permiten resolver el análisis de varianza sin más información de cada grupo que el número de sujetos, la media y la desviación, sin necesidad de disponer de todas las puntuaciones individuales. También, con sólo estos datos (a veces los únicos disponibles), disponemos de programas de Internet para resolver el análisis de varianza.

[6] El ejemplo propuesto aquí está tomado de Downie y Heath (1981); sólo tomamos los datos (se puede comprobar que los resultados son los mismos que con el procedimiento tradicional); el procedimiento que seguimos es el que acabamos de proponer simplificando el cálculo de las *sumas de cuadrados*.

Grupo 1º	Grupo 2º	Grupo 3º
12	18	6
18	17	4
16	16	14
8	18	4
6	12	6
12	17	12
10	10	14
$n_1 =$ 7	$n_2 =$ 7	$n_3 =$ 7
$M_1 =$ 11.714	$M_2 =$ 15.428	$M_3 =$ 8.571
$\sigma_1 =$ 3.917	$\sigma_2 =$ 2.92	$\sigma_3 =$ 4.237

Tabla 1

cálculos previos

1º Calculamos la media y desviación típica de cada uno de los tres grupos;

2º Calculamos la desviación típica de los totales, de $n_1 + n_2 + n_3 = 21$, $\sigma_t = 4.669$

3º Calculamos la *desviación típica de las tres medias* (como si se tratara de un grupo de tres sujetos), $\sigma_M = 2.802$.

Una observación práctica importante que ya hemos hecho anteriormente y que conviene recordar ahora (y lo recordaremos en otras ocasiones). En la tabla 1 tenemos las puntuaciones de *todos* los sujetos y por eso podemos calcular con facilidad la media y varianza del *total* de sujetos (N = 21 en este ejemplo), pero no siempre disponemos de los datos de todos los sujetos, o son demasiados sujetos y resulta incómodo calcular la varianza total con una simple calculadora. A veces de cada grupo sólo conocemos los valores de N, la *media* y la *desviación típica* (que podemos ver, por ejemplo, publicados en un trabajo de investigación, o son datos que hemos ido conservando). A partir de estos datos podemos calcular rápidamente tanto la media (que es la *media ponderada* de los diversos grupos) como la desviación típica de *todos* los datos, que nos van a hacer falta para hacer un análisis de varianza[7].

Antes de hacer las operaciones es conveniente preparar la *tabla de resultados* (que ponemos más adelante, tabla 2) para ir colocando los resultados que vayamos obteniendo en su lugar correspondiente y proceder con orden.

Vamos a calcular tres varianzas:

1º La varianza del *total* de los datos y es esta varianza la que vamos a *descomponer* en otras dos, que son las dos siguientes;

2º La varianza que expresa la variabilidad *entre* los grupos;

3º La varianza que expresa la variabilidad *dentro* de los grupos.

En rigor, y para llevar a cabo todas las operaciones del análisis de varianza, no necesitamos calcular la varianza de todos los datos, pero es preferible hacerlo para poder verificar que la suma de las sumas de cuadrados *entre* y *dentro* de los grupos es igual a la suma de cuadrados de los totales.

Para cada varianza calculamos el numerador *(suma de cuadrados)* y el denominador *(grados de libertad):*

$$\text{Varianza} = \frac{\text{Suma.de.cuadrados}}{\text{Grados.de.libertad}}$$

Indicamos el procedimiento para muestras de idéntico y de distinto tamaño. Cuando se trata de muestras de idéntico tamaño el procedimiento es algo más sencillo; también en este caso se puede utilizar el procedimiento que indicamos para muestras de distinto tamaño, y así lo haremos como ejemplo.

[7] Las fórmula de la desviación típica de los totales calculada a partir de las medias y desviaciones de cada grupo está puesta más adelante (fórmula [32]); estas fórmulas, muy útiles ocasionalmente, están explicadas en el anexo III.

Suma de Cuadrados *(numerados de la varianza):*

1. **del total**: $N\sigma_t^2 = (21)(4.669)^2 =$ 457.79

Multiplicamos la *varianza de los totales* por el *número total de sujetos*. Ya hemos indicado que si no tenemos la varianza de los totales o no disponemos de las puntuaciones de cada sujeto, podemos calcularla a partir de los datos (N, desviación y media) de cada muestra (anexo III).

2. **Dentro** de los grupos:

a) Si los grupos son de *idéntico tamaño:*
$n\Sigma\sigma^2 = (7)[(3.917)^2 + (2.92)^2 + (4.237)^2] =$ 292.75

b) Si los grupos son de *distinto tamaño* (y también si son iguales), la fórmula es:
$\sum n\sigma^2 = [7 \times (3.917)^2] + [7 \times (2.92)^2 + [7 \times (4.237)^2] =$ 292.75

3. **Entre** los grupos:

a) si los grupos son de *idéntico tamaño:* $N\sigma_M^2 = (21)(2.802)^2 =$ 164.87
(multiplicamos la *varianza de las medias* por el número total de sujetos)

b) si los grupos son de *distinto tamaño:* $\Sigma n(M - M_t)^2$

En este caso necesitamos calcular el valor de la media total (M_t); en este ejemplo, como los grupos constan del mismo número de sujetos, la media total es igual a la *media de las medias* = 11.904, y la *suma de cuadrados entre los grupos* será:

$[7 \times (11.714 - 11.904)^2] + [7 \times (15.428 - 11.904)^2] + [7 \times (8.571 - 11.904)^2] =$ 164.94

La pequeña diferencia con respecto al resultado anterior se debe al uso de más o menos decimales y no afecta a los resultados finales del análisis de varianza.

Grados de libertad *(denominador de la varianza)*

del total:	N - 1 (N = todos los sujetos de todos los grupos)	21- 1 = 20
entre los grupos:	k - 1 (k = número de grupos)	3 -1 = 2
dentro de los grupos:	N - k (N menos número de grupos o Σ(n-1)	21 -3 = 18

Los resultados del análisis de varianza se presentan en una tabla semejante a la tabla 2; es la tabla clásica para presentar los resultados del análisis de varianza.

Origen de la variación	Suma de Cuadrados (SC) *(numerador de la varianza)*	Grados de libertad (gl) *(denominador de la varianza)*	Cuadrados Medios $CM = \frac{SC}{gl}$ *(varianza)*	$F = \frac{CM_{entre}}{CM_{dentro}}$
Diferencias **entre** los grupos	$SC_{entre} = 164.87$	2	$CM_{entre} =$ 82.435	5.069 ($p < .05$)
Diferencias entre los sujetos **dentro** de los grupos	$SC_{dentro} =$ 292.75	18	$CM_{dentro} =$ 16.26	
Variabilidad **total**	$SC_{total} = 457.79$	20		

Tabla 2: *Tabla de resultados del análisis de varianza*

La *suma de cuadrados del total* nos sirve para comprobar que no hay errores, pues debe ser igual a la *suma* de cuadrados *entre* los grupos más la suma de cuadrados *dentro* de los grupos. En este caso 164.87 + 292.75 = 457.62. Las sumas no siempre *cuadran* porque solemos redondear decimales, pero la discrepancia debe ser pequeña. En los grados de libertad el resultado debe ser exacto: 2+ 18 = 20 (= N-1).

El valor de F que encontramos en las tablas para 2 grados de libertad de la varianza mayor (k-1) y 18 de la varianza menor (N - k) es de 3.55; si alcanzamos o superamos este valor, podemos afirmar que la diferencia entre ambas varianzas está por encima de lo puramente aleatorio, con una probabilidad de error inferior al 5% ó p< .05 (si hubiéramos superado el valor de 6.01 la probabilidad de error hubiera sido inferior al 1%, p <.01)[8].

La variabilidad *entre* los grupos (entre las *medias)* es en este caso significativamente más alta que la variabilidad *dentro* de los grupos; podemos por lo tanto concluir que entre las medias existen diferencias significativas: la *variabilidad total* se explica más por las diferencias entre las medias (entre los grupos) que por las diferencias intra-grupales. Dicho de otra manera: afirmamos que dos grupos son distintos cuando sus medias difieren entre sí más que los sujetos entre sí.

Un resumen del procedimiento lo tenemos en la tabla 3 (*para muestras de idéntico tamaño*):

Origen de la variación	***Numerador* de la varianza o Suma de Cuadrados *(SC)***		***Denominador* de la varianza (Grados de libertad)**	**Varianza o *Cuadrados Medios* (CM)**	**F**
***Entre* los grupos (varianza parcial: expresa las diferencias entre los grupos)**	***N (número total de sujetos) por la varianza de las medias***	$N\sigma^2_{medias}$	**Número de grupos (k) *menos* uno k - 1**	$\frac{\text{Suma de Cuadrados entre}}{\text{grados de libertad}}$	$\frac{CM_{entre}}{CM_{dentro}}$
***Dentro* de los grupos (varianza parcial: expresa las diferencias entre los sujetos)**	***n (número de sujetos en cada grupo) por la suma de las varianzas de los grupos***	$n\Sigma\sigma^2_{grupos}$	**Número *total* de sujetos menos *número de grupos* N - k**	$\frac{\text{Suma de Cuadrados dentro}}{\text{grados de libertad}}$	
Varianza *total*	***N (número total de sujetos) por la varianza total (de todos los sujetos)***	$N\sigma^2_{de\ todos}$	**Número *total* de sujetos *menos uno* N - 1**		

Tabla 3

Las desviaciones son de la muestra (dividiendo por N) y suponemos muestras de idéntico tamaño. Cuando los grupos son de *distinto tamaño* las *Sumas de Cuadrados* (numerador de la varianza) se calculan así:

Suma de Cuadrados entre los grupos: $\Sigma n(M_n - M_t)^2$: en cada grupo: *número* de sujetos del grupo *por* (Media del grupo *menos* Media total)^2

---

[8] El *valor exacto* de la probabilidad asociada a cualquier valor de F puede verse fácilmente en varias direcciones de Internet puestas en al apartado nº 9.

Suma de Cuadrados dentro de los grupos:$\Sigma n\sigma^2$: en cada grupo: *número* de sujetos del grupo *por* la varianza.

3.2.2. Procedimiento utilizando desviaciones típicas de la población (σ_{n-1})

En el procedimiento, tal como lo hemos expuesto, hemos utilizado las desviaciones de las muestras. El ver las variantes si utilizamos las desviaciones o varianzas de la población (dividiendo por N-1 la suma de cuadrados) nos pueden ayudar a comprender mejor lo que estamos haciendo, y resulta más cómodo si lo que tenemos calculado de las distintas muestras son las desviaciones de la población[9].

a*) Varianza o cuadrados medios entre grupos*

Al menos de manera intuitiva podemos comprender que la diversidad *entre* los grupos *algo tendrá que ver* con la desviación típica o la *varianza de sus medias*, tal como hemos hecho para calcular la *suma de cuadrados entre los grupos*. Multiplicamos esta varianza por N porque todos los sujetos intervienen en el cálculo de la media.

Podemos quizás verlo con más claridad de esta manera: lo que calculamos no es la varianza de las medias, sino la *varianza de toda la muestra* pero *asignando a cada sujeto la media de su grupo, como si no hubiera diferencias dentro de cada grupo* (lo que sucede es que cuando los grupos son de idéntico tamaño nos basta calcular la desviación típica de las medias; como si en cada muestra n = 1). Simplemente estamos calculando la varianza total pero prescindiendo de la diversidad dentro de cada grupo (ésta la tendremos en cuenta en la suma de cuadrados *dentro* de los grupos).

Cuando los grupos son de idéntico tamaño, si calculamos directamente la varianza de las medias pero dividiendo por N-1 (varianza de la población) y ponderamos esa varianza por el número de sujetos que hay en cada grupo, ya tenemos los *cuadrados medios entre los grupos*.

En nuestro ejemplo (muestras de idéntico tamaño):

Varianza de las tres medias (dividiendo por N-1): $\sigma^2_{n-1} = 3.432^2 = 11.78$

Cuadrados Medios entre los grupos: $(n)(\sigma^2_{n-1}) = (7)(11.78) = 82.46$

La variabilidad *entre los grupos* (representados por sus medias) está lógicamente relacionada con la *varianza de las medias.*

La *Suma de Cuadrados entre grupos* es simplemente SC = (CM)(Grados de libertad); nos puede interesar para completar la tabla de resultados convencional y para calcular el coeficiente η^2 que veremos después.

Si las muestras son *de tamaño distinto* tendremos que calcular en primer lugar la *media total* (*media de las medias* ponderadas por el número de sujetos).

En este caso (podemos hacerlo aunque las muestras sean de idéntico tamaño) la media total será:

$$\text{Media}_{\text{total}} = \frac{(7)(11.714) + (7)(15.428) + (7)(8.571)}{21} = 11.90$$

[9] Además en los programas de Internet para hacer el análisis de varianza (reseñados en el apartado 9) las desviaciones típicas requeridas suelen ser las de la población.

La *Suma de Cuadrados* será la que hubiéramos obtenido si todos los sujetos de cada grupo tuvieran idéntica puntuación igual a la media de su grupo. En cada grupo por lo tanto calculamos la diferencia entre la media del grupo y la media total, la elevamos al cuadro y multiplicamos por el número de sujetos del grupo: ésta es la contribución de cada grupo a la suma de cuadrados total.

La Suma de cuadrados entre grupos de tamaño desigual será por lo tanto

$SC_{entre} = \Sigma n(M\text{-}M_{total})$ donde n es el número de sujetos de cada grupo.

En nuestro ejemplo tendremos:

$$SC_{entre} = (7)(11.714\text{-}11.9)^2 + (7)(15.428\text{-}11.9)^2 + (7)(8.571\text{-}11.9)^2 = 164.9$$

Los grados de libertad son los mismos, número de grupos menos uno.

b*) Varianza o cuadrados medios dentro de los grupos*

Por lo que respecta a la varianza *dentro* de los grupos, ésta es simplemente (y *obviamente)* la combinación (la *media*) de las varianzas de todos los grupos[10].

El que el denominador *(grados de libertad)* sea N - k también puede verse con facilidad: el denominador de cada varianza es n -1 (número de sujetos en cada grupo menos 1), luego el denominador de la combinación de todas las varianzas será el número total de sujetos (suma de todos los *n)* menos el número de grupos.

Este procedimiento para calcular los cuadrados medios *dentro* de los grupos (o varianza *dentro* de los grupos) es fácil de ver:

La varianza de un solo grupo, como expresión *descriptiva* de su diversidad, es:

$\sigma^2_n = \dfrac{\Sigma(X - M)^2}{n}$ y ya hemos visto que la suma de cuadrados será $\Sigma(X\text{-}M)^2 = n\sigma^2$,

Esta expresión de la *suma de cuadrados* (o *numerador* de la varianza) $n\sigma^2$ nos es muy útil para facilitar los cálculos, como ya hemos ido viendo.

La varianza de la *población* (el subíndice n-1 expresa ahora que dividimos la *suma de cuadrados* por n –1 al calcular la desviación típica) estimada a partir de esta misma muestra será por lo tanto:

$$\sigma^2_{n\text{-}1} = \frac{n\sigma^2_n}{n-1} \qquad [2]$$

Si *combinamos* las varianzas de dos grupos (1 y 2) sumando sus numeradores y denominadores tendremos que:

$$\sigma^2_{(n\text{-}1)de\ 1+2} = \frac{n_1\sigma^2_{n.1} + n_2\sigma^2_{n.2}}{(n_1 - 1) + (n_2 - 1)} \qquad [3]$$

Y ampliando la fórmula a un número indefinido de muestras tenemos los *cuadrados medios o varianza dentro* de los grupos (k = número de grupos):

10 Esta combinación de varianzas o varianza común es la misma que se utiliza en el contraste de medias para calcular el *tamaño del efecto* propuesto por Cohen (1988); en este caso se trata de la desviación típica.

