Análisis de varianza
Anexos

© Pedro Morales Vallejo
Universidad Pontificia Comillas,
Facultad de Ciencias Humanas y Sociales
(Última revisión: 14 de Noviembre de 2009)

Índice

Anexo I: Por qué no podemos utilizar la *t de Student* para hacer todas las comparaciones posibles cuando tenemos más de dos grupos 3

Anexo II: El test de *Bonferroni:* procedimiento para hacer todas las comparaciones posibles entre pares de grupos sin llevar a cabo un análisis de varianza 5

Anexo III: Cálculo de la media y de la desviación típica del *total* a partir de las medias y desviaciones típicas de las muestras 7

1. *Media de los totales* 7
    1.1. *Con muestras de tamaño distinto* 7
    1.2. *Con muestras de idéntico tamaño* 7
2. *Desviación típica de los totales* 7
    2.1. *Con muestras de distinto tamaño* 7
    2.2. *Con muestras de idéntico tamaño* 8
    2.3. *Deducción de la fórmula* 8

Anexo IV: Contrastes posteriores: la distribución q del *rango estudentizado* 10

Anexo V: Contrastes posteriores en diseños factoriales cuando la interacción es significativa 12

1. *Valores de k para consultar las tablas* 12
2. *Justificación de los nuevos valores de k* 13

Anexo VI: El número de sujetos en el análisis de varianza 17

1. Variables que intervienen en la determinación del tamaño de la muestra 17
2. Tamaño de cada muestra cuando comparamos dos grupos *(t de Student)* 19
3. Tamaño de la muestra cuando tenemos más de dos muestras *(análisis de varianza unifactorial)* 19
4. Tamaño de la muestra en los diseños factoriales 21

Anexo VII: Métodos *no paramétricos* de análisis de varianza 23

1. *Observaciones sobre los métodos no paramétricos* 23
2. *Métodos no paramétricos más importantes análogos al análisis de varianza* 23

Anexo VIII: Análisis de Varianza en Internet 25

Referencias bibliográficas 25

ANEXO I: Por qué no podemos utilizar la *t de Student* para hacer todas las comparaciones posibles cuando tenemos más de dos grupos

Quizás la manera más fácil de captarlo es mediante una *analogía* de fácil comprensión y que tiene que ver con la *distribución binomial*.

Imaginemos que tiramos *una* moneda al aire ¿Que probabilidades tenemos de que nos salga *una cara?* Si solamente tenemos una moneda tenemos dos resultados posibles, o nos sale *cara* o nos sale *cruz:*

		probabilidades:
Una posibilidad:	*cara*	una vez : p = .50 (50%)
Otra posibilidad:	*cruz*	una vez: p = .50 (50%)

Con sólo *dos* resultados posibles (el 100%) tenemos un 50 % de probabilidades de que nos salga *cara* y otro 50 % de probabilidades de que nos salga *cruz.*

Ahora vamos a tirar al aire *dos* monedas a la vez ¿Qué probabilidades tenemos de que nos salga *una cara?* Tenemos estos y solos posibles resultados:

	moneda primera	*moneda segunda*
una posibilidad:	*cara*	*cara*
otra posibilidad:	*cruz*	*cruz*
otra posibilidad:	*cara*	*cruz*
otra posibilidad:	*cruz*	*cara*

Con *dos monedas* tenemos cuatro posibles resultados (*cara-cara, cruz-cruz* y dos veces *cara-cruz).* ¿Qué probabilidades tenemos de obtener *al menos* una cara? tres de cuatro o p = 3/4 =.75 (un 75% de probabilidades de que nos salga *al menos* una cara).

Incidentalmente podemos observar que, simbolizando C como *cara* y X como *cruz* estos son los resultados posibles: *una vez* dos C, *una vez* dos X y *dos veces* XC (da lo mismo *cara-cruz* que *cruz-cara)* es decir, todas las posibilidades son:

$$XX + CC + 2CX$$

Dividiendo cada posible resultado por el número total de resultados distintos (4 en este caso) tenemos la probabilidad de que nos salga una determinada combinación *cara-cruz*. Si utilizamos los símbolos a y b posiblemente caeremos en la cuenta del parecido de la expresión anterior con el *cuadrado de un binomio:*

$$(a + b)^2 = a^2 + 2ab + b^2$$

El *exponente* del binomio sería en este caso el *número* de monedas. Si en vez de dos monedas, utilizamos tres, la resolución de $(a + b)^3$ nos dará cuáles pueden ser los resultados distintos y cuántas veces obtendríamos cada uno.

Al menos hemos visto que si estamos interesados en que nos salga *por lo menos una cara,* no es lo mismo arrojar al aire una moneda que dos o tres… A mayor número de monedas, mayor probabilidad de que *al menos* una nos salga cara. De manera *análoga* si comparamos *tres* grupos tenemos tres comparaciones posibles: el 1º con el 2º, el 1º con el 3º y el 2º con el 3º: en este caso las probabilidades de obtener por azar al menos *una t estadísticamente significativa* suben con respecto a comparar solamente dos grupos.

Si vamos a hacer tres comparaciones y establecemos un nivel de confianza de α =.05 ¿Qué probabilidad tenemos de obtener al menos una t significativa? Esta pregunta tiene que ver con las *probabilidades conjuntas,* tema en el que no entramos pero que puede quedar sugerido o ilustrado viendo todas las posibles diferencias estadísticamente significativas cuando tenemos tres grupos.

Podemos encontrar	*sólo* entre	el 1º y el 2º				
una diferencia	*sólo* entre	el 1º y el 3º				
significativa:	*sólo* entre	el 2º y el 3º				
	entre	el 1º y el 2º	*y*	entre el 1º y el 3º		
	entre	el 1º y el 2º	*y*	entre el 2º y el 3º		
	entre	el 1º y el 3º	*y*	entre el 2º y el 3º		
	entre	el 1º y el 2º	*y*	entre el 1º y el 3º	*y*	entre el 2º y 3º

La probabilidad de encontrar una *t significativa* no es ciertamente del 5%, son muchas más; de hecho, con tres grupos y operando con un nivel de confianza de .05, ya la probabilidad real de obtener al menos una diferencia significativa no es del 5% sino del 9.75%[1].

Algún autor lo explica de manera más sencilla (Kirk (1995): si tiramos al aire dos dados no es tan fácil que nos salgan la primera vez *dos seises;* pero si tiramos al aire 10 dados a la vez tenemos muchas más probabilidades de que nos salgan los *dos seises*…

De todas maneas recordamos que podemos comparar los grupos de dos en dos si se dan estas condiciones: 1º tenemos *hipótesis previas* y justificadas *antes de recoger lo datos* y 2º el número máximo de comparaciones legítimas no puede ser superior a los grados de libertad (= k-1; con tres grupos podaríamos hacer dos comparaciones y no las tres posibles).

[1] Una explicación más detallada sobre cómo obtener las probabilidades reales puede verse en McGuigan (6th Edition, 1994:140ss).

ANEXO II: El test de *Bonferroni:* procedimiento para hacer todas las comparaciones posibles entre pares de grupos sin llevar a cabo un análisis de varianza

Ya hemos indicado que cuando en el mismo planteamiento tenemos más de dos grupos, no podemos contrastar todos los posibles pares de medias entre sí porque aumenta la probabilidad de error. Hemos visto que con tres grupos un nivel de confianza de α = .05 equivale de hecho a un nivel de .0975.

Si deseamos hacer *todos los contrastes posibles de medias* podríamos hacerlos siendo más estrictos en el nivel de confianza sin necesidad de acudir al análisis previo de varianza.

