Introducción al Análisis de Varianza

©Pedro Morales Vallejo
Universidad Pontificia Comillas
Facultad de Ciencias Humanas y Sociales
(Última revisión 19 de Agosto de 2012)

INDICE

1. Finalidad del análisis de varianza

El análisis de varianza lo vamos a utilizar *para verificar si hay diferencias estadísticamente significativas entre medias cuando tenemos más de dos muestras o grupos en el mismo planteamiento.* En estos casos no utilizamos la t de Student que solamente es un procedimiento válido cuando comparamos únicamente las medias de dos muestras. Como explicaremos más adelante, cuando tenemos más de dos muestras y comparamos las medias de dos en dos suben las probabilidades de error al rechazar la hipótesis de no diferencia porque queda suficientemente explicada por factores aleatorios (que también se denomina *error muestral*).

En primer lugar recordamos qué es la *varianza* y qué nos *cuantifica*. La fórmula de la *varianza* ya nos es conocida; es *la desviación típica elevada al cuadrado:*

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

Utilizamos el símbolo X para designar las puntuaciones individuales, y el símbolo M para designar la media aritmética de la muestra; σ va a ser el símbolo de la desviación típica de la muestra si no se indica expresamente que se trata del símbolo de la desviación típica de la población[1].

El *denominador* será N-1 si queremos obtener una estimación de la *varianza de la población.* Esto es lo que haremos habitualmente en el cálculo de las varianzas propias del *análisis de varianza*.

Una varianza grande indica que hay mucha variación entre los sujetos, que hay mayores diferencias individuales con respecto a la media; una varianza pequeña nos indica poca variabilidad entre los sujetos, diferencias menores entre los sujetos. La varianza *cuantifica* todo lo que hay de diferente entre los sujetos u observaciones.

Como iremos viendo *la varianza se puede descomponer* en varianzas parciales y a este descomponer la varianza le denominamos *análisis de varianza.* La varianza expresa variación, y si podemos descomponer la varianza, podemos aislar *fuentes de variación*. Cuando de los sujetos tenemos varios tipos de información, el análisis de varianza nos va a responder a esta pregunta *¿De dónde vienen las diferencias?*

---

[1] Utilizamos M como símbolo de la media aritmética (no $\overline{X}$) y σ (y no s) como símbolo de la *desviación típica de la muestra* (dividiendo por N, no por N -1); por razones de simplicidad y así lo vemos además en otros autores (como Guilford y Fruchter, 1978, que reconocen la falta de una práctica común en el uso de estos símbolos). Es por otra parte frecuente referirse a la desviación típica como *sigma,* el nombre del símbolo. En muchas calculadoras con programación estadística de uso frecuente se utilizan los símbolos σ_n y σ_{n-1} para referirse a la desviación típica de la muestra (dividiendo por N) y de la población (dividiendo por N - 1) respectivamente y son posiblemente los símbolos más claros. Otros autores (como Spatz, 1993) prefieren S (mayúscula) para designar la desviación típica de la muestra y s (minúscula) para la desviación típica de la población; otros en cambio (Rosenthal, 1987, 1991; Rosenthal y Rosnow, 1991) utilizan S para la población y σ para la muestra. Los símbolos σ para la desviación típica de la población y s para la desviación típica de la muestra (la práctica más común) son originarios de William S. Gossett (Pearson y Kendall, Eds., 1978) al que debemos también la distribución de la t de Student. Algún autor prescinde casi de todo tipo de símbolos (Guéguen, 1997). En nuestro caso el símbolo (σ) no se presta a confusión porque prácticamente siempre se trata de la desviación típica de la muestra a no ser que indiquemos expresamente que se trata de la desviación típica de la población; en este caso también utilizaremos ocasionalmente el símbolo σ_{n-1} para referirnos a la desviación típica de la población y σ_n para designar la desviación típica de la muestra.

El *análisis de varianza*[2] no constituye *un* método o procedimiento único; según los diseños y datos disponibles existen diversos modelos de análisis de varianza. En esta introducción nos referiremos al análisis de varianza para *varias muestras independientes,* y más concretamente al análisis de varianza para sólo dos muestras independientes (aunque en este caso solemos utilizar la t de Student) porque es de comprensión más sencilla. La misma explicación básica se puede extrapolar a otras situaciones (más de dos muestras independientes, más de dos muestras relacionadas, diseños factoriales, etc., que iremos viendo más adelante).

2. Por qué utilizamos el análisis de varianza en vez de la t de Student

Cuando tenemos dos muestras y queremos comprobar si difieren significativamente (si proceden de la misma población con una única media) utilizamos la t de Student. Cuando tenemos más de dos grupos utilizamos el análisis de varianza: ¿No podríamos comparar todos los grupos de dos en dos con la t de Student? A primera vista parecería lo más lógico, sin embargo no se hace así por una serie de razones que exponemos a continuación.