Si los grupos son del mismo tamaño: $CM_{dentro} = \frac{n\Sigma\sigma_n^2}{N-k}$ [4]

Si los grupos son de distinto tamaño: $CM_{dentro} = \frac{\Sigma n\sigma_n^2}{N-k}$ [5]

Otra manera de expresar lo mismo es ésta: la variabilidad total *dentro* de los grupos *(cuadrados medios dentro)* viene dada por la *varianza media de los grupos,* calculando estas varianzas dividiendo por N-1 (se trata de la estimación de la *varianza media* de las poblaciones representadas por esas muestras).

Cuando los grupos son de *idéntico tamaño* se ve muy fácilmente. Calculamos con los datos de nuestro ejemplo (tabla 1) las desviaciones típicas de las poblaciones representadas por los tres grupos:

	Grupo 1º	Grupo 2º	Grupo 3º
σ_{n-1} =	4.2341	3.1547	4.5773

La *varianza media* será = $\frac{\Sigma\sigma_{n-1}^2}{k} = \frac{4.2341^2 + 3.1547^2 + 4.5773^2}{3} = 16.27$

Tenemos el mismo resultado obtenido antes *(cuadrados medios dentro de los grupos,* tabla 2). Esto podemos hacerlo siempre que tengamos muestras independientes de idéntico tamaño.

Si las muestras son de *distinto tamaño* se trata de una media *ponderada* por el número de sujetos: para calcular esta varianza media habrá que 1º multiplicar cada varianza por su *n,* y 2º dividir estos productos por N (no por k). Vamos a hacerlo con los mismos datos, ya que los procedimientos para muestras de distinto tamaño son también válidos cuando son del mismo tamaño.

Varianza media = $\frac{\Sigma n\sigma_{n-1}^2}{N} = \frac{(7 \text{ x } 4.2341^2) + (7\text{x } 3.1547^2) + (7\text{x } 4.5773^2)}{21} = 16.27$

En definitiva, lo que hacemos con la razón F es *básicamente* comparar la *varianza de las medias* (= variabilidad *entre*, multiplicada por el número de sujetos pues todos intervienen en la media) con la *varianza media de los grupos* (variabilidad *dentro).*

En el caso más sencillo de varias muestras independientes de idéntico tamaño (= n) podemos expresar así lo que hacemos:

$$F = \frac{\text{diferencias o variabilidad entre las medias}}{\text{diferencias o diversidad de los sujetos dentro de sus grupos}}$$

$$= \frac{(n)(\text{varianza de las medias } \sigma_{n-1}^2)}{\text{media de las varianzas de los grupos } (= \Sigma\sigma_{n-1}^2 / k)} \quad [6]$$

Esta última expresión clarifica el análisis de varianza tanto conceptual como metodológicamente y nos ofrece un procedimiento alternativo para llegar directamente a la razón F, que es lo que nos interesa. Más adelante ofrecemos un ejemplo resuelto siguiendo literalmente esta fórmula. En el numerador tenemos la varianza total prescindiendo de las diferencias dentro de cada grupo (como si todos los sujetos tuvieran una puntuación igual a la media de su grupo), y en el denominador tenemos la varianza total pero prescindiendo de las diferencias entre las medias (entre los grupos).

Habitualmente seguiremos el procedimiento explicado en primer lugar porque suponemos que calculamos las desviaciones típicas (o varianzas) dividiendo por N, ya que es el dato *descriptivo* que solemos calcular rutinariamente o que con más frecuencia encontramos ya publicado, y que además nos permite completar con facilidad la *tabla de resultados* (con las sumas de cuadrados) que es habitual presentar cuando se lleva a cabo un análisis de varianza.

3.2.3. Procedimiento alternativo de análisis de varianza para varias muestras independientes a) de idéntico tamaño y b) utilizando las desviaciones de la población (σ_{n-1})

Cuando tenemos varias muestras independientes y de *idéntico tamaño,* y disponemos de una calculadora con programación estadística, el análisis de varianza puede quedar muy simplificado, yendo directamente al cálculo de los cuadrados medios *entre* los grupos y *dentro* de los grupos y de la razón F.

El procedimiento, utilizando las desviaciones de la población (σ_{n-1}) en vez de las desviaciones de las muestras (σ_n), ya está comentado antes (apartado 3.2, fórmula [6]), pero ahora lo exponemos con un ejemplo resuelto (tabla 4).

	Grupo 1º	Grupo 2º	Grupo 3	*Desviación típica de las medias* (dividiendo por N -1)	*Varianza de las medias*
n = 7	12 18 16 8 6 12 10	18 17 16 18 12 17 10	6 4 14 4 6 12 14	σ_{n-1}	$= \boxed{A}$ σ^2_{n-1}
M	11.714	15.428	8.571	3.4324	11.7817
σ_n	3.917	2.921	4.237	Estas desviaciones típicas (dividiendo por N) son útiles como dato descriptivo	
σ_{n-1}	4.231	3.1547	4.577		
σ^2_{n-1}	17.904	9.952	20.952	*media de las varianzas* (n-1) $= \frac{\Sigma\sigma^2_{n-1}}{k} =$ 16.267	$= \boxed{B}$

Tabla 4

$$F = \frac{\text{Cuadrados Medios } \textit{entre} \text{ grupos}}{\text{Cuadrados Medios } \textit{dentro} \text{ de los grupos}} = \frac{(n)(A)}{B} = \frac{(7)(11.7817)}{16.267} = \frac{82.4719}{16.267} = 5.069$$

Con una calculadora con programación estadística hacemos los cálculos enmarcados en una doble línea con toda facilidad.

1º Calculamos en cada grupo la *media* y las *dos desviaciones típicas;* la desviación típica *normal* (de la muestra, dividiendo por N), como dato descriptivo útil, y la desviación típica dividiendo por N -1 que elevamos al cuadrado directamente (nos interesan las varianzas), y que es lo que anotamos (no hay necesidad de anotar el valor de σ_{n-1} sin elevarlo al cuadrado).

Si tenemos ya calculados los valores de σ_n (y esto es normal que suceda) y no los de σ_{n-1}, pasamos de unos a otros con facilidad [2] (el N es el de cada grupo, 7 en este ejemplo):

$$\sigma^2_{n-1} = \frac{N\sigma^2_n}{N-1}$$

2º Calculamos A: introducimos las tres medias, calculamos *su* σ_{n-1} y lo elevamos al cuadrado. Este valor *(varianza de las medias),* multiplicado por el número de sujetos que hay en *cada* grupo (n), son los Cuadrados Medios *entre* los grupos.

3º Calculamos B: introducimos las varianzas de los grupos y calculamos su media; esta *media de las varianzas* (las que hemos calculado dividiendo por N -1) es el valor de los Cuadrados Medios *dentro* de los grupos.

4º Aplicamos la fórmula de la razón F tal como la hemos puesto antes.

Este método también puede servir simplemente como comprobación. Al presentar los resultados conviene poner una tabla donde aparezcan al menos los Cuadrados Medios, o varianzas, *entre* y *dentro* de los grupos.

Si queremos presentar una *tabla convencional,* donde aparezcan también las *Sumas de Cuadrados* y los *grados de libertad* (con este sistema calculamos directamente los Cuadrados Medios), podemos calcular las Sumas de Cuadrados a partir de estos resultados (Suma de Cuadrados = Cuadrados Medios *por* grados de libertad).

4. Cálculos posteriores

4.1. Contrastes posteriores entre las medias

Un valor de F significativo nos indica que hay diferencias entre las medias por encima de lo aleatorio, o que las diferencias entre las medias (entre los grupos) son mayores que las diferencias entre los sujetos dentro de los grupos, pero no nos dice entre qué medias está o están las diferencias. Para comparar las muestras de dos en dos hay diversos métodos.

En estos casos *no utilizamos la t de Student,* comparando cada media con todas las demás. Si lo hacemos podemos encontrarnos con valores de t significativos, al nivel de confianza que utilicemos, que sin embargo pueden ser casuales; al hacer varias comparaciones de dos en dos aumenta la probabilidad de error. Cuando en el mismo planteamiento tenemos más de dos medias, se utilizan otros tipos de contrastes más rigurosos y apropiados. Estas comparaciones posteriores no se hacen si el valor de F no es significativo.

A pesar de lo que acabamos decir, y como ya indicamos antes, se puede utilizar la t de Student convencional para comparar dos medias en el caso de que tengamos *hipótesis previas y justificadas* acerca de alguna determinada diferencia entre dos medias. Esto sin embargo no suele ser lo frecuente cuando planteamos un análisis de varianza en el que simultáneamente analizamos más de dos muestras.

Sobre estos contrastes debemos tener en cuenta inicialmente:

a) Que existen bastantes procedimientos para hacer estas comparaciones posteriores[11] y que no todos son igualmente válidos, aconsejables o convenientes en todos los casos.

b) Que es frecuente que los programas de ordenador (como el SPSS) nos calculen varios de estos contrastes hechos según distintos métodos, pero el hecho de que los dé rutinariamente un programa de ordenador no quiere decir que dé lo mismo uno que otro.

[11] Según Kirk (1995) actualmente se utilizan más de 30 contrastes; este mismo autor describe 22 procedimientos.

c) Consecuentemente para valorar estos contrastes y escoger el más apropiado es conveniente examinar las peculiaridades de cada uno; aquí expondremos algunos de los más utilizados[12].

Puede llamar la atención el hecho de que haya muchos tipos de contrastes posteriores, pero tiene su explicación. Con estos contrastes se busca controlar y minimizar el error Tipo I (el que se comete al rechazar[13] la Hipótesis Nula de *no diferencia* cuando es verdadera y habría que haber aceptado la *no diferencia),* y hay diversos métodos porque se parte de diversos presupuestos sobre el número y tipos de comparaciones que se van a hacer como se irá indicando en cada caso:

a) Algún procedimiento (como el de Scheffé) supone que se pueden hacer múltiples comparaciones, todas las posibles, incluso combinando medias entre sí, lo que da lugar a un mayor número de contrastes posibles.

b) Con otros procedimientos (como el de Tukey) se supone que sólo se van a comparar las medias de dos en dos (y sin combinar medias entre sí).

c) Otros procedimientos (como el de Dunnett) suponen un número de comparaciones todavía más limitado, como la de varios grupos experimentales con uno solo grupo de control, pero no el contraste de los grupos experimentales entre sí.

Aquí nos limitamos a exponer algunos de los más útiles y frecuentes; puede ser suficiente atenerse a estos métodos; cada uno tiene sus peculiaridades y utilidad específica[14].

Los contrastes que exponemos son los siguientes:

1. El contraste de Scheffé. Cuando las muestras son de tamaño desigual o cuando no hay ningún tipo de hipótesis previa y se *exploran* diferencias por curiosidad, lo más seguro es atenerse al contraste de Scheffé.

2. El contraste de Tukey. Cuando las muestras son de *idéntico tamaño* (o *casi* iguales como veremos), el contraste de Tukey también es útil, sencillo y muy utilizado (y en este caso, muestras de idéntico tamaño, es preferible al contraste de Scheffé).

Estos dos contrastes (Scheffé y Tukey) son probablemente los que con más frecuencia se encuentran en los libros de texto. Aquí los tratamos con mayor amplitud y con ejemplos resueltos, aunque antes de decidirse por un contraste en particular conviene repasar la información disponible sobre otros contrastes.

3. Algunos contrastes que son variantes de Tukey en situaciones específicas como son Tukey-Kramer (para muestras de tamaño distinto), Games-Howell (para muestras de tamaño distinto y varianzas desiguales) y Newman-Keuls (para muestras iguales y más liberal que el contraste original de Tukey). Estos contrastes tienen otra tabla de probabilidades (distribución de q o *rango estudentizado*).

4. El contraste de Fisher (LSD, *Least Significant Difference*), una variante de la t de Student y más liberal que los anteriores.

---

[12] Puede ser útil además consultar alguna obra especializada, como Klockars y Sax (1986) y Toothaker (1993).

[13] Hablando con propiedad habría que decir *no aceptar* en vez de la expresión habitual *rechazar*.

[14] Los contrastes de Bonferroni no los tratamos aquí; nos hemos referido a ellos al comienzo y están comentados en el Anexo II; tampoco fueron pensados específicamente como contrastes posteriores al análisis de varianza. También están programados en el SPSS.

5. El contraste de Dunnett por su utilidad específica, cuando lo que nos interesa es *comparar varias muestras experimentales con otra de control* (no las muestras experimentales entre sí).

5. Sin exponerlos en detalle introduciremos los denominados *contrastes ortogonales*[15].

Veremos que todas las fórmulas de los contrastes posteriores se parecen mucho a la de la de *t de Student*. En el numerador tenemos siempre una diferencia entre dos medias; en el denominador de la *t de Student* tenemos las dos varianzas de los dos grupos que se comparan y en estas fórmulas lo que tenemos en su lugar son los *cuadrados medios dentro de los grupos* que no son otra cosa que la *varianza media* de *todas* las varianzas de las muestras que entran en el planteamiento, y no solamente las de las dos muestras que comparamos en cada caso. Otra diferencia con respecto a la *t de Student* está en las *tablas* que hay que consultar para ver las probabilidades asociadas al valor obtenido que salvo en algún caso (como en el contraste de Fisher) son distintas.

4.1.1. *Contraste de Scheffé*

Es uno de los contrastes más utilizado; no suele faltar en los textos de estadística y a veces es el único que se explica; por estas razones merece un comentario de cierta amplitud. A pesar de su popularidad, y como advierten algunos autores (Klockars y Sax, 1986, entre otros), *no es necesariamente el mejor método en todas las situaciones* (está programado en el SPSS).

a) Valoración general

1) Se trata de un método *seguro,* que se puede aplicar siempre. Es válido para todo tipo de contrastes; se pueden comparar las medias de dos en dos, o una media con una combinación lineal de otras, etc., y las muestras pueden ser tanto de idéntico tamaño como de tamaño desigual. La *flexibilidad* de este contraste lo ha hecho muy popular, aunque esta flexibilidad tiene un precio; como iremos viendo su *rigor* puede ser excesivo.

2) Es un método más fiable y más *seguro* cuando *se violan los supuestos* de normalidad y de homogeneidad de varianzas, y con mayor razón cuando las *muestras son de tamaño muy desigual*. Esta es una buena razón para utilizarlo cuando dudamos de la legitimidad del análisis de varianza porque no se cumplen los requisitos previos.

3) A pesar de la aplicabilidad tan general de este procedimiento (prácticamente siempre es válido aunque no siempre es el mejor o el más recomendable, como vamos indicando) conviene consultar otras posibilidades (que iremos mencionando) antes de aplicar rutinariamente el contraste de Scheffé, sobre todo cuando tenemos hipótesis muy especificas, las muestras son de idéntico tamaño o cuando vamos a hacer menos comparaciones de las posibles.

4) El problema con este método es que es *muy conservador,* y se puede aceptar la Hipótesis Nula cuando podríamos rechazarla legítimamente con otros procedimientos. En términos más propios, se trata de un contraste con poco *poder* (se puede aceptar la Hipótesis Nula de *no diferencia* cuando es falsa y podríamos haber afirmado la diferencia). Es un método muy *blindado* contra el error denominado Tipo I (consiste en rechazar o no aceptar la

---

[15] Todos estos contrastes pueden verse en numerosos textos; también en Internet se encuentran con facilidad expuestos y valorados los diversos contrastes, por ejemplo en Lane, David M. (1993-2007). *HyperStat Online Statistics Textbook* (en Introduction to Between-Subjects ANOVA) y en Dallal (2001) (puede verse además el apartado 9, sobre análisis de varianza en Internet).

Hipótesis Nula cuando es verdadera; este *error* se comete cuando aceptamos la diferencia y en realidad no la hay; las medias proceden de la misma población). Por esto decimos que se trata de un método muy *conservador*: esto significa que *cuesta más* rechazar la Hipótesis Nula.