El procedimiento más sencillo para ajustar el nuevo nivel de confianza es el denominado test de *Bonferroni*.[2] Este procedimiento consiste en dividir el nivel de confianza escogido (α = .05, por ejemplo, o α = .01) por el *número de posibles comparaciones*: la probabilidad resultante es la que debemos utilizar. Ya sabemos que el número de posibles comparaciones de k grupos tomados de dos en dos es igual a k(k-1)/2.

La fórmula general del nuevo nivel de confianza es:

$$\textit{nuevo valor} \text{ de } \alpha = \frac{\alpha}{\frac{k(k-1)}{2}}$$

α = probabilidad escogida (α = .05, .01, etc.)
k = número de grupos

Esta probabilidad es para *pruebas de una cola* o *unilaterales* (nos fijamos solamente en un extremo de la distribución); este valor de α habrá que dividirlo por 2 si utilizamos contrastes bilaterales como es usual.

Por ejemplo, si tenemos tres grupos (k = 3), nuestro nivel de confianza es α = .05, y queremos hacer *todas las comparaciones posibles entre grupos* (sin combinar las medias de varios grupos en una sola) el nuevo valor de α será:

$$\alpha = \frac{.05}{\frac{3(3-1)}{2}} = .0167$$

En las tablas de la distribución normal *(muestras grandes)* vemos que a una probabilidad de .0167 *(área menor)* le corresponde z = 2.39; bastante mayor que el 1.96 al que estamos acostumbrados (y que en pruebas de una cola es 1.64).

De manera más sencilla si disponemos de la probabilidad *exacta* de un contraste (fácilmente disponible en Internet) y la multiplicamos por el número de comparaciones, podemos ver si llega a .05 (suponiendo que éste es nuestro nivel de confianza) (Bland y Alman, 1995). Por ejemplo si obtenemos p = .0167 y tenemos tres comparaciones, nuestra probabilidad será (con un nivel de confianza de .05) p = (.0167)(3) = .05.

En opinión de varios autores este contraste no parece aconsejable porque es considerado *demasiado conservador;* es decir, tiene poca *potencia* para rechazar la Hipótesis Nula cuando realmente es falsa (Hancock y Klockars, 1996; Jaccard, 1998). Por otra parte estos contrastes no están pensados como contrastes posteriores al análisis de varianza sino para siempre que

[2] Carlo Emilio Bonferroni elaboró su teoría sobre la probabilidad (publicada en Florencia en 1936) al margen del análisis de varianza, pero en los años 60 se aplicó a los contrastes posteriores; es el contraste también conocido como de Dunn-Bonferroni. A veces se menciona a Bonferroni como *desconocido* (McGuigan, 1994); pero puede encontrarse citado en Hancock y Klockars (1996) y una breve reseña biográfica en http://www-history.mcs.st-andrews.ac.uk/Biographies/Bonferroni.html (en *The MacTutor History of Mathematics archive* http://www-history.mcs.st-andrews.ac.uk/index.html)

en el mismo planteamiento haya más de dos contrastes de medias (por ejemplo si comparamos dos grupos en una serie de variables) y también serían aplicables si tenemos un conjunto de coeficientes de correlación.

Una crítica bien razonada a los ajustes de Bonferroni puede verse en Perneger (1998) que merece la pena citarle literalmente:

> *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.*

Nuestra valoración personal es que se puede prescindir de los contrastes de Bonferroni con la conciencia tranquila porque son excesivamente conservadores y además esta opinión está avalada por autores relevantes.

Los contrastes de Bonferroni están programados en GraphPad (ver referencias bibliográficas) y también en el SPSS (en análisis de vaianza).

Más que contrastes, como si se tratara de alternativas a la t de Student, se trata de nuevos valores de p e*quivalentes a* α *= .05* cuando hacemos más de un contraste entre dos medias. Estos valores pueden verse en Internet, en BISSONNETTE, VICTOR L., Berry College http://fsweb.berry.edu/academic/education/vbissonnette/ *Some Useful Statistical Tables,* Critical Values of Dunn's (Bonferroni) test (experimentwise α = .05) http://fsweb.berry.edu/academic/education/vbissonnette/tables/dunns.pdf

ANEXO III: Cálculo de la media y de la desviación típica del *total* a partir de las medias y desviaciones típicas de las muestras

Frecuentemente lo que tenemos de las diversas muestras son los datos descriptivos básicos: el *número de sujetos* (N), la *media* (M) y la *desviación típica* (σ). Son datos que podemos ver en una publicación o que nosotros mismos guardamos[3]. Si a partir de estos datos queremos hacer un análisis de varianza con una simple calculadora, necesitamos además tener:

a) La *media del total* de todos los sujetos (al menos para calcular la suma de cuadrados *entre* grupos cuando los grupos son de tamaño desigual).

b) La *desviación típica* del total de las puntuaciones; en realidad necesitamos la *varianza,* que multiplicada por N (número total de datos, si se trata de muestras relacionadas N = n x k) nos dará la *suma de cuadrados del total*.

Tanto la *media total* como la *desviación típica o varianza de los totales* podemos calcularlos a partir de los datos descriptivos de las muestras, sin necesidad de disponer de todas las puntuaciones individuales que normalmente sí necesitaremos si vamos a utilizar un programa de ordenador (como el SPSS). Las fórmulas apropiadas ya las hemos ido viendo en varias ocasiones; ahora las repetimos de manera más ordenada y con una demostración.

1. *Media de los totales*

1.1. *Con muestras de tamaño distinto*

$$M_t = \frac{\Sigma nM}{\Sigma n(\text{o N total})} \qquad [1]$$

Se multiplica cada media por *su* número de sujetos, se suman estos productos y se divide por el número total de sujetos. Se trata de una media ponderada que no necesita una especial demostración.

1.2. *Con muestras de idéntico tamaño*

Cuando los grupos son de *idéntico tamaño,* la media total es simplemente *la media de las medias* (suma de las medias dividida por el número de medias):

$$M_t = \frac{\Sigma M}{k} \qquad [2]$$

2. *Desviación típica de los totales*

2.1. *Con muestras de distinto tamaño*

$$\sigma_t = \sqrt{\frac{\Sigma n(M^2 + \sigma^2)}{N} - M_t^2} \qquad [3]$$

σ_t es la desviación típica de los totales; como realmente nos interesa la varianza, podemos prescindir de la raíz cuadrada. n es el tamaño de cada grupo, que se multiplica por la suma de su media y desviación elevadas al cuadrado. La desviación típica la suponemos calculada dividiendo por N.

La fórmula [3] está tomada de McNemar (1962), pero es fácilmente demostrable como vamos a ver más adelante.

---

[3] A partir de estos datos (N, Media y desviación típica de cada muestra) también tenemos programado en Internet el análisis de varianza para varias muestras independientes (por ejemplo, los programas de Internet de Pezzulo y en Department of Obstretrics And Gynaecology, The Chinese University of Hong Kong; ver referencias bibliográficas).

2.2. *Con muestras de idéntico tamaño*

Cuando los grupos son de *idéntico tamaño* (como sucede siempre que tenemos muestras *relacionadas* y en muchos otros casos) la fórmula de la desviación típica de los totales queda simplificada; k es el número de grupos[4].

$$\sigma_t = \sqrt{\frac{\Sigma M^2 + \Sigma\sigma^2}{k} - M_t^2} \qquad [4]$$

En este caso, muestras de idéntico tamaño, la media de los totales ya sabemos que es *la media de las medias.*

Como se trata de calcular la *varianza* (y no la desviación típica), podemos prescindir de la raíz cuadrada. Estos valores pueden variar según se calculen por un método o por otro, debido al distinto número de decimales que se utilicen en cada caso, pero las diferencias no afectan apreciablemente a los cálculos posteriores. Aun así es preferible utilizar tres o cuatro decimales.