1º La razón *más importante* (y suficiente) para no utilizar la t de Student con más de dos grupos es que, al hacer muchas comparaciones de dos en dos, aumenta la probabilidad de que algunas diferencias resulten significativas por azar y entonces cabe la posibilidad de afirmar que hay una diferencia (de no aceptar la hipótesis nula) cuando realmente no la hay.

Si por ejemplo tenemos tres grupos podríamos hacer tres comparaciones: entre el 1º y el 2º, entre el 1º y el 3º y entre el 2º y el 3º. Operando con un nivel de confianza de α = .05, la probabilidad de encontrar al menos una diferencia significativa *por azar* es de hecho del 9.75% y no del 5% (no es importante el entender ahora el *por qué,* algo aclaramos en el anexo I).

2º Otra razón adicional es que una prueba estadística basada en *todos* los datos utilizados simultáneamente, es más estable que la prueba o análisis que parcializa los datos y no los examina todos juntos. El error típico (que expresa la variación en las medias que podemos encontrar en diversas muestras) es menor cuando el número de sujetos es mayor, como sucede cuando se analizan todos los datos de todos los grupos simultáneamente. En principio es preferible utilizar un método de *análisis global* que abarque todos los datos que se quieren examinar.

Aun así, si se tiene *como hipótesis previa a la recogida de datos* que dos de los grupos difieren estadísticamente, es legítimo utilizar en ese caso y para esos dos grupos la t de Student. Pero lo normal es que el análisis de varianza implique hipótesis relativas a todos los datos tomados simultáneamente, en un único planteamiento.

3º El ahorro de tiempo es otra razón que a veces se aduce, aunque en sí misma no es una razón válida[3]. El número de comparaciones de dos en dos de k elementos es igual a k(k-1)/2; con seis grupos habría que hacer 15 comparaciones y con 10 grupos subirían a 45. El análisis de varianza nos dice *de entrada* si hay o no hay diferencias significativas entre pares de medias, y si no las hay no necesitamos hacer más análisis. En cualquier

[2] También denominado ANOVA; del inglés *ANalysis Of VAriance,* y ANVA en español.

[3] Además la importancia que podría suponer el trabajo extra es nula utilizando programas de ordenador.

caso no sería legítimo comparar todas las medias de dos en dos, en un mismo planteamiento, por las razones dichas antes.

3. Qué comprobamos mediante el análisis de varianza: relación entre la diferencia entre varianzas y la diferencia entre medias

Con la t de Student comprobamos si existe una *diferencia estadísticamente significativa* entre las medias de dos muestras o grupos; es decir, comprobamos si las dos medias difieren más de lo que consideramos normal cuando las muestras proceden de la misma población o, lo que es lo mismo, si las medias no difieren entre sí más de lo que es normal que difieran los sujetos entre sí.

Con el *análisis de varianza* comprobamos si existen diferencias estadísticamente significativas entre *más de dos grupos,* es decir, comprobamos si las diversas muestras podemos considerarlas muestras aleatorias de la misma población. Es el método apropiado cuando tenemos más de dos grupos en el mismo planteamiento; en vez de comparar las medias de dos en dos, utilizamos el análisis de varianza (y ya veremos por qué).

Cuando tenemos solamente dos muestras también podemos utilizar el análisis de varianza para comparar dos muestras en vez de la t de Student, pero con sólo dos muestras es más cómodo utilizar los procedimientos tradicionales del contraste de medias (t de Student).

Lo que directamente comprobamos en el análisis de varianza es si entre dos o más varianzas existen diferencias estadísticamente significativas, pero lo que realmente deseamos comprobar es si hay diferencias entre una serie de medias.

Lo primero que hay que comprender, al menos de una manera simple e intuitiva, es que al comprobar si hay diferencia entre dos varianzas (enseguida veremos de qué dos varianzas estamos hablando), llegamos a una conclusión sobre si hay diferencias entre las medias.

Vamos a verlo en un ejemplo sencillo, con sólo dos muestras de seis sujetos cada una, representadas en la figura 1.