5) Precisamente porque es un procedimiento muy conservador es especialmente útil cuando *no hay hipótesis previas* y simplemente *exploramos* diferencias entre grupos o entre *combinaciones* de grupos. Si no aceptamos la Hipótesis Nula (y afirmamos que sí hay diferencias entre las medias), no nos equivocamos. Es un buen procedimiento cuando comparamos grupos por mera curiosidad.

6) Es un procedimiento adecuado para comparar medias de grupos de *tamaño distinto;* esto hace que con frecuencia sea muy utilizado; en casi todos los otros métodos se supone que las muestras son de idéntico tamaño. Aun así, cuando las muestras son de tamaño distinto, disponemos de otras alternativas que no conviene pasar por alto y que iremos mencionando o explicando[16].

7) Con este procedimiento se puede hacer algo que frecuentemente puede ser de interés: *combinar varias muestras* en una sola para comparar la nueva media con la de otras muestras. Si tenemos, por ejemplo, tres grupos, se puede comparar cada media con las otras dos, pero también podemos unir dos medias para compararla con una tercera.

Estas *combinaciones* de muestras (y medias) pueden ser de mucho interés, porque con frecuencia encontramos que algunas muestras participan de una característica común y tiene sentido unirlas para comparar la nueva muestra (combinación de varias muestras) con una tercera. Precisamente porque este procedimiento está diseñado para hacer múltiples comparaciones (incluso combinando medias entre sí) es un método más conservador que los demás.

Esta posibilidad de combinar muestras (posibilidad a veces poco aprovechada) permite ampliar los análisis y enriquecer la interpretación. Además combinando medias *aumentamos el número de sujetos* y se rechaza con más facilidad la hipótesis nula.

8) Como es un método muy riguroso, sucede a veces que la razón F es significativa, y a ese mismo nivel de confianza no encontramos diferencias significativas entre las medias. Cuando la F es significativa y ninguna comparación posterior lo es, el investigador suele quedarse a veces *perplejo,* y pensando dónde se habrá equivocado; por eso conviene conocer de antemano esta posibilidad.

El que la razón F sea significativa y luego al comparar las medias de dos en dos no encontremos diferencias significativas puede suceder aunque no es lo habitual. Realmente con una F significativa encontraremos al menos una diferencia significativa entre dos medias, pero no necesariamente entre las medias de dos grupos, sino entre *combinaciones de medias* que pueden no tener un interés específico[17]. Cuando encontramos una razón F significativa y

---

[16] Contrastes posteriores que admiten muestras de tamaño desigual son al menos los de Fisher, Tukey-Kramer y Games-Howell; el contraste de Tukey y otros también son válidos cuando las muestras son de tamaño *ligeramente* distinto; en estos casos el número de sujetos que se utiliza (porque *entra* en las fórmulas) es la *media armónica.*

[17] La posibilidad de obtener una F significativa y después no obtener diferencias significativas al comparar los grupos entre sí es una posibilidad no frecuente en los demás contrastes, y lo mismo sucede a la inversa: puede haber un contraste significativo y una F no significativa. Esta posibilidad puede verse discutida en Hancock y Klockars (1996) y en Hancock y Klockars (1998: *Scheffe's test which is commonly used to conduct post hoc contrasts among* k *group means, is unnecessarily conservative because it guards against an infinite number of potential post hoc contrasts when only a small set would ever be of interest to a researcher*). La prueba inicial de F (en definitiva el *análisis de varianza)* ofrece una protección

luego no vemos diferencias estadísticamente significativas entre los grupos, una buena sugerencia es acudir a los contrastes de Games y Howell (explicados después) válidos para muestras con tamaños y varianzas desiguales (programado en el SPSS).

9) Por tratarse de un método *muy conservador,* hay autores que sugieren o recomiendan utilizar con el contraste de Scheffé un nivel de confianza *más liberal,* de $\alpha = .10$ en vez del usual $\alpha = .05$[18] En general en todos los casos lo más recomendable es indicar la probabilidad exacta ($p = .03$, $p = .006$…) en vez de los usuales $p < .05$ ó $p< .01$

b) Procedimiento

Para mayor claridad, y dada la popularidad (y los problemas) de este contraste distinguimos tres situaciones (aunque realmente se trata en todos los casos de lo mismo).

1) Para contrastar las medias de dos muestras

Esto es lo que hacen rutinariamente los programas como el SPSS.

Damos los siguientes pasos:

1º Calculamos este valor de t':

$$t' = \frac{|M_1 - M_2|}{\sqrt{CM_{dentro}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad [7]$$

Por ejemplo, entre los grupos 1º y 2º tendríamos $t' = \dfrac{|11.714 - 15.4281|}{\sqrt{(16.26)(\frac{1}{7} + \frac{1}{7})}} = 1.723$

Cuando los grupos son de igual tamaño se simplifica el cálculo porque el denominador es siempre el mismo. Los valores de t' para los diversos pares de grupos de este ejemplo son:

entre el 1º y el 2º, t' = 1.723
entre el 1º y el 3º, t' = 1.458
entre el 2º y el 3º, t' = 3.181

2º A partir de a) número de grupos (k) y b) de los valores de F que *vienen en las tablas,* y ya consultados en el análisis de varianza previo, calculamos los valores significativos de t' *(construimos nuestras propias tablas)* mediante la fórmula [8]:

$$t' = \sqrt{(k - 1)F} \qquad [8]$$

En nuestro caso (2 y 18 grados de libertad) que encontramos en las tablas son:

---

contra el error tipo I (rechazar (o no aceptar) la Hipótesis Nula cuando es verdadera) que puede resultar excesiva y no faltan autores que recomiendan pasar directamente a los contrastes posteriores cuando hay hipótesis experimentales claras (Jaccard, 1998:28, 67). El encontrar una F significativa y no ver diferencias entre los grupos es más frecuente cuando las muestras son de tamaño desigual (varias explicaciones, comentarios y alternativas metodológicas en *Overseas Chinese Association for Institucional Research* (http://www.ocair.org/) *IR & Statistics, Summaries of On-line Discussion* http://www.ocair.org/files/KnowledgeBase/Statistics/Anova.asp consultado 24, 08, 2007)

[18] Esta recomendación la vemos en Rodrigues (1977), Escotet (1980), Martínez Garza (1988) y en otros autores. En las tablas de la F muchos textos se limitan a $\alpha = .05$ y .01, pero en bastantes se pueden ver también los valores de F para $\alpha = .10$; además tablas con α=.10 las tenemos en Internet, y también en Internet podemos encontrar la probabilidad exacta de cualquier valor de F; al final dedicamos un apartado a recursos de Internet en relación con el análisis de varianza. También convine tener en cuenta que el contaste de Games y Howell puede ser una buena alternativa al de Scheffe.

para p = .05, $t'= \sqrt{(3\text{-}1)(3.55)} = 2.664$

para p = .01, $t'= \sqrt{(3\text{-}1)(6.01)} = 3.466$

3º Estos son nuestros valores de referencia (nuestras nuevas tablas). Comparamos ahora las t' del paso 1º con las que acabamos de calcular para comprobar qué valores de t' llegan o superan los valores de t' significativos (paso 2º). La única diferencia significativa se da entre los grupos 2º y 3º ($p < .05$).

Cuando son muchas las comparaciones entre medias que tenemos que hacer, se puede simplificar el procedimiento calculando en primer lugar cuál debe ser la diferencia mínima entre dos medias para determinados niveles de significación. Para esto nos basta despejar $|M_1 - M_2|$ en la fórmula de t' [7] y así tenemos que:

$$| M_1 - M_2 | = t' \sqrt{CM_{dentro}(\frac{1}{n_1} + \frac{1}{n_2})} \qquad [9]$$

Calculamos los valores de t' tal como lo hicimos antes y hallamos los valores significativos de $|M_1 - M_2|$, que en este ejemplo serán:

para un nivel de confianza de α =.05: $|M_1 - M_2| = 2.664 \sqrt{(16.26)(\frac{1}{7} + \frac{1}{7})} = 5.74$

para un nivel de confianza de α =.01: $|M_1 - M_2| = 3.466 \sqrt{(16.26)(\frac{1}{7} + \frac{1}{7})} = 7.47$

Ahora nos basta calcular las diferencias entre medias y comprobar si alcanzan los valores señalados:

Entre el grupo 1º y 2º: |11.71-15.43| = 3.72, inferior a 5.74 (p> .05); no significativa;
Entre el grupo 1º y 3º: |11.71- 8.57| = 3.14, inferior a 5.74 (p> .05); no significativa;
Entre el grupo 2º y 3º: |15.43- 8.57| = 6.86, *superior* a 5.74 (p< .05); *significativa*.

2) Utilizando un nivel de confianza más liberal (α = .10)

Si deseamos utilizar el contraste de Scheffé con un nivel de confianza más liberal, de α = 10, tenemos que buscar en las tablas de F el valor correspondiente a la probabilidad de α = .10 con 2 y 18 grados de libertad (en este ejemplo). Este valor de F es 2.62 que utilizaremos en la fórmula [8]

$$t'= \sqrt{(3-1)(2.62)} = 2.289$$

La diferencia mínima que debemos encontrar es la que hemos visto en la fórmula [9]:

Para un nivel de confianza de α =.10: $|M_1 - M_2| = 2.289 \sqrt{(16.26)(\frac{1}{7} + \frac{1}{7})} = 4.93$

En nuestro ejemplo y con α = .10 seguimos sin encontrar diferencias estadísticamente significativas entre los grupos 1º y 2º y 1º y 3º.

3) Contrastes combinando medias de grupos

Como ya hemos indicado, el método de Scheffé no sólo sirve para comparar dos grupos entre sí (es lo más habitual) sino también para comparar medias combinando varios grupos.

Por ejemplo, deseamos verificar si la media del grupo 2º supera significativamente a la media combinada de los grupos 1º y 3º:

La media combinada de estos dos grupos es 10.1425 (es la media de las medias, ya que se trata de grupos con idéntico número de sujetos).

El nuevo número de sujetos de este nuevo grupo es ahora 7 + 7 = 14.

Aplicando la fórmula [9] tenemos: $t' = \dfrac{|15.43 - 10.1425|}{\sqrt{(16.26)(\frac{1}{7} + \frac{1}{14})}} = 2.83$, $p < .05$

4.1.2. *Contraste de Tukey para muestras de idéntico (o muy parecido) tamaño*

El procedimiento de Tukey se basa en el estadístico q o del *rango estudentizado;* no es necesario entenderlo para utilizarlo (aunque se aprecia enseguida su semejanza con la t de Student), pero está explicado en el anexo IV (está programado en el SPSS).

a) Valoración general

1) Supone grupos de idéntico tamaño

c) Se pueden comparar todas las medias de dos en dos; el procedimiento está pensado para k(k-1)/2 comparaciones, todas las posibles entre k grupos tomados de dos en dos.

d) Es en principio preferible a Scheffé si se puede garantizar la homogeneidad de varianzas y la distribución normal de las poblaciones; el método de Tukey es *menos tolerante* con las violaciones de estas condiciones previas que el procedimiento de Scheffé.

Cuando es claro que no se cumplen estas condiciones, disponemos de adaptaciones de este contraste que exponemos en los apartados siguientes (contrastes de Tukey-Kramer y de Games y Howell).

e) Con el contraste de Tukey se rechaza con más facilidad la Hipótesis Nula que con el contraste de Scheffé; es preferible también cuando se está interesado en todos los posibles contrastes entre pares de medias.

b) Procedimiento

1º *La diferencia honestamente significativa* (DHS *Honestly Significant Difference)*

Una diferencia es *estadísticamente significativa* si es igual o mayor que el valor simbolizado como DHS (*diferencia honestamente significativa*). Habitualmente se calcula un solo valor con el que se comparan todas las diferencias.

$$DHS = q\sqrt{\frac{CM_{dentro}}{n}} \qquad [10]$$

DHS = *Diferencia honestamente significativa* o *diferencia mínima necesaria* para poder rechazar la Hipótesis Nula

q = Valor del *rango estudentizado* que hay que consultar en las tablas apropiadas para los grados de libertad del término del error (Cuadrados Medios *dentro* de los grupos) y para k (número de grupos o de medias)[19].

[19] Tablas de Tukey en el documento con tablas estadísticas y las direcciones de Internet del apartado nº 9.

CM_{dentro}= Varianza (Cuadrados Medios) *dentro* de los grupos;

n = Número de sujetos o datos en *cada* grupo.

En nuestro ejemplo el valor de q para tres grupos y 18 grados de libertad es 3.61 ($p < .05$) por lo que para rechazar la Hipótesis Nula habría que alcanzar una diferencia de:

$$DHS = 3.61\sqrt{\frac{16.26}{7}} = 5.50$$

Esta diferencia es algo menor que la que nos indica el método de Scheffé. El contraste de Tukey detecta mejor las diferencias significativas que el de Scheffé y se puede utilizar en cualquier planteamiento de análisis de varianza, con tal de que el número de sujetos sea igual (o *casi)* en todos los grupos.

Aunque quizás lo más frecuente sea calcular la *diferencia mínima* necesaria para rechazar la Hipótesis Nula que acabamos de ver [10], también se puede calcular directamente el valor de q (como hacemos con la t de Student); podemos despejar el valor de q en la fórmula [10]; DHS es ahora la diferencia entre las medias que queremos contrastar:

$$q = \frac{|M_1 - M_2|}{\sqrt{\frac{CM_{dentro}}{n}}} \qquad [11]$$

Vamos a comparar las medias de los grupos 2º (M = 15.428) y 3º (M = 8.571) (tabla 1); los *cuadrados medios dentro de los grupos* son igual a 16.26 (tabla 2):

$$q = \frac{|15.428 - 8.571|}{\sqrt{\frac{16.26}{7}}} = 4.499$$

En las tablas de q para k (número de medias) = 3 y 18 grados de libertad tenemos los valores de 3.61 (.05) y 4.70 (.01), por lo que en nuestro caso $p < .05$

Cuando se trata de un análisis de varianza con más criterios de clasificación (como cuando disponemos los datos en un cuadro de doble entrada), el valor de k no es siempre el número de grupos. Aquí nos limitamos a exponer el procedimiento para el caso de varias muestras independientes con un sólo criterio de clasificación; no para los casos en que hay más (como en los cuadros de doble entrada de los diseños factoriales; esto lo veremos en su lugar).

2º *Cuando el número de sujetos es ligeramente desigual*

Cuando el número de sujetos en cada grupo es desigual (pero no *muy* desigual), en vez de n (número de sujetos en cada grupo cuando son de idéntico tamaño), puede utilizarse la *media armónica* (n*), de los diversos n (k es el número de grupos)[20]:

[20] Cuando se utiliza la *media harmónica* suele emplearse la expresión *unweighted means analysis*, y no debe utilizarse con muestras muy pequeñas (<10) y con diferencias entre las muestras mayores de 2 en número de sujetos (Dallal, 2001). La recomendación de utilizar la media armónica de *n* cuando los grupos son de tamaño ligeramente desigual se encuentra en muchos autores (por ejemplo Wildt y Ahtola, 1978 y Klockars y Sax, 1986).

media armónica de $n^* = \dfrac{k}{\dfrac{1}{n_1} + \dfrac{1}{n_2} + \dfrac{1}{n_3} \ldots + \dfrac{1}{n_k}}$ o $n^* = \dfrac{k}{\Sigma(\frac{1}{n})}$ [12]

Otra posibilidad que recomiendan algunos autores[21] es utilizar la media armónica de *n* de *solamente* los *dos grupos que vamos a comparar* si las diferencias en tamaño son algo mayores, o simplemente podemos utilizar como valor de *n* el número de sujetos del grupo más pequeño; nuestro test será en este caso más conservador (como sucede siempre; si nos fijamos en la fórmula del contraste de Tukey veremos que a mayor *n* es más fácil rechazar la Hipótesis Nula de no diferencia).

4.1.3. *Contraste de Tukey-Kramer para muestras de distinto tamaño y varianzas iguales*

Este contraste es válido para comparar las medias de muestras de distinto tamaño, pero *con la condición de que las varianzas de las poblaciones a las que pertenecen las muestras sean iguales* (por *iguales* entendemos sin diferencias estadísticamente significativas).