2.3. *Deducción de la fórmula*

¿De dónde viene esta fórmula [3] para combinar desviaciones típicas a partir de los valores de las desviaciones, medias y número de sujetos de los diversos grupos?

La fórmula más directa de la desviación típica es $\sigma = \sqrt{\frac{\Sigma(X - M)^2}{N}}$ [5]

Hay otras fórmulas, que se derivan de [5], para calcular la desviación típica sin necesidad de calcular las puntuaciones diferenciales (X-M). Lo que sucede es que estas fórmulas que simplifican operaciones son menos útiles, ya que disponemos de calculadoras con programación estadística.

Una de estas fórmulas para calcular la desviación típica a partir de las puntuaciones directas y de la media, es la que nos interesa para poder demostrar la fórmula que nos permite combinar desviaciones típicas de varios grupos; es la fórmula [6] que podemos encontrar en muchos textos.

A partir de esta fórmula [6] llegamos con facilidad a la fórmula [3] para combinar desviaciones típicas o varianzas:

$\sigma = \sqrt{\frac{\Sigma X^2}{N} - M^2}$ [6] la varianza será $\sigma^2 = \frac{\Sigma X^2}{N} - M^2$ y $\Sigma X^2 = N(\sigma^2 + M^2)$ [7]

En [7] podemos ver ya el parecido con la fórmula [3]. Si de dos muestras conocemos los valores de N, M y σ, podemos utilizar la fórmula [6] para calcular la desviación típica de las dos muestras combinadas en una única muestra.

En esta fórmula [6] tenemos ahora que $N = N_1 + N_2$; la media será la media total de las dos (o más) muestras y ΣX^2 será la expresión [7] calculada en las dos (o más) muestras y sumadas. Es decir, substituyendo [7] en [6] llegamos a la fórmula [3].

Podemos preguntarnos ahora de dónde sale la fórmula [6], que es la que hemos utilizado para demostrar la fórmula [3] que nos permite calcular la desviación típica de dos (o más) muestras combinadas en una sola.

---

[4] $\sum M^2 + \sum\sigma^2$ se calcula con mucha facilidad con una calculadora con programación estadística: basta introducir todas las medias y todas las desviaciones y obtener el resultado en la función Σx^2

En la fórmula de la desviación típica, la habitual, tenemos en el numerador un binomio elevado al cuadrado [$\Sigma(X\text{-}M)^2$]. No hay más que aplicar el procedimiento usual: $(a - b)^2 = a^2 - 2ab + b^2$, y así llegamos a [6]:

Utilizando la varianza para simplificar, tenemos que:

$$\sigma^2 = \frac{\Sigma(X - M)^2}{N} = \frac{\Sigma(X^2 - 2MX + M^2)}{N} = \frac{\Sigma X^2}{N} - 2M\frac{\Sigma X}{N} + \frac{\Sigma M^2}{N} \qquad [8]$$

Como en [8] tenemos que $\frac{\Sigma X}{N} = M$ y $\frac{\Sigma M^2}{N} = \frac{NM^2}{N} = M^2$

tendremos que $\sigma^2 = \frac{\Sigma X^2}{N} - 2M^2 + M^2$ $\quad \sigma^2 = \frac{\Sigma X^2}{N} - M^2$ y así llegamos a [6].

ANEXO IV: Contrastes posteriores: la distribución q del *rango estudentizado*

En los contrastes posteriores de *Tukey* utilizamos la distribución de **q** o *rango estudentizado* (a veces se utiliza el símbolo *qk*; k suele simbolizar el número de medias o de muestras). Las tablas de q también se utilizan en otros contrastes posteriores (como en los de Duncan y Newman-Keuls)[5].

Es fácil entender la fórmula y distribución de q porque se trata de algo análogo a la fórmula y distribución de t o z cuando contrastamos dos medias.

Comparemos en primer lugar las fórmulas de t (para dos muestras de idéntico número de sujetos = n) y de q para varias muestras también de idéntico tamaño (n):

$$t = \frac{M_1 - M_2}{\sqrt{\dfrac{\sigma_1^2 + \sigma_2^2}{n-1}}} \qquad\qquad q = \frac{M_i - M_k}{\sqrt{\dfrac{CM_{dentro}}{n}}}$$

Los *numeradores* son idénticos; se trata de la diferencia entre dos medias.

Los *denominadores* no son idénticos, pero casi. El que en la t de Student tengamos n-1 en el *denominador del denominador* quiere decir que las dos varianzas se han calculado dividiendo por n y no por n-1; si las hubiéramos calculado dividiendo por n-1 tendríamos n en el denominador; ahí no hay diferencia.

Los *cuadrados medios dentro de los grupos* que vemos en la fórmula de q no son otra cosa que la combinación de las varianzas de todos los grupos, mientras que en la t de Student sólo tenemos dos varianzas porque solamente se trata de dos grupos.

Por lo tanto el estadístico q es *análogo* a la t de Student o a una puntuación típica z: es en este caso un valor asociado a la probabilidad de que se dé una diferencia entre dos medias en el caso de que ambas procedan de la misma población. De la misma manera que tenemos una *distribución* de t, tenemos una distribución de q. Tanto t como q expresan una diferencia entre dos medias en *desviaciones típicas* o (con más propiedad *errores típicos,* pues es eso lo que tenemos en los denominadores). En ambos casos esta reducción de una diferencia a una *puntuación típica* (eso viene a ser lo que realmente hacemos) nos permite conocer las probabilidades de que esa diferencia sea casual o, lo que es lo mismo, debida a fluctuaciones normales.

En la distribución de q se tiene en cuenta el que en la medida en que aumenta el número de medias (no se trata solamente de dos medias), aumenta también la probabilidad de error al rechazar la Hipótesis Nula de *no diferencia.* La diferencia entre las fórmulas de t y q está en que en el denominador de t tenemos el *error típico de la diferencia entre dos medias* y en el caso de q lo que tenemos en el denominador es el *error típico del conjunto de medias* que tenemos en nuestro planteamiento; se basa en la variabilidad o diferencias normales y esperadas en un conjunto de medias.

El término *rango estudentizado* que suele aplicarse a q puede resultar poco claro. *Rango* viene a ser aquí un término análogo de *diferencia* (entre dos medias) y *estudentizado* (por analogía con la *t de Student,* quizás más claro sería decir *tipificado)* quiere decir que dividimos la diferencia entre dos medias por una desviación típica (o *error típico* que es como denominamos a las desviaciones típicas de las distribuciones muestrales). El término *rango* tiene su sentido porque expresa *orden:* ordenamos las medias de mayor a menor y lo que

[5] La distribución de q se la debemos a William Sealey Gossett (Kirk, 1995), lo mismo que la de la *t de Student.*

verificamos es la probabilidad de que la media más alta y la más baja difieran significativamente (procedan de la misma población). Si las medias extremas proceden de la misma población, también consideramos que las medias intermedias no difieren significativamente entre sí. Si las medias mayor y menor difieren significativamente, podemos seguir comparando entre sí el resto de las medias.