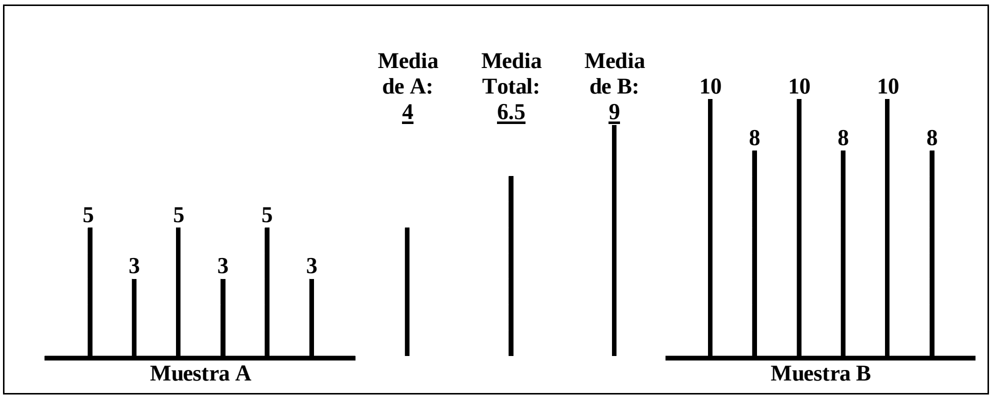


Figura 1

En la figura 1 tenemos representados dos grupos o muestras, muestra A y muestra B, cada una con su media. La media del grupo A es $M_a = 4$ y la media del grupo B es $M_b = 9$.

Si consideramos a todos los sujetos como pertenecientes a un único grupo, A+B, tenemos que la media total es $M_{a+b}= (M_a + M_b)/2 = 6.5$.

Este considerar a *todos los sujetos como hipotéticamente pertenecientes a una única muestra* es importante para entender el procedimiento de análisis de varianza; porque *es esta varianza del grupo total la que vamos a analizar o descomponer.*

En la figura 2 tenemos la representación de los mismos sujetos de los dos grupos de la figura 1, pero ahora unidos gráficamente en un solo grupo.

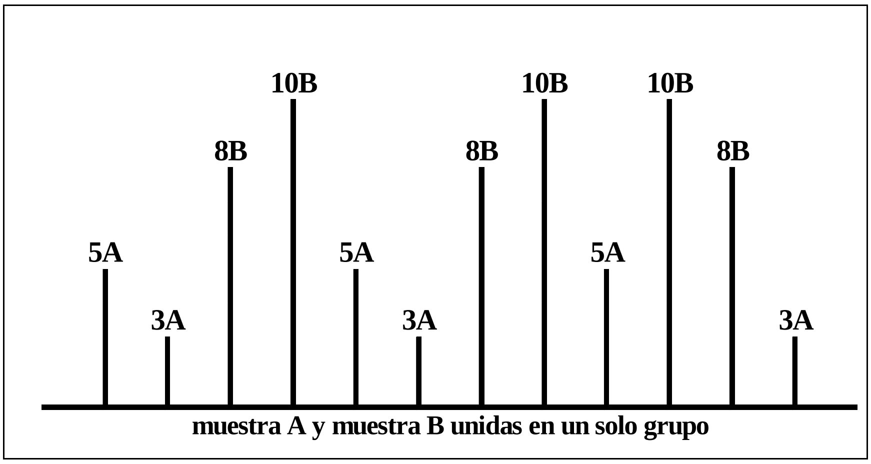


Figura 2

Cuando pensamos en términos del análisis de varianza la *imagen* de la figura 1 debería ser la de la figura 2, donde tenemos *un solo grupo* integrado por los dos grupos iniciales: *es la varianza de este nuevo grupo la que vamos a analizar o 'descomponer'*. De cada sujeto conservamos en esta figura la información sobre su grupo inicial de pertenencia (A o B).

Observando las diferencias entre los sujetos de este grupo *total* podemos preguntarnos: *¿De dónde vienen* las diferencias en este *grupo total* formado por las muestras A y B?

¿De que los sujetos son muy distintos entre sí *dentro* de cada grupo? No, en este ejemplo los sujetos *dentro* de cada grupo tienen un grado semejante de homogeneidad o variabilidad: *dentro* de cada grupo las diferencias entre sujetos (las varianzas) son iguales (si nos fijamos en la figura 1, vemos que en ambos grupos las diferencias entre cualquier par de sujetos o son igual a 0 o son igual a 2).

Lo que sucede es que *las medias son distintas: las medias de los grupos difieren entre sí más que los sujetos entre sí dentro de cada grupo*. Si calculamos la varianza *dentro* de cada uno de los dos grupos (representados en las figuras 1 y 2), veremos que su valor es igual a 1; en cambio si calculamos la varianza *entre* los grupos (utilizando las dos medias como si se tratara de datos de dos sujetos, o utilizando los datos de todos los sujetos, pero asignando a cada uno la media de su grupo) veremos que la varianza es igual a 6.25: es mayor la varianza (*diferencias entre*) de los grupos que la de los sujetos.