Esta condición no siempre es fácil de confirmar por lo que este procedimiento *no se debe utilizar* (Toothaker, 1993) si a) las desviaciones típicas (o varianzas) de las muestras no son muy parecidas o b) si no disponemos de más datos de otros estudios que nos confirmen que las varianzas de las poblaciones son de magnitud semejante.

Aunque esta condición es bastante restrictiva se trata de un contraste aplicable en muchas ocasiones en las que tenemos muestras de tamaño desigual. Además con un ajuste en los grados de libertad es un contraste válido cuando el tamaño de las muestras es distinto y también las varianzas son distintas (y tenemos aquí una alternativa al procedimiento de Scheffé); se trata en este caso del contraste de Games y Howell que veremos en el apartado siguiente.

Este contraste sigue también la distribución de q (como el contraste habitual de Tukey) y suele presentarse de dos maneras (obviamente equivalentes); presentamos las dos para evitar confusiones.

a) Calculamos en primer lugar este valor de t':

$$t' = \frac{M_i - M_k}{\sqrt{CM_{dentro}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}} \quad [13]$$

M y n son respectivamente las medias que vamos a comparar y el número de sujetos de *cada* muestra; CM_{dentro} son, como siempre, los *cuadrados medios dentro de los grupos.*

Calculamos el valor crítico de t' a partir del valor de q (al nivel de confianza deseado; tablas de q):

$$t' = \frac{q}{\sqrt{2}} \quad [14]$$

Buscamos el valor de q correspondiente al número de medias de nuestro planteamiento (número de muestras) y los grados de libertad de los *cuadrados medios dentro de los grupos* (como es usual en estas tablas).

---

[21] Por ejemplo Klockars y Sax (1986).

También podemos calcular directamente la *diferencia mínima* que tenemos que encontrar para afirmar la diferencia; para esto nos basta con despejar el numerador de la fórmula [13] sustituyendo q por el valor que encontremos en las tablas:

$$|M_i\text{-}M_k| = \frac{q}{\sqrt{2}}\sqrt{CM_{dentro}\left(\frac{1}{n_i}+\frac{1}{n_k}\right)} \qquad [15]$$

Si vamos a hacer varias comparaciones entre medias lo único que irá variando es el tamaño de los grupos; el resto de los valores (q y CM_{dentro}) son constantes.

b) Dada la relación entre t' y q cuando se trata de dos muestras ($t' = q/\sqrt{2}$) podemos también calcular directamente el valor de q (fórmula [16]) y consultar las tablas de q.

$$q = \frac{M_i - M_k}{\sqrt{CM_{dentro}\left[\frac{(1/n_i)+(1/n_k)}{2}\right]}} \qquad [16]$$

Conviene caer en la cuenta de que ambas fórmulas ([13] y [16]) son en última instancia equivalentes, para evitar confusiones si encontramos distintas fórmulas en distintos lugares.

Vamos a ver la equivalencia de estos procedimientos con un ejemplo que además clarifica el uso de estas fórmulas. En la tabla 4 tenemos los datos de tres grupos de tamaño distinto (A, B y C).

	A	B	C
n =	10	9	8
M =	8.00	11.55	14.125
σ_n =	3.898	4.7868	4.780

Tabla 4

El número total de sujetos es N = 27. Suponemos que la razón F es significativa (realmente lo es).

Los *cuadrados medios dentro de los grupos* (que necesitamos para aplicar las fórmulas [13] y [16]) son, como ya sabemos (fórmula [5]), igual a $\Sigma n\sigma^2/N - k = 22.539$

Contrastamos las medias de los grupos C y A, primero con la fórmula [13]:

$$t' = \frac{14.125 - 8.00}{\sqrt{22.539\left(\frac{1}{8}+\frac{1}{10}\right)}} = 2.7198$$

Los grados de libertad son: 3 (número de grupos)
24 (N-k, grados de libertad *dentro* de los grupos)

En las tablas de q vemos que con un nivel de confianza de $\alpha = .05$ necesitamos q = 3.53, por lo que el valor crítico de t' (fórmula [14]) es por lo tanto:

$$t' = \frac{3.53}{\sqrt{2}} = 2.496$$

Como nuestro valor de t' (2.719) supera este valor, rechazamos la Hipótesis Nula y concluimos que las muestras A y C difieren significativamente (proceden de poblaciones distintas con distinta media).

Para hacer este mismo contraste vamos a calcular ahora directamente el valor de q (fórmula [16]):

$$q = \frac{14.125 - 8.00}{\sqrt{22.539\left[\frac{(1/10)+(1/8)}{2}\right]}} = 3.846$$

Superamos el valor de 3.53 que vemos en las tablas para 3 y 24 grados de libertad y llegamos a la misma conclusión (p< 05)

Verificamos por último que en el caso de dos muestras $t = \frac{q}{\sqrt{2}}$

Hemos obtenido un valor de t = 2.719 y de q = 3.846; efectivamente $\frac{3.846}{\sqrt{2}} = 2.719$

4.1.4. *Contraste de Games y Howell (GH) para muestras de distinto tamaño y varianzas desiguales*

Cuando no estamos seguros de que las varianzas de las poblaciones a las que pertenecen nuestras muestras sean iguales, disponemos del contraste de Games y Howell (suele simbolizarse simplemente como GH) (está programado en el SPSS).

El procedimiento y las fórmulas son idénticos al contraste anterior de Tukey-Kramer [13]; la diferencia está en los *grados de libertad dentro de los grupos* para buscar el valor de q en las tablas; estos grados de libertad quedan reducidos y hará falta un valor mayor de F.

En este caso la fórmula de los grados de libertad es la siguiente (Klockars y Sax, 1986; Toothaker, 1993)[22]:

$$\textit{grados de libertad} = \frac{\left[\left(\sigma_i^2/n_i\right)+\left(\sigma_k^2/n_k\right)\right]^2}{\frac{(\sigma_i^2/n_i)^2}{n_1-1}+\frac{(\sigma_k^2/n_k)^2}{n_k-1}} \qquad [17]$$

Esta fórmula se puede expresar de una manera más simple que facilita el cálculo:

$$\textit{grados de libertad} = \frac{(d_i - d_k)}{\frac{d_i^2}{n_i - 1}+\frac{d_k^2}{n_k - 1}} \qquad [18] \qquad \text{donde } d = \frac{\sigma^2}{n}$$

Como vemos en [17] los grados de libertad para consultar las tablas de q no son N-k, sino que los calculamos a partir de los valores de las desviaciones típicas (de las muestras) y del número de sujetos que tenemos en los grupos que comparamos. Lo normal es que obtengamos una cifra con decimales que podemos redondear al número entero más próximo.

Podemos hacer la misma comparación entre las medias de los grupos C y A (tabla 4) asumiendo que las varianzas son muy diferentes. El valor de q ya lo tenemos calculado (3.846), lo que necesitamos son los grados de libertad; aplicando la fórmula [17] tenemos:

---

22 Es la fórmula de Satterthwaite (Dallal, 2001, en *Significance Tests / Hypothesis Testing*); como indica el mismo autor (Dallal, 2001) éste es el procedimiento utilizado por el programa SPSS en esta situación (muestras de tamaño distinto y varianzas desiguales).

$$\textit{grados de libertad} = \frac{\left[(\sigma_C^2/n_C)+(\sigma_A^2/n_A)\right]^2}{\frac{(\sigma_C^2/n_C)^2}{n_C-1}+\frac{(\sigma_A^2/n_A)^2}{n_A-1}} = \frac{\left[(4.78^2/8)+(3.898^2/10)\right]^2}{\frac{(4.78^2/8)^2}{8-1}+\frac{(3.898^2/10)^2}{10-1}} = 13.46$$

Los grados de libertad son ahora 13 (redondeando decimales), bastante menos que los 24 (N-k) que teníamos antes. Para grados de libertad 3 (número de grupos) y 13, y $\alpha = .05$ necesitamos un valor de q de 3.73 (mayor que q = 3.53 con 24 grados de libertad que teníamos antes), por lo que nuestro valor de q = 3.846 (mayor que 3.73) sigue siendo suficiente para no aceptar (rechazar) Hipótesis Nula.

Esta fórmula de los *grados de libertad de los cuadrados medios dentro de los grupos* es aparentemente laboriosa, pero con muestras de tamaño muy distinto y varianzas claramente desiguales, este procedimiento es una buena alternativa al de Scheffé[23].

4.1.5. Contraste de Newman-Keuls

El test de Newman-Keuls es muy popular, también requiere muestras de idéntico tamaño y es algo más liberal que el de Tukey; también utiliza las tablas de q.

La fórmula es idéntica a la de Tukey [12], la diferencia está al consultar las tablas de q: 1º las medias *se ordenan de más a menos* y 2º el *número de medias* para consultar las tablas no es el número de medias que tenemos en el análisis de varianza sino las dos que se comparan *más* las que estén entre estas dos una vez ordenadas; es decir quedan excluidas las medias mayores o menores de las dos que comparamos. Los valores críticos de q bajan según baja el número de medias, por lo que resulta más fácil no aceptar la Hipótesis Nula; es por lo tanto un contraste más liberal que el de Tukey.

El contraste de Duncan (*new multiple range test*) es una modificación del de Newman-Keuls, también se utilizan las tablas de q y es algo más liberal (programado en el SPSS).

4.1.6. Contraste *Least Significant Difference* (LSD) de Fisher.

Con este contraste se utilizan las mismas tablas de la t de Student y se puede utilizar con muestras de distinto tamaño. La diferencia con la *t de Student* está en que en el denominador de la t (donde figura el *error típico* de la diferencia entre medias) colocamos los Cuadrados Medios *dentro,* por lo tanto la fórmula es:

$$\text{t (Fisher)} = \frac{M_1 - M_2}{\sqrt{CM_{\text{dentro}}\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}} \qquad [20]$$

Cuando n = n esta fórmula es idéntica a la de Dunnett pero las tablas que hay que consultar en el contraste de Dunnett son distintas.

Este contraste (LSD) sólo se debe utilizar si la F es significativa; es uno de los contrastes más liberales (es *más fácil* no aceptar la Hipótesis Nula).

---

[23] Estos *grados de libertad* también son válidos, y recomendables, para consultar las tablas de la *t de Student* cuando comparamos dos muestras con tamaño y varianzas muy desiguales (Hinkle, Wiersma y Jurs, 1994; Coombs, Algina y Oltman, 1996).

4.1.7. *Contraste de Dunnett*

a) *Válido para comparar las medias de los grupos experimentales (tratamientos) con un grupo de control*; no para comparar los tratamientos entre sí (está pensado para k-1 comparaciones; por eso se rechaza la Hipótesis Nula con más facilidad; están previstas menos comparaciones). Es útil precisamente en estos casos, cuando tenemos varios grupos experimentales y un grupo de control (está programado en el SPSS).

b) Supone *idéntico número de sujetos en cada grupo* (n) aunque también podemos utilizar la *media armónica* de n, como vimos en el método de Tukey, cuando el número de sujetos es desigual pero muy parecido.

Fórmula del contraste de Dunnett: $$t' = \frac{\text{Media}_{\text{tratamiento}} - \text{Media}_{\text{control}}}{\sqrt{\dfrac{2CM_{\text{dentro}}}{n}}} \qquad [19]$$

CM_{dentro}, o *cuadrados medios del error,* son, como siempre en estos casos, los *cuadrados medios* (o varianza) *dentro* de los grupos. Este contraste tiene sus *propias tablas* para verificar la probabilidad de un determinado valor de t'[24].

4.1.8. Contrastes ortogonales

Sin entrar en una explicación del procedimiento, mencionamos señalamos las características de estos contrastes.

1) Se planifican de antemano, *antes de recoger los datos* y responden a hipótesis muy específicas,
2) Responden a preguntas *independientes* que no aportan información redundante (por eso se denominan *ortogonales,* sin relación)[25],
3) Se pueden comparar tanto medias entre sí como combinaciones de medias,
4) El número de comparaciones que es permisible hacer es limitado y no puede ser superior a k -1 (número de grupos menos uno).

No los explicamos aquí pero los tenemos en programas de ordenador y se encuentran con facilidad en otros textos[26].

4.1.9. Valoración de los diferentes contrastes

No todos los contrastes vistos son igualmente rigurosos, con unos es más fácil que con otros rechazar la Hipótesis Nula.

En la lista siguiente[27] aparecen los más comunes, situados en un continuo según sean más *liberales* o más *conservadores* (según sea más fácil o más difícil rechazar la Hipótesis Nula).

---

[24] Tablas de Dunnett en el documento con tablas estadísticas y en las direcciones de Internet indicadas en el apartado nº 9.

[25] Si por ejemplo nuestra hipótesis es que los grupos A+C (dos grupos unidos en uno solo) tienen una media mayor que los grupos B+D (otros dos grupos unidos en uno solo), no podemos plantearnos *también* si la media de A es mayor que la de B, porque parcialmente esta comparación está incluida en la anterior. El término *ortogonal* (*orthogonality*) fué usado por primera vez por Yates en 1933 (StatSoft, Inc., 2007).

[26] Los contrastes ortogonales no suelen venir explicados en textos de carácter más bien básico, pero sí en los de un nivel medio o superior, pues son algo más complejos que los que exponemos aquí. Pueden verse explicados en Klockars y Sax (1986), Guilford y Fruchter (1978) y Kirk (1995) entre muchos otros.

más liberal	Fisher (LSD) *(Least Significant Difference)*
	Duncan *(new multiple range test)*
	Newman-Keuls
	Tukey *(Honestly Significant Difference)*
más conservador	Scheffé

Algunos autores recomiendan no utilizar los contrastes de Fisher, Duncan y Newman-Keuls cuando tenemos *más de tres grupos;* en este caso son contrastes *demasiado liberales* (se rechaza con demasiada facilidad la Hipótesis Nula)[28]. Esta lista no incluye el test de Dunnett porque no está concebido para comparaciones múltiples, sino sólo para comparar distintas muestras experimentales con una sola muestra de control.

Al disponer de tantas posibilidades distintas de hacer los contrastes posteriores, puede parecer que este paso, necesario por otra parte para llegar a las conclusiones finales, es complicado. En la práctica una buena orientación es escoger entre Scheffé y Tukey (son los que con más frecuencia aparecen en los libros de texto) o el de Dunnett cuando se trata de comparar varios grupos experimentales con un grupo de control o de referencia. Los contrastes de Tukey-Kramer y Games y Howell (variantes de Tukey) también son los oportunos en las circunstancias indicadas al comentar estos contrastes (ambos contrastes coinciden en el tamaño desigual de las muestras). Al decidir sobre la elección de alguno de estos contrastes es conveniente tener a la vista las observaciones hechas a propósito de los mismos[29].

Estos contrastes suelen dividirse en dos tipos: planificados de antemano *(a priori)* y no planificados de antemano *(a posteriori* o *post hoc)*; el término *post hoc* se emplea frecuentemente para designar los contrastes hechos después de inspeccionar los datos. Los diversos contrastes para comparar las medias de dos en dos (como el de Tukey) suelen incluirse ente los métodos *a posteriori,* pero también pueden planificarse de antemano, por lo que esta distinción no es tan nítida; en un análisis exploratorio también se pueden especificar *a priori* los contrastes de interés[30].

---

[27] Tomada de Huck, Cormier, y Bounds (1974); Black (1999) y Gerstman (2003) presentan cuadros semejantes y los mismos contrastes aparecen con el mismo orden.