Es útil caer en la cuenta de esta relación entre t y q para captar que lo que hacemos es semejante a lo que hacemos cuando comparamos dos medias mediante la t de Student. Si observamos las tablas de q y nos fijamos en los valores correspondientes a *dos medias* nada más (en cuyo caso podríamos utilizar la t de Student) y para muestras grandes (los grados de libertad máximos que vienen en las tablas) tenemos que:

niveles de confianza	valor de q para k =2 y gl = ∞	valor de t = $\frac{q}{\sqrt{2}}$	
α = .05	2.77	$\frac{2.77}{\sqrt{2}} = 1.96$	Estos son los valores de t que encontramos en las tablas de la distribución normal para muestras grandes.
α = .01	3.64	$\frac{3.64}{\sqrt{2}} = 2.57$	

ANEXO V: Contrastes posteriores en diseños factoriales cuando la interacción es significativa

Tratando del análisis de varianza con dos criterios de clasificación (diseños factoriales) vimos cómo hacer los contrastes posteriores; se trata de una aplicación del contraste de Tukey. Cuando comparamos medias del mismo factor (medias de *fila*s entre sí o de *columnas* entre sí), el valor de k para consultar las tablas de q es el número de medias (o niveles) que hay en el factor.

Cuando la *interacción* es significativa nos puede interesar contrastar las medias de los distintos niveles de un mismo factor en un nivel del otro. En la tabla 1 si la interacción es significativa nos puede interesar comparar, por ejemplo las medias de A_1B_1 y A_3B_1 (dos medias de A *en uno* de los niveles de B).

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

Tabla 1

1. *Valores de k para consultar las tablas*

En estos casos los valores de k para consultar las tablas no son el número de medias, sino que dependen del número de comparaciones de interés, que a su vez dependen del número de niveles que hay en los factores[6] (tabla 2).

tamaño de la tabla	Valor de k' para consultar las tablas de q
2x2	3
2x3	5
2x4	6
2x5	8
3x3	7
3x4	8
3x5	10
4x4	10
4x5	12
5x5	15

Tabla 2

Los valores de k que tenemos que consultar son los puestos en la tabla 2 que hemos preparado según la justificación puesta más adelante (y que es la misma tabla puesta al tratar de los contrastes posteriores en los diseños factoriales). En esta tabla 2 tenemos el valor de k

[6] Por esta razón en la tabla 1 utilizamos el símbolo k' en vez de k. Aunque como indica Toothaker (1993), en vez de tener en cuenta el número de comparaciones *posibles,* podemos contabilizar solamente (para buscar k en las tablas de q) el número de comparaciones que nos *interesan* o que tienen sentido en nuestro planteamiento. Como criterio *conservador* podemos tener en cuenta *todas* las comparaciones *posibles*.

en las tablas de Tukey según el tamaño de nuestra tabla nxn, desde tablas 2x2 hasta tablas 5x5.

2. *Justificación de los nuevos valores de k*

Lo que pretendemos ahora es justificar esta tabla y explicar cómo está hecha, de manera que podamos ampliarla apara planteamientos con más 5 niveles en uno o en los dos factores.

Para explicar de dónde viene la tabla 2, seguimos dos pasos metodológicos:[7]

1º Calculamos el número de comparaciones posibles de interés (valor de C, fórmula [2], puesta más adelante)

2º A partir de este número podemos calcular los valores de k (análogo al número de medias en los diseños unifactoriales) para consultar las tablas de q con *grados de libertad* = N - ab (número total de sujetos menos número de subgrupos).

1º Calculamos el número de *comparaciones posibles* entre medias del mismo nivel y factor

Una cosa es *todas las comparaciones posibles,* y otra *todas las comparaciones posibles de interés.* En general el número total de comparaciones posibles tomando las medias de dos en dos ya sabemos que es:

$$\textit{número de comparaciones posibles} = \frac{k(k-1)}{2} \quad [1]$$ donde k es el número de grupos.

Aquí no nos interesan *todas* las comparaciones posibles sino solamente todas las posibles *entre los niveles de cada factor*. Antes de presentar la fórmula correspondiente [2] vamos a ver, paso a paso, cuántas comparaciones podemos hacer en el planteamiento inicial, donde tenemos dos factores, A dividido en tres niveles, y B dividido en dos niveles (tabla 2 x 3). De esta manera entendemos mejor la fórmula.

Comparaciones en los tres niveles del factor A

Número de comparaciones en A_1 (dos subgrupos) = $\frac{2(2-1)}{2} = 1$

En cada *columna* cabe una comparación (sólo hay dos medias) y como tenemos tres columnas (los tres niveles de A) el número de comparaciones posibles en el factor A es 1 x 3 = 3

Comparaciones en los dos niveles del factor B

Número de comparaciones en B_1 (tres subgrupos) = $\frac{3(3-1)}{2} = 3$

En cada fila caben 3 comparaciones, y como tenemos dos filas (B_1 y B_2), el número de comparaciones posibles en el factor B son 3 x 2 = 6

El número total de comparaciones posibles es igual a 3 + 6 = 9 (siempre dentro de la misma fila o columna).

Por lo tanto la fórmula general para calcular el número de comparaciones posibles *entre medias pertenecientes al mismo nivel de un mismo factor* cuando tenemos dos factores es:

---

[7] El procedimiento es de Cicchetti (1972) aunque la tabla III (reproducida parcialmente) está tomada de Linton, Gallo y Logan (1975, que a su vez la adaptan de Cicchetti). La justificación del procedimiento puede verse también en Cicchetti (1972) y en Toothaker (1991, 1993).

$$C = \left[\frac{b(b-1)}{2} \text{ x a}\right] + \left[\frac{a(a-1)}{2} \text{ x b}\right] \qquad [2]$$

C = comparaciones posibles entre medias del mismo nivel
a = número de niveles en el factor A
b = número de niveles en el factor B

Es sencillo confeccionar una tabla con el número de comparaciones posibles cuando tenemos dos factores divididos en varios niveles.

En la tabla 3 figuran las comparaciones posibles (contando solamente las que se pueden hacer dentro de cada nivel de cada factor) cuando tenemos divididos los factores entre 2 y 5 niveles:

Tablas de dos factores	Número de contrastes posibles en el mismo nivel (C)
2 x 2	4
2 x 3	9
2 x 4	16
2 x 5	25
3 x 3	18
3 x 4	30
3 x 5	45
4 x 4	48
4 x 5	70
5 x 5	100

Tabla 3

El número de contrastes posibles es un paso previo para poder consultar las tablas de q. Para consultar estas tablas necesitamos dos valores, el de k (*número de medias* o su equivalente como en este caso) y los *grados de libertad.*

2º *Valor de k*

Sabemos ya que el número de comparaciones posibles (C), tomándolos de dos en dos, de k elementos es igual a:

$$C = \frac{k(k-1)}{2}$$

En nuestro caso partimos ya del valor de C, calculado previamente (tabla 2). En nuestro ejemplo, una tabla de 2 x 3, C = 9, por lo que 9 = [k (k -1)/2]. Lo que tenemos que hacer es despejar el valor de k:

Si $9 = \frac{k(k-1)}{2}$ tenemos que $k^2 - k - (9 \text{ x } 2) = 0.$

Se trata de una ecuación de segundo grado, cuya expresión general es:

$ax^2 + bx + c = 0$ y la doble solución para x es $x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$

En nuestro caso particular tenemos que siempre a = +1, b = -1 y c = - (2 x C) (multiplicamos por dos el número de comparaciones posibles ya calculado y con signo *menos),* por lo que nuestra solución para x (tendremos siempre *dos* soluciones o valores de x que cumplen la ecuación) es:

$$x = \frac{1 \pm \sqrt{1 + (4x2C)}}{2} \qquad [3]$$

En nuestro ejemplo C = 9, por lo que las dos soluciones de x son 4.77 y -3.77; redondeando estos números y en valores absolutos tenemos 5 y 4.

Con estos valores, 5 y 4, aplicamos la fórmula de C, y nos quedaremos, para utilizarlo como k para consultar las tablas, con el valor de x que nos dé un resultado más próximo al valor de C ya calculado (tabla 3):

$$C = \frac{5(5-1)}{2} = 10 \qquad\qquad C = \frac{4(4-1)}{2} = 6$$

El valor de x (5 ó 4) que al aplicar la fórmula de C nos dé un valor más próximo a nuestro C (= 9) es 5 (10 se aproxima a 9 más que 6), luego escogemos 5 como valor de k para consultar las tablas.

Con este procedimiento podemos ampliar las tablas si en algún factor tenemos más de seis niveles.

Supongamos que tenemos un cuadro de 6 x 6, dos factores cada uno dividido en 6 niveles. El número posible de comparaciones (siempre dentro de cada factor) sería igual a C = 180 (según vimos antes en [2]), y obtendríamos estos valores de x: 19.47 y 18.47 (en números absolutos) y redondeando 19 y 18, con los que tendríamos estos valores de C:

$$C = \frac{19(19-1)}{2} = 171 \qquad\qquad C = \frac{18(18-1)}{2} = 153$$

El valor más próximo a nuestro C (= 180) nos lo da x = 19, luego tenemos que k = 19 al consultar las tablas de q.

El proceso puede parecer complicado, pero no lo es si nos fijamos en los tres sencillos pasos de este *resumen:*

1º Calculamos el número de comparaciones posibles entre medias pertenecientes al mismo factor; es el valor de C de la fórmula [2]

2º Calculamos los dos valores de x de la ecuación anterior; fórmula [3]

3º Utilizamos los dos valores de x en la fórmula general [1] y nos quedamos con el valor de x que nos dé un número más próximo a C: éste va a ser el valor de k que utilizaremos al consultar las tablas.

La tabla 4 está confeccionada siguiendo los pasos indicados.

Número de comparaciones posibles *no confundidas* (C)	Valor de k' para consultar las tablas de q
3 - 4	3
5 - 8	4
9 - 12	5
13 - 17	6
18 - 24	7
25 - 32	8
33 - 40	9
41 - 50	10
51 - 60	11
61 - 72	12
73 - 84	13
85 - 98	14
99 - 112	15

Tabla 4

En la práctica lo más cómodo es disponer de una tabla que combine las tablas 3 y 4, para poder consultar el valor de k en las tablas de q a partir del número de niveles que tenemos en nuestro planteamiento; esto es lo que hemos hecho en la tabla 2 puesta al comienzo de este anexo, y que también figura en el lugar correspondiente al tratar de los diseños factoriales.

ANEXO VI: El número de sujetos en el análisis de varianza

En los estudios de carácter empírico o experimental, en sentido amplio, las normas que suelen darse para determinar el número apropiado de sujetos depende de la finalidad del estudio.

a) Posiblemente lo más frecuente es encontrar normas y fórmulas para determinar el tamaño de la muestra cuando se quieren extrapolar los resultados a la población, como es normal en los estudios de carácter más sociológico (encuestas, sondeos pre-electorales, etc.).

b) Normas o recomendaciones distintas suelen darse con otras finalidades específicas, como el número de sujetos necesario cuando se trata de construir un test o escala, o el número de sujetos necesario o recomendable cuando se va a hacer un análisis correlacional, o más concretamente, un análisis factorial.

c) Aquí nos interesa cómo determinar el número de sujetos en los estudios experimentales, los más relacionados con el análisis de varianza. Incluiremos también lo relativo al tamaño de las muestras cuando sólo se requiere la *t de Student* para comparar dos medias, por cercanía con el análisis de varianza que también se podría utilizar en este caso, y completamos así las orientaciones sobre el número de sujetos cuando se utiliza un diseño experimental o cuasi-experimental en un sentido más restringido.

1. Variables que intervienen en la determinación del tamaño de la muestra

Aunque en la práctica podemos limitarnos a consultar unas tablas, es muy conveniente conocer con qué criterios están hechas estas tablas. Se trata de las variables de las que depende el tamaño de la muestra.

1. *El nivel de confianza* (que solemos expresar así: $\alpha = .05$, $\alpha = .01$). Si escogemos un nivel de confianza de .05 (como es práctica común) queremos decir que aceptamos un 5% de probabilidades de error al rechazar la Hipótesis Nula (de no diferencia). Se trata de minimizar el denominado error Tipo I (aceptamos pocas probabilidades de equivocarnos cuando afirmamos una diferencia).

2. La *potencia de la prueba*. Por potencia entendemos la probabilidad de no cometer el error denominado Tipo II: no rechazar la Hipótesis Nula cuando podríamos haberla rechazado. La probabilidad de cometer este tipo de error se simboliza como β, y la potencia es por lo tanto $1-\beta$. Podemos definir la *potencia* como *la probabilidad de rechazar una Hipótesis Nula que es falsa.*

De la misma manera que un *nivel de confianza* de $\alpha = .05$ es habitualmente aceptado como razonable, por lo que respecta a la potencia ($1-\beta$) se estima que es razonable establecer una potencia de .80, es decir tener un 80% de probabilidades de detectar una diferencia (o relación) de una determinada magnitud[8]. Si deseamos una potencia mayor (.90 o incluso 1) el tamaño requerido de la muestra puede ser ya excesivamente grande.

El error Tipo I (decir *sí* cuando habría que decir que *no* hay diferencia, relación, etc.) es más serio que el error Tipo II (decir *no* cuando podríamos haber dicho que *sí),* de ahí la práctica generalizada de utilizar unos niveles de confianza muy estrictos, como son .05 ó .01:

[8] La recomendación de una potencia de .80 la propone y justifica Cohen (1988:56; Jacob Cohen es la fuente principal que suele seguirse en este tema). El peligro de *cometer* el error Tipo II queda reducido a .20 (20% de probabilidades) y está en equilibrio con $\alpha = .05$: suponemos que el error Tipo I es cuatro veces más serio que el error Tipo II (.20 es cuatro veces .05). Esta recomendación no es tan seguida como la de establecer un nivel de confianza de .05, porque con frecuencia no se tiene en cuenta el error Tipo II, que por otra parte ya se controla bien cuando el número de sujetos es grande.

aceptamos muy pocas probabilidades de equivocarnos cuando afirmamos una diferencia. Si establecemos un nivel de significación muy estricto (un valor de α muy bajo) es muy improbable que cometamos el error Tipo I: si rechazamos el *azar* (o la variabilidad normal debida al *error muestral)* como explicación de una diferencia es muy poco probable que nos equivoquemos.

Lo que sucede es que con un valor muy bajo de α podemos caer en el error Tipo II: puede ser que la Hipótesis Nula sea falsa, pero como somos muy estrictos no llegamos a rechazarla. Con un nivel de significación de α = .001 las probabilidades de *no* rechazar una Hipótesis Nula que sea realmente falsa pueden ser muy pocas. En la práctica hay que *sopesar* ambos tipos de error. El minimizar el error Tipo I no significa que no tengamos que prestar atención al error Tipo II. Aunque las decisiones sobre el tamaño de la muestra se toman frecuentemente en función de los datos disponibles, o imitando lo que han hecho otros, no es racional, (como señala Cohen, 1988:55), el determinar el tamaño de la muestra sin tener en cuenta el error Tipo II.

3. La *magnitud de la diferencia* (o de la relación, etc.) que deseamos detectar y que solemos denominar *tamaño del efecto*. El término *efecto* no implica *causalidad,* sino simplemente el *grado* en que un fenómeno (diferencia, relación, etc.) está presente.