[28] Toothhaker (1993); Kirk (1995); Hancock y Klockars (1996)

[29] Hancock y Klockars (1996) tienen una buena revisión crítica de todos estos contrastes. Entre los pocos que en última instancia recomiendan son Scheffé (para *explorar),* Tukey (para comparar entre sí *todas las medias de dos en dos* en muestras de idéntico tamaño*)* y Dunnett (no para hacer todas las comparaciones posibles sino únicamente para comparar *muestras experimentales con una muestra de control).* Keselman *y* otros (1998) tienen un interesante estudio sobre las *preferencias metodológicas en los diversos tipos del análisis de varianza,* tal como aparecen en investigaciones publicadas en 17 revistas de prestigio *(Child Development, Developmental Psychology, American Educational Research Journal, Journal of Applied Psychology,* etc.). Los *contrastes posteriores* más utilizados son por este orden: Tukey, Newman-Keuls, Scheffé y Fisher *Least Significant Difference.* También es posible *utilizar contrastes distintos para comparar distintos pares de medias entre sí:* "Most analysts agree that Fisher's LSD is too liberal. Some feel that Tukey's HSD is too conservative. While it is clear that the largest difference between two means should be compared by using Tukey's HSD, it is less obvious why the same criterion should be used to judge the *smallest* difference. The [Student]-Newman-Keuls Procedure is a compromise between LSD and HSD… I use Tukey's HSD for the most part… One general approach is to use both Fisher's LSD and Tukey's HSD. Differences that are significant according to HSD are judged significant; differences that are not significant according to LSD are judged nonsignificant; differences that are judged significant by LSD by not by HSD are judged open to further investigation" (Dallal, 2001). Este autor tiene una buena exposición (*online*) sobre los distintos contrastes.

[30] Hancock y Klockars (1996) reservan el término *a posteriori* exclusivamente al contraste de Scheffé; en Jaccard (1998:27) pueden verse discutidos estos términos.

En los programas de ordenador con frecuencia aparecen de manera rutinaria una serie de contrastes (aquí *no* están expuestos o mencionados todos los posibles); lo recomendable es repasar la información disponible sobre estos contrastes para escoger e interpretar el (o los) que se estime más idóneo para una situación determinada.

4.2. Relevancia *práctica* de los resultados: proporción de varianza relacionada con la variable-criterio de clasificación y tamaño del efecto

Mediante el análisis de varianza propiamente dicho y los contrastes posteriores verificamos si las diferencias entre las medias son *estadísticamente significativas.* Con esta expresión lo que queremos decir es si podemos considerar que las diferencias observadas entre los grupos en la variable dependiente (aquella en la que hemos medido a los sujetos) son mayores de lo que podemos esperar por azar, es decir, si no están dentro de la variabilidad *normal.* En ese caso podemos atribuir las diferencias a la pertenencia a alguno de los grupos (a un tratamiento, etc.). Ahora bien, esta *significación estadística* no coincide necesariamente con la *significación* (o *relevancia*) *práctica:* el *efecto* de un tratamiento (o de la pertenencia a un grupo determinado) puede ser estadísticamente significativo pero pequeño e irrelevante. Valores grandes de F y pequeños de p (como $p<.001$) *no* indican *efectos* o *diferencias grandes*. Un $p <.05$ con un grupo pequeño puede ser *más importante* que un $p<.001$ con una muestra grande.

Para juzgar sobre la *relevancia práctica* de los resultados no tenemos un indicador preciso y hay que hacer juicios cualitativos, pero es útil poder apreciar la *magnitud* de ese efecto. Disponemos al menos de dos orientaciones metodológicas relacionadas entre sí:

a) Podemos estimar la *proporción de varianza* en la variable dependiente (la que hemos medido) asociada con el criterio que ha servido para clasificar a los sujetos (pertenencia a un grupo u otro).

b) Podemos calcular un *tamaño del efecto* análogo al que hacemos en el contraste de dos medias.

En el contexto del análisis de varianza y para poder interpretar mejor los resultados, posiblemente lo más habitual es calcular la proporción de varianza que podemos atribuir a los criterios de clasificación (a la variable independiente).

4.2.1. Proporción de varianza relacionada con la variable-criterio de clasificación

Esta proporción de varianza en la variable dependiente asociada o atribuible a los criterios de clasificación nos la dan una serie de coeficientes que pueden considerarse como un *tamaño del efecto.* Los coeficientes propuestos aquí son los coeficientes ω^2 y η^2. Estos coeficientes y otros análogos son importantes porque ayudan a poner de relieve la *significación práctica* de los resultados y son comparables en su interpretación al *tamaño del efecto* en el contraste de medias. Lo que no podemos hacer es calcular la media de estos coeficientes, como se hace en el *meta-análisis,* para *resumir* resultados de diversos estudios, ya que son siempre positivos y no indican la dirección de la diferencia.

4.2.1.1. El coeficiente ω^2

Este coeficiente ω^2 puede aplicarse en el análisis de varianza unifactorial (un único factor o criterio de clasificación, el que estamos viendo ahora) siempre que las categorías de clasificación sean *fijas* (como suelen ser habitualmente, es decir, escogidas con criterios lógicos, y no escogidas aleatoriamente de una población mayor, por ejemplo de centros escolares, etc.).

El coeficiente ω^2 es un índice general de asociación entre dos variables, y, como ya se ha indicado, aporta una información análoga al *tamaño del efecto* en el contraste de medias. El valor de ω equivale a un coeficiente de correlación, y elevado al cuadrado nos indica la proporción de varianza compartida por las dos variables. En realidad no puede hablarse con propiedad de coeficiente de correlación, pues las categorías de clasificación (pertenencia a un grupo u otro) no son necesariamente continuas; en cualquier caso ω^2 nos indica la *proporción de varianza* en la variable dependiente (la que hemos medido) atribuible a la pertenencia a uno u otro grupo.

1. *Cuando los grupos son de idéntico tamaño*

En el análisis de varianza de una clasificación simple (un criterio de clasificación dividido en varios niveles, que es el modelo que estamos viendo) y con grupos de *idéntico tamaño* el cálculo es el siguiente (Guilford y Fruchter, 1973:245):

$$\omega^2 = \frac{(k-1)(F-1)}{(k-1)(F-1)+kn} \qquad [19]$$

k = número de grupos,
n = número de sujetos en cada grupo,
F = la razón F obtenida en el análisis.

En este caso tendremos que $\omega^2 = \dfrac{(3-1)(5.069-1)}{(3-1)(5.069-1)+(3)(7)} = .28$

Este resultado nos indica que aproximadamente (se trata de una *estimación)* el 28% de la varianza (diferencias en los datos analizados) está relacionado con la pertenencia a uno u otro grupo. Con una F significativa sólo sabíamos que esta relación era superior a cero (que había relación, pero no cuánta).

Esta otra fórmula del coeficiente ω^2 da una *estimación* semejante:

$$\varpi^2 = \frac{SC_{entre} - (k-1)CM_{dentro}}{SC_{total} + CM_{dentro}} \qquad [22]$$

En nuestro ejemplo tenemos que:

$$\omega^2 = \frac{164.87-(3-1)16.26}{457.79+16.26} = .28$$

2. *Cuando los grupos son de distinto tamaño*

En las fórmulas anteriores se asume que los grupos son de idéntico tamaño (o no muy diferente). Cuando los grupos son de tamaño claramente desigual la fórmula que suele proponerse como *estimación* de ω^2 es ésta (Kirk, 1995):

$$\omega^2 = \frac{SC_{entre} - (k-1)CM_{dentro}}{SC_{entre} + [k(\frac{N}{k}-1)CM_{dentro}] + CM_{dentro}} \qquad [23]$$

k = número de grupos
N/k = *tamaño medio* de los grupos

En nuestro ejemplo, aunque los grupos son de idéntico tamaño, aplicando esta fórmula tendríamos:

$$\varpi^2 = \frac{164.87-(3-1)(16.26)}{164.87+[3(\frac{21}{3}-1)16.26]+16.26} = .28$$

Para interpretar ω^2 hay que tener en cuenta que[31]:

1. Se trata siempre de *categorías fijas* (posiblemente el caso más común),
2. Un coeficiente ω^2 negativo se considera igual a cero;
3. El coeficiente ω^2 sólo debe calcularse si F es significativo;
4. El coeficiente ω^2 sólo puede aplicarse a las categorías de clasificación utilizadas; puede variar si se quitan o aumentan categorías. No sucede lo mismo con la interpretación del valor de F, que, si es significativo y las muestras son aleatorias, puede extrapolarse a la población.
5. El coeficiente ω^2 se puede extrapolar a la población, al menos como estimación; en cambio el coeficiente η^2 que veremos a continuación se refiere solamente a las muestras analizadas.

4.2.1.2. El coeficiente η^2

Otro coeficiente, muy sencillo y muy utilizado y que da una aproximación menos exacta de la proporción de la variable dependiente atribuible a los criterios de clasificación es el coeficiente *eta al cuadrado:*

$$\eta^2 = \frac{SC_{\text{entre}}}{SC_{\text{total}}} \qquad [24]$$

SC_{entre} es la Suma de Cuadrados *entre grupos* en este caso, pero en otros modelos de análisis de varianza puede ser cualquier criterio de clasificación (puede haber más de un criterio de clasificación, como vemos en los cuadros de doble entrada propios de los diseños factoriales)

La misma fórmula ya nos está indicando que se trata de una *proporción:* la proporción de varianza de la variable dependiente explicada por la variable cuya suma de cuadrados está en el numerador[32]. En este planteamiento, varias muestras independientes, lo que tenemos en el numerador son las diferencias asociadas a pertenecer a uno u otro grupo. No se puede extrapolar a la población, se refiere solamente la muestra. Suele dar valores mayores que el coeficiente ω^2

En nuestro ejemplo (tabla 2) tenemos que $\eta^2 = \dfrac{164.87}{457.79} = .36$

Si solamente conocemos del valor de F (podemos encontrarlo publicado) disponemos de esta otra fórmula (parecida a la [21]):

$$\eta^2 = \frac{(\text{F})(\text{grados de libertad } \textit{entre})}{(\text{F})(\text{grados de libertad } \textit{entre}) + \text{grados de libertad } \textit{dentro}} \qquad [25]$$

Aplicando esta fórmula a los datos de la tabla 2 tenemos: $\eta^2 = \dfrac{(5.069)(2)}{(5.069)(2)+18} = .36$

No sobra observar el parecido de la fórmula [25] con la que utilizamos para transformar el valor de t en un coeficiente de correlación [26]:

---

31 Guilford y Fruchter (1973:260); Kirk (1995:180)

32 Normalmente utilizamos η^2 que es análogo a r^2 (que expresa la proporción de varianza común a dos variables); η (sin elevar al cuadrado) (también denominado *correlation ratio* en inglés) viene a ser un coeficiente de correlación *no lineal* (el coeficiente r de Pearson cuantifica relaciones lineares, no curvilíneas). Si se calcula η (grado de relación) solamente con dos variables continuas y relacionadas linealmente *(a más de una, más de la otra),* η equivale a r (explicación más detallada en Nunnally y Bernstein, 1994:135 y Rosenthal y Rosnow, 1991:323, 351).

$$r_{bp} = \sqrt{\frac{t^2}{t^2 + (N_1 + N_2 - 2)}} \qquad [26]$$

Realmente se trata de la misma fórmula si la elevamos al cuadrado (r^2 expresa la proporción de varianza compartida): en el caso de dos grupos t^2 = F y r = η; los grados de libertad entre los grupos son 2-1 = 1 y los grados de libertad al comparar dos grupos son N_1+N_2-2

η^2 se puede utilizar con muestras de idéntico o distinto tamaño; es un estadístico meramente descriptivo, referido a la muestra, y su cálculo suele ser habitualmente suficiente; para extrapolar esta misma información a la población podemos calcular el coeficiente ω^2

El coeficiente η^2 si lo referimos a la población nos da una estimación demasiado alta; una estimación más ajustada del valor de η^2 en la población la tenemos con este ajuste[33]:

$$\tilde{\eta}^2 = 1 - \frac{N-1}{N-k}(1-\eta^2) \qquad [27]$$

Aplicado a nuestros datos (tabla 1) tenemos que: $\tilde{\eta}^2 = 1 - \frac{21-1}{21-3}(1-.36) = .289$

Este tipo estimaciones de la proporción de varianza de la variable dependiente explicada por la variable independiente pueden expresarse de varias maneras y conviene conocer las más usuales para evitar confusiones, ya que se trata de lo mismo, así la fórmula [27] es equivalente a esta otra[34] [28]:

$$\tilde{\eta}^2 = \frac{CM_{total} - CM_{dentro}}{CM_{total}} \qquad [28]$$

Tomamos los datos de la tabla 2. Para obtener los *cuadrados medios (o varianza) del total* nos basta dividir la *suma de cuadrados* por los *grados de libertad:* 457.79/20 =22.8895

Aplicando la fórmula [28] tenemos $\qquad \tilde{\eta}^2 = \frac{22.8895 - 16.26}{22.8895} = .289$

4.2.1.3. Valoración de estos coeficientes

1) Estos coeficientes (que también se aplican a otros modelos de análisis de varianza) son importantes como análisis complementarios, porque nos ayudan a juzgar sobre la *relevancia* de los resultados y su *importancia relativa* cuando tenemos varios coeficientes.

2) Se trata de coeficientes *no direccionales;* es decir, aunque nos facilitan el valorar la importancia de las diferencias entre las medias de la variable independiente, ya hemos indicado que no nos dicen en qué *dirección* está la diferencia. Estos coeficientes también se denominan *tamaño del efecto* (es un término de hecho muy genérico) porque expresan *magnitud* y no probabilidad (t o F nos remiten a una probabilidad).

3) Para los mismos datos los dos coeficientes, ω^2 y η^2, dan valores diferentes, por lo que no deben compararse entre sí (η^2 suele ser mayor que ω^2).

4) ¿Cuándo se puede decir que estas proporciones de varianza explicada son *grandes?*

---

[33] Hedges y Olkin (1985:102); Kirk (1995:180)

[34] Estas y otras fórmulas pueden verse comentadas en Hedges y Olkin (1985:100ss.)

No hay criterios definidos pero a la vista de lo que suele encontrase en la literatura experimental, si la variable independiente explica menos de un 5% de la variable dependiente se trata de una relación muy baja aunque sea *real* (*no casual*…), y un 10% de la varianza explicada puede considerarse mayor de lo que es normal encontrar (valoración de Linton, Gallo y Logan, 1975; para Runyon y Haber, 1984, es difícil encontrar valores que lleguen al 35%).

Una *valoración orientadora*[35] referida a ω^2 es ésta:

.01 expresa una asociación *baja,*
.06 expresa una asociación *media*
.14 o más expresa una asociación *grande* o apreciable.

5) En la práctica siempre es ilustrativo e informativo comparar unos valores con otros cuando proceden de planteamientos semejantes, e interpretar esos valores en términos relativos.

Estos coeficientes se utilizan con *categorías fijas* (escogidas por el investigador); si se trata de *categorías aleatorias* (escogidas aleatoriamente de una población mayor, como sería el caso de una muestra de centros escolares, de profesores, etc., escogidos aleatoriamente para representar una población), se utilizan los coeficientes de *correlación intraclase* que pueden verse tratados en diversos autores[36].

4.2.2. El *tamaño del efecto*

Los coeficientes de asociación que hemos visto también son denominados *tamaño del efecto*, pero este término puede ser aquí equívoco en el sentido de que no cuantifican la diferencia entre *dos* medias; ahora nos referimos al *tamaño del efecto* en cuanto *diferencia tipificada,* propia del contraste de *dos medias* entre sí.

En el análisis de varianza vemos dos tipos de tamaño del efecto: uno es semejante al habitual en el contraste de medias y nos cuantifica la diferencia entre *dos* medias; el otro nos da una *apreciación global* (como el coeficiente η) y en su uso e interpretación es análogo a los coeficientes de asociación.