La implicación de la magnitud en el tamaño de la muestra es obvia: cuando las diferencias son grandes, nos bastan pocos sujetos para detectarlas, pero cuando son muy pequeñas necesitamos muchos sujetos; si solamente nos interesan diferencias grandes, necesitaremos muchos menos sujetos. Podemos intuirlo con un ejemplo muy claro. Si estamos interesados en comprobar si difieren en *altura* los *escandinavos* y los *twa* (pigmeos de Ruanda y Burundi) no necesitaremos muchos sujetos en las muestras; nos bastarán muy pocos sujetos de cada grupo para caer en la cuenta de que se trata de poblaciones muy distintas en altura. En cambio si se trata de encontrar diferencias pequeñas entre las medias de poblaciones que no difieren mucho entre sí, no nos bastará con comparar muestras de tamaño pequeño. Es claro por otra parte que con muestras grandes es fácil encontrar diferencias *estadísticamente significativas* pero pequeñas y con frecuencia irrelevantes.

Al planificar cualquier tipo de experimento o análisis debemos tener en cuenta también en qué tipo de *magnitud* estamos interesados, porque si solamente son de interés magnitudes más bien grandes podemos ahorrar costes y trabajo utilizando muestras relativamente pequeñas. Si queremos garantizar a *toda costa* que nos salgan unas diferencias (o relaciones) *estadísticamente significativas,* aunque sean muy pequeñas (y con frecuencia de interés muy dudoso o cuestionable), necesitaremos muestras muy grandes.

4. La *varianza de la población:* ya sabemos que si los sujetos son muy iguales dentro de cada grupo, necesitaremos muestras menores para detectar diferencias (si todos son de idéntica altura, o todos piensan lo mismo, etc., nos bastaría un solo sujeto de cada grupo para ver si hay alguna diferencia entre los grupos).

Estas cuatro variables se combinan en las fórmulas apropiadas para determinar el tamaño óptimo de las muestras. Aunque en principio son preferibles las muestras grandes, por razones de economía (costos, trabajo) podemos *calibrar* el tamaño de la muestra de acuerdo con nuestras especificaciones en estas cuatro variables.

No necesitamos aplicar las fórmulas para conocer el tamaño de la muestra porque ya disponemos de tablas para las situaciones más frecuentes; las tablas que ponemos aquí están muy reducidas pero pueden ser suficientes como orientación sobre el tamaño de la muestra que debemos buscar (tablas más completas pueden encontrase en los autores que citamos y en otros). Sí es de interés conocer qué variables inciden en el número de sujetos que necesitamos.

No sobra recordar aquí que el *tamaño de la muestra* es importante, pero no es la única característica de la muestra que nos interesa. En diseños experimentales en sentido propio necesitaremos *muestras aleatorias,* y en cualquier caso siempre debemos preguntarnos a qué población pueden estar representando las muestras que utilizamos.

2. Tamaño de cada muestra cuando comparamos dos grupos *(t de Student)*

En la tabla 4 tenemos el tamaño de *cada muestra* necesario para comparar *dos* muestras[9]. Tablas semejantes, más o menos extensas o adaptadas, pueden encontrarse en diversos autores; no siempre coinciden exactamente las cifras del número de sujetos debido al distinto redondeo de decimales al aplicar las fórmulas.

Suponemos: varianzas iguales,
muestras de idéntico tamaño,
hipótesis bilaterales
potencia (1-β) de .80

nivel de confianza	d =.20	d = .30	d =.50	d = .70	d =.80	d =.1.0	d = 1.20
.05	392	174	63	32	25	16	12
.01	586	260	93	48	36	23	18

Tabla 5

Estamos suponiendo muestras de *idéntico tamaño,* pero si tenemos ya una muestra con un determinado número de sujetos, podemos calcular el tamaño necesario en la otra muestra.

La fórmula [Cohen, 1988:59] es ésta: $n_{nuevo} = \dfrac{(n_{disponible})(n_{tablas})}{2n_{disponible} - n_{tablas}}$

Vamos a suponer, por ejemplo, que tenemos ya un grupo experimental de 40 sujetos que ha tenido una determinada experiencia y deseamos compararlo con otro (grupo de control, o al menos como término de comparación); estamos interesados en detectar al menos una diferencia moderada (d = .50) a un nivel de confianza de α = .05 ¿Cuántos sujetos deberemos incluir en el nuevo grupo de control? En las tablas vemos que necesitaríamos 63 sujetos en *cada* grupo; el tamaño del nuevo grupo deberá ser:

$$n_{nuevo} = \frac{(40)(63)}{(2x40) - 63} = 148$$

3. Tamaño de la muestra cuando tenemos más de dos muestras *(análisis de varianza unifactorial)*

En la tabla 6 tenemos el número de sujetos necesario en *cada muestra* cuando tenemos más de dos muestras (entre tres y seis muestras).

En esta tabla hemos puesto como orientación los valores correspondientes a α = .05 y 1-β *(potencia)* de .70 y .80; suponemos también un número idéntico de sujetos en cada muestra.

Podemos tomar como referencia de magnitud o el valor de ω^2 o el valor de f, el tamaño del efecto propuesto por Cohen (1988) cuando tenemos más de dos grupos[10].

[9] Valores seleccionados de la tabla 2.4.1 de Cohen (1988).

[10] Ponemos los dos valores porque podemos encontrar los dos como referencia en otras tablas.

a) El coeficiente ω^2 nos cuantifica el *grado de asociación* entre la variable independiente (el pertenecer a uno u otro grupo) y la variable dependiente.

b) El *tamaño del efecto f* propuesto por Cohen (1988) cuando tenemos más de dos grupos[11].

Cuando tenemos solamente dos grupos, ya sabemos que el tamaño del efecto es igual a la diferencia entre las dos medias dividida por la desviación típica combinada. Cuando hay más dos grupos el denominador es el mismo, pero lo que tenemos en el numerador es la dispersión o desviaciones de todas las medias con respecto a la media común (un valor análogo a la desviación típica de las medias). En la práctica el cálculo más sencillo de *f* es a partir de ω^2 (que es habitual calcular como complemento al análisis de varianza); ambos valores están relacionados de esta manera (fórmula 2.27):

$$f = \sqrt{\frac{\omega^2}{1-\omega^2}}$$

Realmente si hemos calculado ω^2 ya no necesitamos conocer el valor de f, pues no va a aportar una información que nos lleve a una interpretación o a una valoración distinta. Por lo que respecta a tener una orientación sobre el tamaño de la muestra, nos basta consultar las tablas teniendo en cuenta, al menos de manera aproximada, el tipo de magnitud en la que estamos interesados. Las valoraciones (magnitud *pequeña, moderada y grande)* son de Cohen y constituyen una referencia comúnmente aceptada como guía orientadora[12].

Suponemos que las k muestras son de idéntico tamaño; si son de tamaño desigual podemos utilizar el *tamaño medio* de las muestras (N/k)[13].

número de grupos	*potencia*	*pequeña* $\omega^2 = .01$ f = .10	*magnitud* *moderada* $\omega^2 = .06$ f = .25	*grande* $\omega^2 = .14$ f = .40
3	.70	258	42	17
	.80	322	52	21
4	.70	221	36	15
	.80	274	45	18
5	.70	195	32	13
	.80	240	39	16
6	.70	175	29	12
	.80	215	35	14

Tabla 6

[11] Explicado por Cohen (1988:274ss, 284). Se trata de un tamaño del efecto *global,* teniendo en cuenta todas las diferencias de las medias con respecto a la media total (no se trata de la diferencia entre *dos* medias, como sucede en el tamaño del efecto convencional).

[12] Los valores de referencia seleccionados están tomados de Cohen (1988, tablas 8.4.4 y 8.4.5); también pueden verse en Kirk (1995:186 y tabla E.13). Las tablas de Cohen (válidas hasta 25 muestras) son más fáciles de consultar, y utiliza f como criterio de magnitud; otros autores como Kirk (1995) utilizan ambos valores f y ω^2. Cohen utiliza el símbolo η^2 en vez de ω^2 (y comenta la falta de unanimidad en los símbolos en p. 282).

[13] Las implicaciones del tamaño desigual pueden vese comentadas en Cohen (1988:360ss). Si las muestras mayores tienen también las mayores medias, el tamaño del efecto será mayor que si las muestras fueran de idéntico tamaño y también será mayor la potencia (y a la inversa también es verdad).

Los valores tabulados son el número de sujetos en *cada muestra.* En el caso de tres muestras, si estamos interesados en verificar solamente si hay diferencias valoradas como *grandes* y con una probabilidad de encontrarlas (si las hay) del 80%, necesitaremos una *muestra total* de 21x3 = 63 sujetos; si en cambio consideramos un buen resultado el encontrar diferencias *pequeñas* (pero significativas), necesitaremos una *muestra total* de 322x3 = 966 sujetos.

Comparando los valores correspondientes a una potencia de .70 y 80 podemos apreciar cómo al disminuir el número de sujetos disminuyen las probabilidades de rechazar la Hipótesis Nula. En tablas más extensas podríamos ver que a) si tenemos tres muestras, b) estamos interesados en descubrir pequeñas diferencias (pero significativas) porque en nuestro caso las consideramos relevantes y c) sólo tenemos en torno a 20 sujetos en cada muestra, las probabilidades de tener que aceptar la Hipótesis Nula de no diferencia siendo falsa son del 90%.

4. Tamaño de la muestra en los diseños factoriales

En la tabla 7 tenemos el número necesario de sujetos en *cada celda* cuando tenemos *dos criterios de clasificación* (o *factores)* divididos en entre dos y cuatro niveles. Suponemos que en cada clasificación hay un idéntico número de sujetos, como es usual en estos planteamientos.

Suponemos también un *nivel de confianza* de α = .05 y una *potencia* (1-β) de .70 o de .80 En estas tablas los niveles (o subclasificaciones) de cada factor pueden ser 2, 3 ó 4.

Para valorar la *magnitud* utilizamos los mismos criterios de la tabla 5[14]

El número *total* de sujetos será igual al número de sujetos que aparece en la tabla multiplicado por el número de subclasificaciones o celdas. En una tabla 2x3 tenemos 6 celdas; si estamos interesados en detectar diferencias moderadas el número total de sujetos será 6x18 = 108.

El número de sujetos especificado en la tabla 7 es suficiente para detectar si uno de los dos factores (o los dos) es estadísticamente significativo (si hay diferencias entre los niveles de cada factor, o, lo que es lo mismo, entre las medias de cada columna o de cada fila), pero en el caso de la interacción con estos mismos números la potencia es menor porque intervienen menos sujetos (no los totales de cada fila o columna sino los que hay en cada clasificación).

[14] Los valores del tamaño de la muestra (en *cada* clasificación), puestos como referencia orientadora, están seleccionados de las extensas tablas de Kirk (1995:401 y tabla E.13). La disposición de la tabla es una adaptación muy simplificada.

Tamaño de la tabla	*potencia*	*pequeña* $\omega^2 = .01$ f = .10	*magnitud moderada* $\omega^2 = .06$ f = .25	*grande* $\omega^2 = .14$ f = .40
2x2	.70	152	25	11
	.80	193	32	13
2x3	.70	127	21	9
	.80	158	26	11
2x4	.70	109	18	8
	.80	134	22	9
3x3	.70	85	14	6
	.80	106	18	7
3x4	.70	73	12	5
	.80	90	15	6
4x4	.70	55	9	4
	.80	67	12	5

Tabla 7

Anexo VII: Métodos *no paramétricos* de análisis de varianza

Disponemos de una serie de alternativas *no paramétricas* al análisis de varianza. No las exponemos aquí porque se trata de procedimientos sencillos que pueden encontrarse en muchos textos[15]; en cambio sí interesa tener una visión de conjunto de estas otras posibilidades metodológicas, al menos de las más importantes.

1. *Observaciones sobre los métodos no paramétricos*

a) En realidad en estas alternativas no se trata de *análisis de varianza* propiamente dicho, porque no descomponemos ninguna varianza total en varianzas parciales; con estos análisis sin embargo llegamos a conclusiones del mismo orden, sobre si entre varias muestras hay *diferencias estadísticamente significativas* o no las hay y pueden considerarse extraídas de la misma población.

b) El término *no paramétrico* viene del hecho de que en estos métodos no hacemos ninguna suposición sobre las condiciones que deben darse en las poblaciones de donde proceden las muestras (un *parámetro* es una medida o característica de una *población)*. Estas condiciones o supuestos ya los conocemos:

1º Unidades de intervalo en las medidas,
2º Distribución normal en las poblaciones,
3º Homogeneidad de varianzas.

c) Aunque los métodos paramétricos, y el análisis de varianza en particular, son *en términos generales* métodos seguros aun cuando no se cumplan los supuestos previos, las alternativas *no paramétricas* pueden ser preferibles cuando estos supuestos se violan de manera clara, sobre todo con muestras pequeñas (y con mayor razón si son de tamaño desigual), con distribuciones claramente no normales y sobre todo con varianzas muy desiguales.

d) No hay que confundir métodos *no paramétricos* con *métodos ordinales,* aunque muchos de los métodos no paramétricos son ordinales; es decir el dato que se utiliza de cada sujeto es el *rango* o *número de orden* (en otros se utilizan frecuencias acumuladas; el *ji cuadrado,* y otros métodos semejantes para tratar variables categóricas, son también no paramétricos).

e) Los métodos *no paramétricos* no son una mera alternativa a los métodos paramétricos; son buenos procedimientos de análisis en sí mismos, con la ventaja añadida de su *simplicidad*.

2. *Métodos no paramétricos más importantes análogos al análisis de varianza*

Los más conocidos y los que con más frecuencia se encuentran en los textos son las alternativas a la *t de Student,* para dos muestras independientes (el más conocido es la U de Mann-Whitney aunque hay otros métodos, como la prueba de Kolmogorov-Smirnov que utiliza frecuencias acumuladas y la prueba de las *rachas* de Wald-Wolfowitz) o relacionadas (como la T de Wilcoxon y la *prueba de los signos* que es una aplicación directa de la distribución binomial*).* Por lo que respecta a lo métodos no paramétricos que pueden considerarse alternativas al análisis de varianza (para analizar más de dos muestras independientes o relacionadas) los más conocidos son los que figuran en la tabla 8.

---

[15] En muchos textos de estadística es normal encontrar algún capítulo dedicado a estos métodos *no paramétricos* en los que se explican al menos los más importantes. De la abundante bibliografía dedicada a los métodos no paramétricos es útil mencionar a Siegel (1972; la primera edición en inglés es de 1956 y sigue siendo una obra de referencia importante), Siegel, y Castellan, (1988), Gibbons (1993) y Kanji (1993).