4.2.2.1. El tamaño del efecto en la diferencia entre dos medias

No es tan frecuente ver calculado el habitual *tamaño del efecto* como complemento a los contrastes posteriores, pero podemos hacerlo como en cualquier contraste de medias. Ya sabemos que el tamaño del efecto es una *diferencia estandarizada:* la diferencia entre dos medias dividida por la desviación típica común a ambas muestras. Nos expresa la diferencia entre dos medias en unas magnitudes fácilmente interpretables y comparables con cualquier otro tamaño del efecto aunque las escalas de medición sean muy distintas. Por otra parte disponemos de los criterios de Cohen (1988), muy seguidos habitualmente, para valorar estas magnitudes (en torno a .20 diferencia *pequeña,* en torno a .50 diferencia *moderada* y .80 o más diferencia *grande).*

En el caso del análisis de varianza la desviación típica que ponemos en el denominador puede ser la desviación típica combinada de las *dos* muestras que comparamos, pero también podemos utilizar (y es más sencillo) la desviación típica combinada de *todas* las muestras de

[35] Estas valoraciones las propone Cohen (1988:284-288) y son habitualmente tenidas en cuenta (un ejemplo indicativo es el conocido texto de Kirk, 1995:178); también se aplican a coeficientes análogos como η^2

[36] Por ejemplo en Hedges y Olkin (1985:101ss), Kirk (1995) y otros.

nuestro planteamiento sobre todo si desconocemos las desviaciones típicas de las muestras o si las desviaciones típicas de todas las muestras no difieren mucho entre sí[37]. Esta desviación típica, como ya sabemos, no es otra cosa que *la raíz cuadrada de los cuadrados medios dentro de los grupos* (porque estos *cuadrados medios* son la varianza común, no la desviación típica común).

La fórmula del tamaño del efecto (que simbolizamos como *d*) es por lo tanto[38]:

$$d = \frac{|M_i - M_k|}{\sqrt{CM_{dentro}}} \qquad [29]$$

Podemos calcular el tamaño del efecto de los grupos correspondiente a la diferencia entre los grupos 2 y 3 (tabla 1); el denominador *(cuadrados medios dentro de los grupos)* lo tomamos de la tabla 2.

$d = \frac{|15.428 - 8.571|}{\sqrt{16.26}} = \frac{6.857}{4.03} = 1.70$; podemos valorar esta diferencia como *grande.*

Si utilizamos en el denominador la desviación típica combinada de estos dos grupos nada más (algo que también podemos hacer), el tamaño del efecto que obtenemos es de 1.88, ligeramente mayor.

Cuando la información disponible es sólo la razón F y el tamaño de los grupos, Thalheimer y Cook (2002) proponen esta fórmula para calcular una estimación del tamaño del efecto.

$$d = \sqrt{F\left[\frac{n_1 + n_2}{(n_1)(n_2)}\right]\left[\frac{n_1 + n_2}{(n_1 + n_2 - 2)}\right]} \qquad [30]$$

4.2.2.2. El tamaño del efecto como *apreciación global* de la magnitud de las diferencias entre todas las medias

Cuando tenemos solamente dos grupos, el tamaño del efecto es igual a la diferencia entre las dos medias dividida por la desviación típica combinada. Cuando hay más dos grupos simultáneamente, como sucede en el análisis de varianza, disponemos de un tamaño del efecto que no se refiere a la diferencia entre dos medias, sino a todas las diferencias de las medias con respecto a la media total. Lo que en este caso tenemos en el numerador es una cuantificación de la dispersión o de las desviaciones de *todas* las medias con respecto a la media común; comparamos estas diferencias entre todos los grupos con las diferencias entre los sujetos[39].

Expresado este concepto del tamaño del efecto en términos no convencionales, la fórmula sería ésta [31]:

$$\textit{tamaño del efecto} = \frac{\textit{variabilidad de las medias}}{\textit{variabilidad de los sujetos}} \qquad [31]$$

---

[37] Cortina y Mauri (1999:11ss)

[38] Jaccard (1998:36); Cortina y Nouri (2000:13); en estos dos últimos autores puede verse un tratamiento más extenso y específico del tamaño del efecto en el contexto del análisis de varianza.

[39] Explicado en Cohen (1988:274ss., 284). El numerador es análogo a la desviación típica de las medias; el denominador es el mismo visto en el apartado anterior (la raíz cuadrada de los *Cuadrados Medios dentro de los grupos).* El modo exacto de cálculo puede verse en Kirk (1995, 180ss) con un ejemplo resuelto; en la práctica es más sencillo atenerse al coeficiente ω^2

En la práctica el cálculo más sencillo es a partir de η^2 o ω^2, pues ambos valores están relacionados de esta manera (Cohen ,1988:284).

$$f = \sqrt{\frac{\omega^2}{1-\omega^2}} \qquad [32]$$

Este tamaño del efecto asociado a la razón F se simboliza como *f* (Rosenthal y Rosnow (1991:450; que en esta fórmula proponen η^2 en vez de ω^2). Naturalmente *este* tamaño del efecto no nos dice a quién favorece la diferencia, y no se puede utilizar para calcular el *tamaño del efecto medio* tal como se hace en el *meta-análisis*.

Los valores de referencia para valorar la magnitud de este tamaño del efecto corresponden a los ya vistos de ω^2:

tamaño del efecto	*pequeño*	*moderado*	*grande*
ω^2	.01	.06	.14
f	.10	.25	.40

En el caso de *dos grupos* nada más, y utilizando η^2 en vez de ω^2, f se relaciona con el tamaño del efecto (d de Cohen) de esta manera: f = d/2 (Rosenthal y Rosnow, 1991:450).

Es más frecuente calcular ω^2 o η^2 que f; de comprensión más sencilla es η^2 (una simple proporción) aunque es útil conocer también este *tamaño del efecto* (f) pues también se utiliza y además podemos encontrarlo en diversas tablas como referencia para calcular el número de sujetos necesario para un determinado experimento o estudio (la magnitud deseada es una variable que entra en la determinación del tamaño de la muestra).

5. Análisis de varianza cuando solamente conocemos los valores de las medias y de las desviaciones típicas

Hacemos un primer lugar una observación de interés y que justifica este apartado. Los procedimientos que suelen exponerse para resolver el análisis de varianza parten del supuesto de que conocemos *todos los datos de todos los sujetos*. Lo mismo sucede con los programas de ordenador (o de una hoja de cálculo como EXCEL); hay que introducir *todos los datos individuales*. Sin embargo éste no es siempre el caso: de varias muestras podemos conocer solamente (o simplemente tener a mano) los valores del *número de sujetos,* la *media* y la *desviación típica*. O podemos encontrar estos datos publicados en algún lugar.

En estos casos, sin más datos de cada muestra (solamente n, M, σ), podemos resolver el análisis de varianza, y además de una manera muy simple; de hecho ya lo hemos visto en los apartados anteriores aunque en todos los ejemplos disponíamos de todos los datos individuales.

Cuando no tenemos los datos de todos los sujetos, lo único que hay que tener previsto es cómo combinar medias y desviaciones típicas, y es lo que vamos a exponer aquí (las fórmulas están recogidas y demostradas en el anexo III).

Vamos a verlo con dos ejemplos; en uno tenemos en cada grupo un número distinto de sujetos y en otro tenemos muestras de idéntico tamaño.

5.1. Cuando el número de sujetos es distinto en cada grupo

Datos en un ítem de un cuestionario de evaluación de una universidad por los alumnos *(valoración global de los profesores*, en una escala de 1 a 5) en tres tipos de carreras de la misma universidad (tabla 5).

	A	B	C
n	1112	1214	400
Media	3.48	3.97	4.34
Desviación típica	1.09	1.06	1.10

Tabla 5

Lo primero que hacemos es calcular la media total (media *ponderada* por el número de sujetos en cada grupo):

$$M_t = \frac{\Sigma nM}{\Sigma n} = \frac{(1112x3.48)+(1214x3.97)+(400x4.34)}{1112+1214+400} = \frac{104225.34}{2726} = 3.8244$$

Suma de Cuadrados *entre* los grupos = $\Sigma n(M - Mt)^2$
= $[1112\ (3.48 - 3.824)^2] + [1214(3.97 - 3.824)^2] + [400\ (4.34 - 3.824)^2] =$ 264.8525

Suma de Cuadrados *dentro* de los grupos =
$\Sigma n\sigma^2 = (1112 \text{ x } 1.09^2) + (1214 \text{ x } 1.06^2) + (400 \text{ x } 1.10^2) =$ 3169.2176

Los resultados los tenemos en la tabla 6:

Origen de la variación	*Suma de Cuadrados* (SC)	*Grados de libertad* (gl)	*Cuadrados Medios* $CM = \frac{SC}{gl}$	$F = \frac{CM_{entre}}{CM_{dentro}}$
Entre grupos	264.8525	3 -1 = 2	132.426	113.78 ($p < .01$)
Dentro de los grupos	3169.2176	2726 - 3 = 2723	1.1638	

Tabla 6

Podemos afirmar *con mucha seguridad* que los grupos pertenecen a poblaciones distintas por lo que respecta a cómo juzgan la calidad global del profesorado; el criterio de clasificación (variable independiente, facultades) tiene que ver con la variable dependiente, aunque está relación no es grande ($\eta^2 = .08$). Los *contrastes posteriores* adecuados serían los de Scheffé (muestras de tamaño muy desigual).

Lo que no hemos hecho es calcular la *suma de cuadrados del total;* nos puede interesar para calcular η^2, o para comprobar que las sumas de cuadrados están bien hechas, o para presentar completa la tabla de resultados. Podemos hacerlo sin dificultad, pero antes necesitamos la varianza de los totales (de las tres muestras juntas).

Para combinar varianzas utilizamos esta fórmula[40]: $\sigma_t^2 = \frac{\Sigma n(M^2+\sigma^2)}{\Sigma n} - M_t^2$ [33]

$$\sigma^2_{A+B+C} = \frac{1112(3.48^2+1.09^2)+1214(3.97^2+1.06^2)+400(4.34^2+1.10^2)}{2726} -24^2 = 1.2594$$

Esta varianza de los totales multiplicada por el número total de sujetos nos da la suma de cuadrados del total = (1.2594)(2726) = 3433.12.

[40] Esta y otras fórmulas para combinar medias y varianzas están explicadas en el Anexo III.

Si sumamos las dos sumas de cuadrados *entre* y *dentro* tenemos 264.8525 + 3169.2176 = 3434. Las dos sumas de cuadrados del total no son exactamente idénticas (calculadas por caminos distintos con distinto redondeo de decimales) por el diferente redondeo de decimales en cada paso, pero la diferencia es negligible.

5.2. Cuando el número de sujetos es el mismo en cada grupo

El procedimiento es similar, aunque algo más sencillo. Vamos a utilizar los datos del primer ejemplo (tabla 1, reproducidos en la tabla 7).

	A	B	C
n	7	7	7
Media	11.714	15.428	8.571
Desv. típica	3.917	2.921	4.237

Tabla 7

Como cálculo auxiliar previo nos interesa la *varianza de las medias,* $\sigma^2_M = 7.854$

Esta varianza, multiplicada por el número total de sujetos (N = 21) nos va dar la suma de cuadrados *entre* los grupos, y la suma de las tres varianzas de los grupos multiplicada por el número de sujetos que hay en cada grupo (n = 7) nos va dar la suma de cuadrados *dentro* de los grupos.

Podemos calcular la razón F directamente:

$$F = \frac{\dfrac{\textit{suma de cuadrados entre}}{\textit{grados de libertad entre}}}{\dfrac{\textit{suma de cuadrados dentro}}{\textit{grados de libertad dentro}}} = \frac{\dfrac{N\sigma^2_M}{k-1}}{\dfrac{n\Sigma\sigma^2}{N-k}} = \frac{\dfrac{(21)(7.854)}{2}}{\dfrac{(7)(41.8273)}{21-3}} = \frac{\dfrac{164.934}{2}}{\dfrac{292.79}{18}} = \frac{82.467}{16.266} = 5.069$$

Tenemos el mismo resultado que hemos obtenido en tabla 2.

Como antes, no nos ha hecho falta calcular la *suma de cuadrados de los totales*. Nos puede interesar calcularla para comprobar que las sumas de cuadrados *entre* y *dentro cuadran* o para presentar completa la tabla de resultados.

Si vamos a calcular la *suma de cuadrados total,* nos va a hacer falta de nuevo la varianza de los totales, que multiplicada por N nos va a dar la suma de cuadrados total. Como se trata de muestras de *idéntico tamaño,* la fórmula [33] queda simplificada así:

$$\sigma^2_t = \frac{\Sigma M^2 + \Sigma\sigma^2}{k} - M^2_t \qquad [34]$$

La media total (M_t) , cuando las muestras son de idéntico tamaño, es igual a la *media de las medias* = 11.90

En nuestro caso:

$$\sigma^2_{A+B+C} = \frac{11.714^2 + 15.428^2 + 8.571^2 + 3.917^2 + 2.921^2 + 4.2378^2}{3} - 11.90^2 = 21.902$$

La *suma de cuadrados de los totales* será = (21 = N)(21.902) = 459.95 que es *casi* (por no utilizar en todos los casos los mismos decimales) igual a las sumas de cuadrados obtenidas antes (164.934 + 292.79 = 458, tabla 2).

Queda claro que a partir del *número de sujetos, la media y la desviación típica de cada muestra* (datos que con frecuencia están disponibles, porque los hemos calculado previamente, están publicados, etc.) es rápido y sencillo hacer todos los cálculos del análisis de varianza, sin necesidad de partir de todas las puntuaciones individuales (ni de utilizar un programa de ordenador).

También con sólo estos datos, N, M y σ, disponemos de programas de Internet que nos resuelven el análisis de varianza (direcciones en el último apartado); en estos programas de Internet la desviación típica que hay que introducir es la de la población (σ_{n-1}, dividiendo por N-1).

6. Análisis de varianza para *dos* muestras independientes

Normalmente cuando tenemos dos muestras independientes utilizamos el contraste de medias *(t de Student),* pero podemos utilizar igualmente el análisis de varianza con idénticos resultados. En el caso de *dos muestras* independientes tenemos que $t = \sqrt{F}$; con ambos procedimientos llegamos a los mismos resultados y a las mismas conclusiones[41].

6.1. Utilizando las desviaciones de las muestras

Desarrollamos el procedimiento con un ejemplo (tabla 9).

Las operaciones son las mismas ya vistas a propósito de más de dos grupos (resultados en la tabla 8). Las operaciones para grupos de distinto tamaño son también válidas cuando son del mismo tamaño.

Grupo A	Grupo B
22	12
18	16
24	10
22	10
16	4
18	6
13	17
18	14
19	14
22	10
n_1 = 10	n_2 = 10
M_1 = 19.2	M_2 = 11.3
σ_1 = 3.156	σ_2 = 3.95

Tabla 8

Cálculos previos

1º Calculamos la media y desviación típica (de la muestra; dividiendo por N) de cada uno de los dos grupos;

2º Calculamos la desviación típica de los *totales,* de $n_1 + n_2 = N = 20$;

$\sigma_{total} = 5.328$

3º Calculamos la *desviación típica de las dos medias* (como si se tratara de un grupo de dos sujetos), $\sigma_M = 3.95$

Sumas de Cuadrados (numerador de las varianzas):

1. Del *total:* $N\sigma_t^2 = (20)(5.328)^2 =$ 567.75

2. *Dentro* de los grupos (SC_{dentro})
   de idéntico tamaño: $n(\sigma_1^2 + \sigma_2^2) = (10)(3.156^2 + 3.95^2) =$ 255.63
   de distinto tamaño: $n_1\sigma_1^2 + n_2\sigma_2^2 = (10 \times 3.156^2) + (10 \times 3.95^2) =$ 255.63

---

41 Cabe preguntarse si cuando se enseña y aprende el contraste de medias *(t de Student)* no sería preferible comenzar con el análisis de varianza, que se podría después aplicar con menor dificultad de comprensión a otras situaciones. De hecho la explicación dada en la *introducción al análisis de varianza* está referida al caso de dos muestras por ser más fácil de captar. Una limitación de utilizar el análisis de varianza con sólo dos muestras puede estar en que los valores de F se refieren solamente a pruebas bilaterales (Guéguen. 1997), pero también es verdad que son las pruebas que habitualmente utilizamos.

3. *Entre* los grupos (SC_{entre}):

de idéntico tamaño: $N\sigma^2_M = (20)(3.95)^2 =$ 312.05

de distinto tamaño: $n_1(M_1 - M_t)^2 + n_2(M_2 - M_t)^2$

$= [10(19.2\text{-}15.25)^2] + [10(11.3\text{-}15.25)^2] =$ 312.05

Los *grados de libertad (denominador de las varianzas)* son:

del *total* = N-1 ($N = n_1 + n_2$): 20-1= 19

dentro de los grupos = N - número de grupos: 20-2= 18

entre los grupos = número de grupos - 1: 2-1= 1

Origen de la variación	*Suma de Cuadrados* (SC)	*Grados de libertad* (gl)	*Cuadrados Medios* varianza = $\frac{SC}{gl}$	$F = \frac{CM_{entre}}{CM_{dentro}}$
Entre de los grupos	312.05	1	312.05	21.97 ($p < .01$)
Dentro de los grupos	255.63	18	14.20	
Variabilidad total	567.75	19		

Tabla 9

Con sólo dos grupos de *idéntico tamaño* los Cuadrados Medios *entre* los grupos se pueden calcular directamente de manera más sencilla (coinciden con la Suma de Cuadrados, numerador de la varianza, porque los grados de libertad, el denominador, es igual a 1):

$$CM_{entre} = \frac{n(M_1 - M_2)^2}{2} \quad [33] \qquad = \frac{10(19.2 - 11.3)^2}{2} = 312.05$$

Si calculamos la t de Student: $t = \dfrac{19.2 - 11.3}{\sqrt{\dfrac{3.156^2}{9} + \dfrac{3.95^2}{9}}} = 4.687$

y $4.687^2 = 21.967$ pues $t^2 = F$ (cuando sólo tenemos dos muestras).

6.2. Utilizando las desviaciones de las poblaciones

Si calculamos las varianzas de las poblaciones (σ^2_{n-1}) en vez de las varianzas de las muestras (σ^2_n) podemos calcular directamente y con toda facilidad los *cuadrados medios* (varianzas *entre* y *dentro* de los grupos) y la razón F, tal como hemos visto antes y tenemos resumido en la fórmula [6] y que repetimos aquí:

$$F = \frac{(n)(\text{varianza de las medias } \sigma^2_{n\text{-}1})}{\text{media de las varianzas de los grupos } (= \Sigma\sigma^2_{n-1} / k)} \qquad [6]$$

1. *Cuadrados medios dentro de los grupos*

Necesitamos las varianzas de los grupos dividiendo por N-1; podemos calcularlas directamente si se trata de pocos sujetos, pero si ya disponemos de las varianzas dividiendo por N y se trata de muchos sujetos y habría que introducir los datos en una hoja de cálculo o son datos que encontramos publicados o que tenemos guardados de otras ocasiones, podemos utilizar la fórmula [2], que a modo de ejemplo vamos a aplicar aquí:

$$\sigma^2_{n-1} = \frac{n\sigma^2_n}{n-1} \qquad [2]$$

Grupo A: $\sigma_n = 3.156$ $\qquad \sigma^2_{n-1} = \frac{(10)(3.156^2)}{10-1} = 11.067$

Grupo B: $\sigma_n = 3.95$ $\qquad \sigma^2_{n-1} = \frac{(10)(3.95^2)}{10-1} = 17.336$

$CM_{dentro} = \Sigma\sigma^2_{n-1} / k = \frac{11.067 + 17.336}{2} = 14.20$

2. *Cuadrados medios entre los grupos*

Para calcular los cuadrados medios *entre* grupos necesitamos la varianza (σ^2_{n-1}) de las dos medias:

σ^2_{n-1} de M_1 (19.2) y M_2 (11.3) = $5.586^2 = 31.205$

Cuadrados medios entre los grupos = n x σ^2_{n-1} =(10)(31.205) = 312.05

3. Y finalmente $\quad F = \frac{n \text{ x } \sigma^2_{n-1} \textit{(de las medias)}}{\Sigma\sigma^2_{n-1} \textit{(de los grupos)/número de grupos}} = \frac{312.05}{14.20} = 21.97$

Hemos llegado a los mismos resultados que tenemos en la tabla 9.

Aunque en este caso se trataba de solamente de dos grupos, el procedimiento es igualmente válido para cualquier número de grupos de idéntico tamaño.

7. Cómo presentar los resultados del análisis de varianza

Es importante presentar los resultados con claridad. Las tablas de resultados convencionales que se hacen en cada análisis de varianza son claras para presentar cada análisis de varianza, sobre todo si se trata de un solo análisis o de unos pocos, pero cuando se comparan *varias muestras en una serie de variables,* hay que buscar un método de presentación de conjunto que de manera casi intuitiva facilite la comprensión e interpretación de los resultados tanto del análisis de varianza como de los contrastes posteriores.

En el ejemplo puesto como ilustración (tabla 10)[42] se han comparado entre sí cinco muestras en nueve variables (hay por lo tanto nueve análisis de varianza).

En la *primera columna* aparecen el nombre de las variables y los valores de F, p y η^2 de cada análisis. Aunque estos datos se pueden presentar de diversas maneras, lo que no debe faltar, siguiendo las recomendaciones de la APA (2001), es algún indicador del tamaño del efecto o algún coeficiente de asociación (como η^2 en el ejemplo de la tabla 11) (APA, 2001).

En la *segunda columna* están los grupos (dos muestras normativas) cuyas medias son significativamente mayores que las medias de las muestras puestas *en la misma fila*. Aparecen también los valores del contraste de Scheffé y la probabilidad asociada a estos valores.

Los datos descriptivos de las muestras (número de sujetos, media y desviación) pueden figurar en otra tabla distinta.

[42] Tomado de Gismero, Elena (1995). *La conducta asertiva y su relación con la anorexia nerviosa.* Madrid: Universidad Pontificia Comillas. En esta presentación hubiera sido útil incluir el tamaño del efecto.

De manera intuitiva se ve qué muestras tienen en general medias mayores (segunda columna) y qué muestras tienden a tener medias más bajas.

VARIABLE	*muestras con medias más altas*	*muestras con medias más bajas*		
EXPRESIÓN DE DESACUERDO $F = 5.29, \eta^2 = .114$ $p < .001$	CONTROL	ANOREXIA 3.407, $p < .03$		TERAPIA 3.471, $p < .02$
	DIETA			TERAPIA 2.885 $p < .08$
AUTOAFIRMACIÓN ANTE EXTRAÑOS $F = 3.18$, $\eta^2 = .07$ $p < .05$	DIETA	ANOREXIA 3.05, $p < .06$		
AUTOEXPRESIÓN SOCIAL $\eta^2 = .12$ $F = 5.632$ $p < .001$	CONTROL	ANOREXIA 3.984, $p < .01$		
	DIETA	ANOREXIA 3.474, $p < .02$		
ASERTIVIDAD GENERAL $F = 6.34$, $\eta^2 = .13$ $p < 0001$	CONTROL	ANOREXIA 4.1657, $p < .001$		
	DIETA	ANOREXIA 3.658, $p < .05$		
SATISFACCIÓN PROPIO CUERPO $F = 7.07$, $\eta^2 = .15$ $p < 0001$	CONTROL	ANOREXIA 4.786, $p < .001$		
	DIETA	ANOREXIA 3.474, $p < .05$		
	TERAPIA	ANOREXIA 3.483, $p < .02$		
AUTOCONCEPTO $F = 23.22$, $\eta^2 = .36$ $p < 0001$	CONTROL	ANOREXIA 4.818, $p < .0001$	BULIMIA 2.826, $p < .10$	TERAPIA 3.931, $p < .01$
	DIETA	ANOREXIA 8.149, $p < .0001$	BULIMIA 3.579, $p < .02$	TERAPIA 4.70, $p < .001$
COMPRENSIÓN PARENTAL $F = 23.22$, $\eta^2 = .21$ $p < 0001$	CONTROL	ANOREXIA 3.393, $p < .03$	BULIMIA 3.547, $p < .01$	TERAPIA 5.24, $p < .0001$
	DIETA	ANOREXIA 3.313, $p < .04$	BULIMIA 3.587, $p < .01$	TERAPIA 5.69, $p < .0001$
ACEPTACIÓN SOCIAL $F = 7.71$, $\eta^2 = .16$ $p < 0001$	CONTROL	ANOREXIA 5.101, $p < .0001$		
	DIETA	ANOREXIA 4.389, $p < .01$		
AUTOCONCEPTO $F = 24.44$, $\eta^2 = .37$ $p < 0001$	CONTROL	ANOREXIA 8.03, $p < .0001$	BULIMIA 3.306, $p < .04$	TERAPIA 4.54, $p < .001$
	DIETA	ANOREXIA 8.04, $p < .0001$	BULIMIA 3.83, $p < .01$	TERAPIA 5.02, $p < .0001$

Tabla 10

En este ejemplo (con variables en el ámbito de la *asertividad* y del *autoconcepto)* se puede observar, casi de un golpe de vista, que dos grupos, el de *control* (N = 57) y *dieta* (N = 33) no difieren entre sí y a la vez ambas muestras superan en casi todas las variables al grupo *anorexia* (N = 45) y en menos variables (posiblemente debido al menor número de sujetos en estas muestras) a las muestras *bulimia* (N = 12) y *terapia* (N = 33); la muestra *anorexia* es la que aparece en peores condiciones en estas variables[43].

Debajo del nombre de cada grupo está puesto el valor de Scheffé y su probabilidad; hubiera quedado más completa la información añadiendo el *tamaño del efecto* correspondiente a cada contraste.

Otra manera de presentar los resultados es la que figura en la tabla 12. Además de presentar la tabla habitual de resultados y otra información relevante (puede ser suficiente una información semejante a la puesta en la primera columna de la tabla 11), si se quiere poner el énfasis en la *magnitud de las diferencias* entre los grupos, se puede hacer una tabla en la que figuren las diferencias estadísticamente significativas (la probabilidad *p* corresponde en este caso a los contrastes de Scheffé) y los tamaños del efecto al comparar las medias de dos en dos. Un ejemplo puede ser el de la tabla 11, comparando alumnos de diversas facultades en *la importancia que se da a tener puestos de autoridad* en distintas profesiones.[44]

	Medias más bajas					
Medias más altas	*Pisco-pedagogía*	*Enfermería*	*Derecho y Económicas*	*Ingeniería*	*Trabajo Social*	*Filosofía Teología*
Derecho y Económicas	p = .001 d = .56	p = .000 d = .65	---	---	p = .000 d = .67	p = .04 d = .50
Ingeniería		p = .16 d = .45	---	---	p = .096 d = .47	

Tabla 11

Esta información puede ir acompañada de algún gráfico como el de la figura 1 en el que se aprecian con facilidad las diferencias entre las medias.

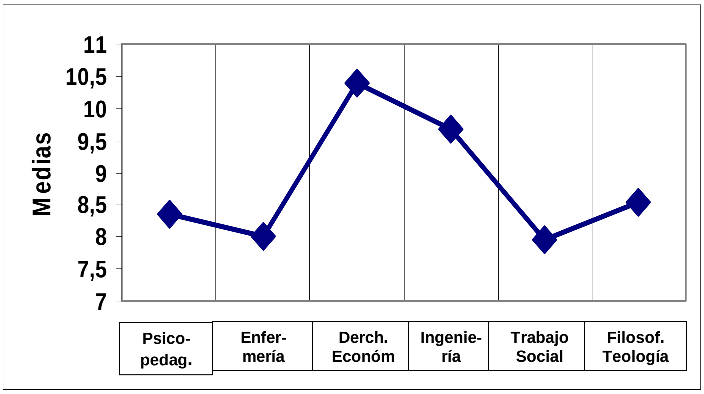

Figura 1

Otra manera de presentar un resumen de los resultados es ésta:

1º Se ordenan las medias de más a menos
2º Se unen con la misma línea continua las medias que no difieren entre sí

---

[43] En un trabajo académico (como una tesis doctoral) las tablas completas de los resultados de *cada* análisis de varianza pueden ir en un anexo; lo que se presenta en esta tabla es un resumen significativo con la información que se va a interpretar o comentar..

[44] Datos tomados de la tesis doctoral de Rufino Meana *La experiencia subjetiva de sentido y su relación con variables psicológicas y sociodemográficas,* Universidad Pontifica Comillas (2003).

Por ejemplo (ejemplo ficticio):

Medias de cuatro grupos *ordenadas de más a menos*: I III II IV

-No difieren entre sí el I y el III, el III y el II y el II y el IV (unidos por línea continua)
-El I supera al II y al IV y el III solamente al IV (no unidos por línea continua).

También caben cuadros y resúmenes en los que se pueden incluir comentarios *cualitativos* como complemento a los datos cuantitativos.

8. El análisis de varianza en programas informáticos y en Internet

8.1. Análisis de varianza para muestras independientes en EXCEL y en el SPSS

Para utilizar estos recursos hay ya manuales específicos, aquí nos limitamos a dar algunas indicaciones. Para utilizar EXCEL o el SPSS (y en general cualquier paquete informático con programas estadísticos) necesitamos introducir en primer lugar todos los datos de todos los sujetos; no podemos operar a partir solamente del número de sujetos, media y desviación típica de cada grupo. En este apartado nos referimos exclusivamente al análisis de varianza para varias muestras independientes.

EXCEL. El análisis de varianza para muestras independientes se denomina en EXCEL *análisis de varianza de un factor*. Además de los datos descriptivos de cada grupo nos da la tabla final de resultados, pero no los contrastes posteriores.

SPSS. En *analizar* (barra superior) tenemos la opción *comparar medias* y allí *ANOVA de un factor*. De los *contrastes posteriores* comentados (*post hoc* en el cuadro de diálogo) tenemos, entre otras opciones, los contrastes de Scheffe, Tuckey, Duncan (comentado en el apartado dedicado al contraste de Newman-Keuls del que es una variante; los contrastes de Duncan y Newman-Keuls para muestras de idéntico tamaño son algo más liberales que el de Tukey), Games-Howell (tamaño distinto y varianza desiguales, puede ser una buena alternativa a Scheffé) y Dunnett.

Otros cálculos complementarios, como el coeficiente η^2 y tamaño del efecto, se completan fácilmente con una calculadora teniendo las fórmulas a la vista.

8.2. Recursos en Internet relacionados con el Análisis de Varianza

En Internet disponemos además de numerosos recursos para llevar a cabo análisis estadísticos. Pueden ser especialmente útiles en varias situaciones:

a) Cuando no tenemos disponibles hojas de cálculo como EXCEL, o programas informáticos como el SPSS.

b) Cuando de las distintas muestras tenemos disponibles el número de sujetos, las medias y las desviaciones; sólo con estos datos no podemos utilizar programas como el SPSS que requieren la introducción de todos los datos de todos los sujetos, pero sí podemos llevar a cabo un análisis de varianza con toda facilidad.

Por lo que respecta al análisis de varianza en Internet hacemos una selección de páginas Web (hay muchas más) útiles con distintas finalidades[45].

[45] Las direcciones de Internet son con frecuencia inestables pero disponiendo de información sobre títulos y autores no es difícil encontrar estas u otras con información semejante en los *buscadores* de Internet.

8.2.1. Test de Bartlett para comprobar la homogeneidad de varianzas

Si queremos verificar si las varianzas no difieren significativamente podemos utilizar *el test de Bartlett,* sencillo y programado en Internet, http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/BartletTest.htm

Basta introducir de cada muestra el número de sujetos y las varianzas (no las desviaciones típicas) y si p > .05 podemos aceptar la homogeneidad de varianzas (esta dirección, con otras muchas posibilidades de análisis estadísticos, se encuentra en la *Home Page* de Hossein Arsham http://home.ubalt.edu/ntsbarsh/Business-stat/home.html en JavaScript E-labs Learning Objects)

8.2.2. Tablas de la F de Snedecor, Tukey, Dunnett y Bonferroni

Las tablas convencionales que suelen estar en libros de texto también están disponibles en Internet.