	Tipos de hipótesis	
	Hipótesis habitual: si las muestras *proceden de la misma población*	Hipótesis sobre si se da una *tendencia* a crecer o decrecer
Más de dos muestras *independientes.*	H de Kruskal-Wallis *(se utilizan rangos)*	Prueba de Jonckheere
Más de dos muestras *relacionadas*	Prueba de Friedman *(se utilizan rangos)* Q de Cochran *(con puntuaciones dicotómicas, 1 ó 0)*	Prueba L de Page *(extensión de la prueba de Friedman)* Prueba de Mann *(cuando n = 1)*

Tabla 8

Las pruebas no paramétricas para comprobar *tendencias* son menos populares; de hecho no se encuentran en la mayoría de los textos.[16]

La H de Kruskal-Wallis, la alternativa al análisis de varianza para varias muestras independientes más utilizada, es insensible a la *no normalidad* de las distribuciones, pero no es tan claro que lo sea también a la falta de homogeneidad de varianzas.[17]

Un inconveniente de las dos pruebas más populares como alternativa no paramétrica al análisis de varianza, la H de Kruskal-Wallis y el test de Friedman, es que la mayoría de los textos no suelen incluir contrastes posteriores o cálculos complementarios (coeficientes de asociación del tipo η^2) que ayuden a la interpretación de los resultados y por este motivo con frecuencia se estudian de manera incompleta[18].

En Internet se encuentra con facilidad cómo hacer los análisis no paramétricos más comunes (también el SPSS).

La H de Kruskal-Wallis y la prueba de Friedman (además de los equivalentes a la t de Student, U de Mann-Whiteny y T de Wilcoxon) se encuentran al menos en:

Lowry, Richard, VassarStats: Web Site for Statistical Computation,Vassar College, Poughkeepsie, NY, USA; http://faculty.vassar.edu/lowry/VassarStats.html (menú *ordinal data*)

Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong http://department.obg.cuhk.edu.hk/index.asp?scr=1024 (menú *frequency tables*)

[16] Las pruebas de Page y Jonckheere para comprobar tendencias no están descritas en el texto más conocido de Siegel y tampoco es fácil encontrarlas en textos convencionales de estadística. Las dos pueden verse al menos en Siegel y Castellan (1988) y en Hollander y Wolf (1973, un texto de nivel alto) y en Green y d'Oliveira (1984, un texto sencillo); la de Jonckheere está tratada con mayor extensión en Leach (1982); la de Mann (muestras relacionadas y n = 1) también se encuentra en Leach (1982).

[17] Lix, Keselman y Keselman, (1996).

[18] Sobre cómo llevar a cabo los contrastes posteriores en el análisis de varianza *no paramétrico* (algo que no es fácil encontrar en la mayoría de los textos) pueden consultarse Linton, Gallo y Logan (1975), Pett (1997), Black (1999).

ANEXO VIII: Análisis de Varianza en Internet

Son bastantes los programas que diversas instituciones ponen en Internet con acceso libre y que permiten llevar a cabo diversos tipos de análisis de varianza.

Muchos de estos programas permiten importar datos (de EXCEL, tabla de Word, etc.), otros resuelven el análisis de varianza para muestras independientes a partir de los datos descriptivos básicos *(from summary data:* media, desviación típica y número de sujetos). Algunos de estos programas están indicados en el último apartado de *análisis de varianza para muestras independientes.* En el mismo lugar se indica dónde se pueden consultar diversas tablas (F, Tukey, Dunnet) y las probabilidades exactas asociadas a cualquier valor de F.

Una dirección con numerosos enlaces de análisis estadísticos en las que podemos encontrar con programas para llevar a cabo distintos tipos de análisis de varianza pueden verse en **Web Pages that Perform Statistical Calculations!,** http://statpages.org/ Esta dirección se encuentra en *John C. Pezzullo's Home Page* (de Georgetown University, Washington, DC.) http://statpages.org/JCPhome.html (en Interactive Statistics Pages). Esta página presenta un índice bien estructurado con enlaces a los diversos modelos de análisis de varianza y otros análisis estadísticos. Muchos de estos enlaces remiten a estas direcciones de interés general:

College of Saint Benedict/Saint John's University [http://www.csbsju.edu/ **]**
http://www.physics.csbsju.edu/stats/

Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong
http://department.obg.cuhk.edu.hk/index.asp?scr=1024

GraphPad, San Diego, CA [http://graphpad.com/] Free Online Calculators for Scientists, http://graphpad.com/quickcalcs/index.cfm

Institute of Phonetic Sciences, Amsterdam (en Demos, tests, experiments →statistics)
http://www.fon.hum.uva.nl/Welcome.html

Lane, David M. *HyperStat Online Statistics Textbook*
http://davidmlane.com/hyperstat/index.html

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

SISA (*Simple Interactive Statistical Analysis,* de Consultancy for Research and Statistics, Hilversum, The Netherlands) http://www.quantitativeskills.com/sisa/

Referencias bibliográficas

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

BLAND, J. MARTIN and ALTMAN, DOUGLAS G. (1995). Multiple significance tests: the Bonferroni method. *British Medical Journal* 1995;310:170 (21 January). http://www.bmj.com/cgi/content/full/310/6973/170

CICCHETTI, DOMINIC V. (1972). Extensions of multiple-range tests to interaction tables in the analysis of variance: A rapid approximate solution. *Psychological Bulletin,* 77, 405-408.

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

DEPARTMENT OF OBSTRETRICS AND GYNAECOLOGY, THE CHINESE UNIVERSITY OF HONG KONG http://department.obg.cuhk.edu.hk/ResearchSupport/OWAV.asp

GIBBONS, JEAN DICKINSON, (1993). *Nonparametric Tests, an Introduction.* Quantitative Applications in the Social Sciences. Newbury Park: N.J.: Sage

GRAPHPAD. San Diego, CA [http://graphpad.com/] Free Online Calculators for Scientists http://graphpad.com/quickcalcs/posttest1.cfm

GREENE, JUDITH y D'OLIVEIRA, MANUELA, (1984). *Pruebas estadísticas para psicología y ciencias sociales: una guía para el estudiante*. Bogotá, Colombia: edit. Norma

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.

HOLLANDER, MYLES and WOLFE, DOUGLAS A., (1973). *Nonparametric Statistical Methods.* New York: Wiley and Sons.

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

KANJI, GOPAL, (1993). *100 Statistical Tests.* London: Sage.

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

LANE, DAVID M. *HyperStat Online Statistics Textbook* http://davidmlane.com/hyperstat/index.html

LEACH, CHRIS, (1982). *Fundamentos de estadística, enfoque no paramétrico para ciencias sociales.* México: Limusa.

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

LIX, LISA M., KESELMAN, JOANNE C. and KESELMAN, H.J., (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test. *Review of Educational Research*, 66 (4) 579-619.

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

MCGUIGAN, F. J., (1994) *Experimental Psychology, Methods of Research.* Sixth edition. Englewood Cliffs, N.J.: Hall.

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

PETT, MARJORIE A. (1997). *Nonparametric Statistics for Health Care Research. Statistics for Small Samples and Unusual Distributions.* Thousand Oaks & London: Sage.

PEZZULLO, JHON C. *Web Pages that Perform Statistical Calculations*. http://members.aol.com/johnp71/javastat.html; Analysis of Variance from Summary Data http://members.aol.com/johnp71/anova1sm.html

SIEGEL, SIDNEY and CASTELLAN, JR., N. JOHN (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill

SIEGEL, SIDNEY, (1972). *Estadística no paramétrica aplicada a las ciencias sociales.* México: Trillas.

TOOTHAKER, LARRY E., (1991). *Multiple Comparisons for Researchers*: Sage University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park: Sage.

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