Tablas de la F de Snedecor

Las tablas de la F de Snedecor se encuentran en muchas direcciones de Internet, con distintas presentaciones y para distintos niveles de confianza, por ejemplo:

ALEXEI, SHAROV. Department of Entomolgy, Virginia Tech, Blacksburg, VA *On-Line Lectures*, http://www.ento.vt.edu/~sharov/PopEcol/tables/f005.html para α = .05, .01 y .001

GERSTMAN, B. BURT (2003). *StatPrimer*, http://www.sjsu.edu/faculty/gerstman/StatPrimer/ en *Probability Tables* para α = .10, .05, .025, .01 y .001

SIX SIGMA Reference Tables http://www.micquality.com/reference_tables/index.htm para α = .01, .025, .05, .10 y .25

STATSOFT, INC., Distribution tables, http://www.statsoft.com/textbook/sttable.html tablas para α = .10, .05, .025 y .01

Tablas de Tukey

http://www.stat.duke.edu/courses/Spring98/sta110c/qtable.html

BISSONNETTE, VICTOR L., Berry College
http://fsweb.berry.edu/academic/education/vbissonnette/ *Some Useful Statistical Tables,*
http://fsweb.berry.edu/academic/education/vbissonnette/tables/posthoc.pdf

CHAVES, COSME *Introducción a la Estadística* http://costaricalinda.com/Estadistica/ (tablas)

Tablas de Dunnett

ARMSTRONG, J. SCOTt http://www.forecastingprinciples.com/tables.pdf Reprinted from Armstrong, J. Scott, *Long-Range Forecasting*. New York, John Wiley 1985 (reproducidas de la publicación original de Dunnett (niveles de confianza .05 y .01 para pruebas de una y dos colas)

LANE, DAVID M. HyperStat Online Statistics Textbook
(http://davidmlane.com/hyperstat/index.html)
http://davidmlane.com/hyperstat/table_Dunnett.html (en Introduction to Between-Subjects ANOVA→ Comparing means with a control)
Table Critical values for the Dunnett test http://www.watpon.com/table/dunnetttest.pdf

BISSONNETTE, VICTOR L., Berry College http://facultyweb.berry.edu/vbissonnette/ *Some Useful Statistical Tables,*
Table Critical values for the Dunnett test http://www.watpon.com/table/dunnetttest.pdf

CHAVES, COSME *Introducción a la Estadística* http://costaricalinda.com/Estadistica/ (tablas)

Tablas de Bonferroni

BISSONNETTE, VICTOR L., Berry College
http://fsweb.berry.edu/academic/education/vbissonnette/ *Some Useful Statistical Tables,*
Critical Values of Dunn's (Bonferroni) test (experimentwise $\alpha = .05$)

8.2.3. Probabilidades exactas de la razón F en Internet

Las probabilidades exactas de la razón F (y otros estadísticos) pueden verse al menos en estas direcciones:

a) THE CHINESE UNIVERSITY OF HONG KONG, DEPARTMENT OF OBSTETRICS AND GYNAECOLOGY, http://department.obg.cuhk.edu.hk/researchsupport/statmenu.asp; en *Statistics Tool Box* se busca *Statistical Tests* y allí *Statistical Significance*, o directamente http://department.obg.cuhk.edu.hk/researchsupport/F_Test.asp.

b) HYPERSTAT ONLINE TEXTBOOK (de David M. Lane) http://davidmlane.com/hyperstat/ en *Analysis Tools* buscar *Tables;* (hay varios programas; se puede ir directamente a http://davidmlane.com/hyperstat/F_table.html o a http://members.aol.com/johnp71/pdfs.html

c) SISA, *Simple Interactive Statistical Analysis* (Daan Uitenbroek PhD, Research and Statistical Consultancy, Hilversum, The Netherlands) http://www.quantitativeskills.com/sisa/ (en *significance testing*).

d) GRAPH PAD SOFTWARE http://graphpad.com/quickcalcs/PValue1.cfm

8.2.4. Cómo llevar a cabo un Análisis de Varianza en Internet

Aunque en las páginas seleccionadas nos fijamos en el análisis de varianza para muestras independientes, podemos encontrar también programas para hacer otros tipos de análisis de varianza.

Esta selección (es sólo una selección pues hay muchos más programas en Internet) la dividimos en dos apartados: a) cuando tenemos disponibles medias y desviaciones de las muestras y b) cuando vamos a introducir todos los datos de todos los sujetos.

8.2.4.1. A partir del número de sujetos, medias y desviaciones de las muestras

Estos programas son muy útiles porque con frecuencia no disponemos de los datos de cada sujeto y los únicos datos disponibles de cada muestra (a veces en obtenidos en análisis previos) son el número de sujetos, la media y la desviación típica; con sólo estos datos no podemos acudir a EXCEL o al SPSS.

Seleccionamos tres programas muy sencillos (hay muchos más); nos basta introducir el número de sujetos, la media y la desviación típica de cada muestra.

a) PEZZULLO, JHON C. *Web Pages that Perform Statistical Calculations*. http://members.aol.com/johnp71/javastat.html (éste es el modo de citar esta fuente).

Analysis of Variance from Summary Data
http://members.aol.com/johnp71/anova1sm.html (en *Web Pages that Perform Statistical Calculations!* http://members.aol.com/johnp71/javastat.html)

La desviación típica que hay que introducir es la de la *población* (dividiendo por N-1); es lo normal en los programas de Internet. Si el número de sujetos es muy grande las dos desviaciones apenas difieren (dividir por 200 o por 199 no altera gran cosa los resultados) pero si se tiene calculada la desviación típica de la muestra (σ_n) es muy sencillo calcular la de la población (σ_{n-1}):

$$\sigma_{n-1} = \sqrt{\frac{\mathbf{N}\sigma_{\mathbf{n}}^2}{\mathbf{N}-1}}$$

Este programa no calcula los contrastes posteriores.

b) DANIEL SOPER.com http://www.danielsoper.com/default.aspx en *Statistics calculators*→ ANALYSIS OF VARIANCE →One-Way ANOVA from Summary Data

Este programa no calcula los contrastes posteriores

c) DEPARTMENT OF OBSTRETRICS AND GYNAECOLOGY, THE CHINESE UNIVERSITY OF HONG KONG http://department.obg.cuhk.edu.hk/ResearchSupport/OWAV.asp La referencia citada en esta fuente es: Armitage P. Statistical Methods in Medical Research (1971). Blackwell Scientific Publications. Oxford. P.189-207

Este programa da además simultáneamente el contraste de Tukey y también están programados los contrastes posteriores de Scheffé en http://department.obg.cuhk.edu.hk/ResearchSupport/Least_sig_diff_Scheffe.asp (en el menú de la izquierda, en *categories* buscar *group differences*).

8.2.4.2. Introduciendo los datos de todos los sujetos

Cuando hay que introducir todos los datos de todos los sujetos, lo habitual es utilizar EXCEL o el SPSS, pero también disponemos de programas en Internet.

Un programa sencillo lo tenemos en KIRKMAN, T.W , College of Saint Benedict/Saint Johns University [http://www.csbsju.edu/] http://www.physics.csbsju.edu/stats/anova.html; el índice de análisis estadísticos de este centro está en http://www.physics.csbsju.edu/stats/

Los datos se pueden introducir de dos maneras, copiándolos de una tabla o introduciéndolos directamente. Este programa no calcula los contrastes posteriores.

Otro programa para llevara cabo un análisis de varianza introduciendo los datos de todos los sujetos (se pueden copiar y pegar de una tabla WORD o de EXCEL) es LOWRY, RICHARD, VASSARSTATS: Web Site for Statistical Computation,Vassar College, Poughkeepsie, NY, USA; http://faculty.vassar.edu/lowry/VassarStats.html (ANOVA en el menú de la izquierda); incluye los contrastes posteriores de Tukey.

8.2.4.3. Contrastes posteriores

Los contrastes de **Bonferroni** están programados en GraphPad, San Diego, CA [http://graphpad.com/] *Free Online Calculators for Scientists* http://graphpad.com/quickcalcs/posttest1.cfm (en *how the calculations are performed* se especifica que se trata del test de Bonferroni) pero se trata de un contraste considerado demasiado conservador (tiene poca *potencia* para rechazar la Hipótesis Nula cuando realmente es falsa (Hancock y Klockars, 1996; Jaccard, 1998)[46]. No se trata propiamente de contrastes posteriores, sino de las probabilidades equivalentes a .05 cuando hacemos múltiples contrastes, y tampoco es un procedimiento pensado específicamente para el análisis de varianza.

En DEPARTMENT OF OBSTRETRICS AND GYNAECOLOGY, THE CHINESE UNIVERSITY OF HONG KONG http://department.obg.cuhk.edu.hk/ResearchSupport/OWAV.asp tenemos programados los contrastes posteriores de **Scheffé**, **Tukey** y los correspondientes a la prueba de **Kruskal-Wallis** que es la alternativa no paramétrica (utilizando rangos) al análisis de varianza para muestras independientes.

9. Referencias bibliográficas

AMERICAN PSYCHOLOGICAL ASSOCIATION. (2001). *Publication Manual of the American Psychological Association* (5th edition). Washington, DC: Author.

ARSHAM, HOSSEIN, *Homepage* http://home.ubalt.edu/ntsbarsh/Business-stat/home.html (en *JavaScript E-labs Learning Objects*)

BISSONNETTE, VICTOR L., Berry College (2000). *Course Resources* http://fsweb.berry.edu/academic/education/vbissonnette/

BLACK, THOMAS R. (1999). *Doing Quantitative Research in the Social Sciences.* London: Sage.

CHAVES, COSME *Introducción a la Estadística* http://costaricalinda.com/Estadistica/

COHEN, JACOB (1988). *Statistical Power Analysis for the Behavioral Sciences*. Second Edition. Hillsdale, N.J.: Lawrence Elrbraum.

COOMBS, WILLIAM T.; ALGINA, JAMES and OLTMAN, DEBRA OLSON, (1996). Univariate and Multivariate Omnibus Hypothesis Tests Selected to Control Type I Error Rates When Population Variances Are Not Necessarily Equal. *Review of Educational Research, 66* (2), 137-179.

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

DALLAL, GERARD E. (2001) *The Little Handbook of Statistical Practice* (en *Multiple Comparisons*) http://www.tufts.edu/~gdallal/LHSP.HTM (consultado 5 de Octubre, 2007).

DEPARTMENT OF OBSTRETRICS AND GYNAECOLOGY, THE CHINESE UNIVERSITY OF HONG KONG, http://department.obg.cuhk.edu.hk/index.asp?scr=1024

DOWNIE, N.M. Y HEATH, R.W., (1971). *Métodos estadísticos aplicados*. Madrid: Ediciones del Castillo

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

GERSTMAN, B. BURT (2003). *StatPrimer*, http://www.sjsu.edu/faculty/gerstman/StatPrimer/

GISMERO, ELENA (1995). *La conducta asertiva y su relación con la anorexia nerviosa.* Madrid: Universidad Pontificia Comillas.

[46] Pueden verse los comentarios hechos antes a propósito de estos contrastes y en el Anexo II.

GORDON, LEONARD V. (1973). One-Way Analysis of Variance Using Means and Standard Deviations. *Educational and Psychological Measurement,* 12 1973; vol. 33: pp. 815 - 816

GRAPH PAD SOFTWARE *Online calculators for scientists* http://graphpad.com/quickcalcs/PValue1.cfm

GUÉGUEN, NICOLAS (1997). *Manuel de Statistique pour Psychologues*. Paris: Dunod.

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].

HANCOCK, GREGORY R. AND KLOCKARS, ALAN J., (1996). The Quest for α: Developments in Multiple Comparison Procedures in the Quarter Century Since Games (1971). *Review of Educational Research,* 66, (3). 269 - 306.

HEDGES, LARRY V. and OLKIN, INGRAM**,** (1985). *Statistical Methods for Meta-Analysis*. New York: Academic Press.

HINKLE, DENNIS E.; WIERSMA, WILLIAM and JURS, STEPHEN G. (1994). *Applied Statistics for the Behavioral Sciences*. Boston: Houghton-Mifflin.

HUCK, SCHUYLER W. and MALGADY, ROBERT G., (1978). Two-Way Analysis of Variance Using Means and Standard Deviations. *Educational and Psychological Measurement,* 38, 235-237.

HUCK, SCHUYLER W., CORMIER, WILLIAM H. AND BOUNDS, WILIAM G., (1974), *Reading Statistics and Research*, New York, Harper & Row

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

KESELMAN, H. J.; HUBERTY, CARL J.; LIX, LISA M.; OLEJNIK, STEPHEN; CRIBBIE, ROBERT A.; DONAHUE, BARBARA; KOWALCHUK, RHONDA K.; LOWMAN, LAUREEN L.; PETOSKEY, MARTHA D. and KESELMAN, JOANNE, C. (1998). Statistical Preferences of Educational Researchers: An Analysis of Their ANOVA, MANOVA and ANCOVA Analyses. *Review of Educational Research,* 68 (3), 350-385.

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

KIRKMAN, T.W. (1996) College of Saint Benedict/Saint Johns University. *Statistists to Use*. http://www.physics.csbsju.edu/stats/

KLOCKARS, ALAN J. and HANCOCK, GREGORY R. (1998). A More Powerful Post Hoc Multiple Comparison Procedure in Analysis of Variance. *Journal of Educational and Behavioral Statistics*, Vol. 23 (3).

KLOCKARS, ALAN J. and SAX, GILBERT**,** (1986). *Multiple Comparisons: Sage* University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park: Sage.

LANE, DAVID M. *HyperStat Online Statistics Textbook* (*Last updated: 3/16/2009*)http://davidmlane.com/hyperstat/index.html (consultado 07, 010, 2007)

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

MARTÍNEZ GARZA, ÁNGEL**,** (1988). *Diseños Experimentales*. México: Trillas.

NUNNALLY, JUM C. and BERNSTEIN, IRA H. (1994). *Psychometric Theory*, 3rd. ed., New York, McGraw-Hill.

PERNEGER, THOMAS V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal* 1998;316:1236-1238 http://www.bmj.com/cgi/content/full/316/7139/1236

PEZZULLO, JHON C. *Web Pages that Perform Statistical Calculations*. http://members.aol.com/johnp71/javastat.html

RODRIGUES, AROLDO, (1977). *Investigación experimental en psicología y educación*. México: Trillas.
ROSENTHAL, ROBERT and ROSNOW, RALPH L. (1991). *Essentials of Behavioral Research, Methods and Data Analysis.* Boston: McGraw-Hill.
RUNYON, RICHARD P. y HABER, AUDREY, (1984) *Estadística para las Ciencias Sociales,* México, Fondo Educativo Interamericano.
SHAROV, ALEXEI (1996). *Quantitative Population* Ecology, *On-Line Lectures*, http://home.comcast.net/~sharov/PopEcol/
SIX SIGMA *Reference Tables* http://www.micquality.com/reference_tables/index.htm
STATSOFT, Inc. (2007). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/stathome.html.
THALHEIMER, WILL & COOK, SAMANTHA. (2002). *How to Calculate Effect Sizes From Published Research Articles: A Simplified Methodology.* A Work-Learning Research Publication Available online: http://www.learningaudit.com/white_papers/effect_sizes/Effect_Sizes_pdf4.pdf (consultado 30 de Mayo, 2011).
TOOTHAKER, LARRY E., (1993). *Multiple Comparison Procedures,* Sage University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park: Sage.
UITENBROEK, DAAN SISA, *Simple Interactive Statistical Analysis*. http://www.quantitativeskills.com/sisa/
WILDT, ALBERT R. and AHTOLA, OLLI T., (1978). *Analysis of Covariance,* Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills: Sage.