La media total ((4+9)/2) es de 6.5; las medias de cada grupo se apartan más de la media total que los sujetos de *su* propia media. Y ésta será la *conclusión importante:*

Si las medias entre sí difieren más que los sujetos entre sí, podemos concluir que las medias son distintas.

Dicho de otra manera, si las medias difieren entre sí más que los sujetos entre sí, concluiremos que las medias pertenecen a muestras que proceden de poblaciones distintas con distinta media; hay una variabilidad mayor entre las medias que entre los sujetos.

En la figura 3 tenemos un caso distinto, con otros dos grupos de seis sujetos. Los dos grupos tienen idéntica media, no difieren *en cuanto grupos,* pero entre los sujetos, dentro de cada grupo, sí hay diferencias.

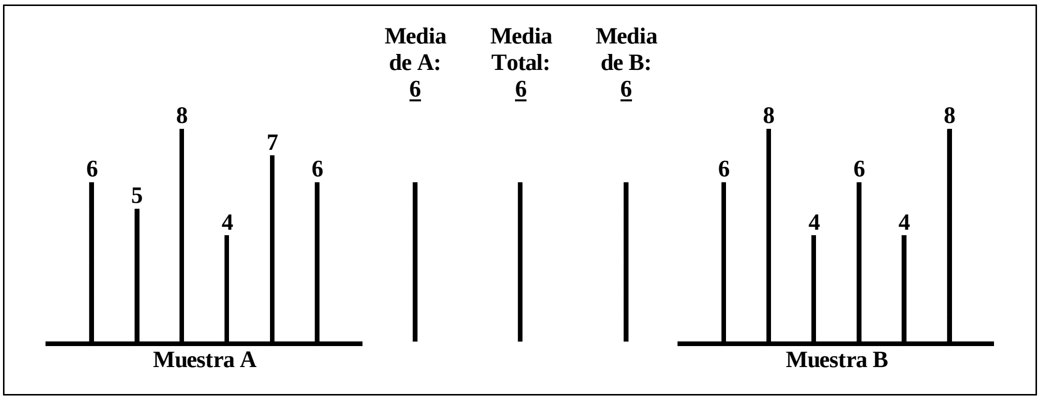


Figura 3

Uniendo ambos grupos, podríamos calcular la varianza total, y preguntarnos de nuevo: ¿De dónde viene esa varianza (esas diferencias)? ¿De que los grupos son distintos, con media distinta, como en el caso anterior? ¿O las diferencias en el *grupo total* vienen simplemente de que los sujetos dentro de cada grupo son distintos?

En este caso las diferencias **no** vienen de *diferencias **entre** los grupos,* que tienen idéntica media, sino de que los sujetos ***dentro*** de cada grupo son muy distintos.

Vamos a suponer que estas puntuaciones son de *autoestima,* y que los dos grupos pertenecen a dos aulas distintas de alumnos. Si comprobamos que la varianza o diversidad *dentro de los grupos* es mayor, o más o menos igual, que la varianza o diversidad *entre los grupos,* nuestra conclusión sería que, *por lo que respecta a la autoestima,* estamos ante un único grupo, o ante dos muestras que representan a la misma población. La hipótesis de *dos grupos,* o de dos muestras procedentes de poblaciones distintas con distinta media en autoestima, no se sostendría.

Podemos imaginar un ejemplo todavía más sencillo: tenemos dos grupos, uno de *enanos* y otro de *gigantes:*

Cada grupo tiene *su* media en altura; la media de los gigantes es mayor que la media de los enanos.

Dentro de cada grupo hay también diferencias; no todos los enanos son igualmente bajitos ni todos los gigantes son igualmente altos.

Pero ¿cuál sería nuestra conclusión si comprobamos que la diferencia entre las medias de los gigantes y de los enanos es más o menos igual a las diferencias que podemos encontrar entre los sujetos dentro de cada grupo?… Pues sencillamente que no tenemos ni enanos ni gigantes, la hipótesis es falsa, y por lo que respecta a estatura, podemos considerar que todos pertenecen al mismo grupo (o hablando con más propiedad, que todos pertenecen a la misma población por lo que respecta a la altura).

El razonamiento para explicar el análisis de varianza (consideramos que dos grupos son distintos cuando la variabilidad *entre* los grupos, entre las medias, es mayor que la variabilidad *dentro* de los grupos) es sencillo y además aplicable a otras situaciones al margen del análisis estadístico. *Dentro* de grupos *oficialmente* distintos en la percepción social (distintos en estatutos, ideario o cualquier otra variable) puede haber diferencias mayores o iguales que las diferencias que se dan por ciertas *entre* los grupos; la única diferencia puede estar en el *cartel* utilizado para designarlos, sin base real para afirmar que en una determinada característica esos grupos tienen medias distintas y constituyen *poblaciones distintas.*

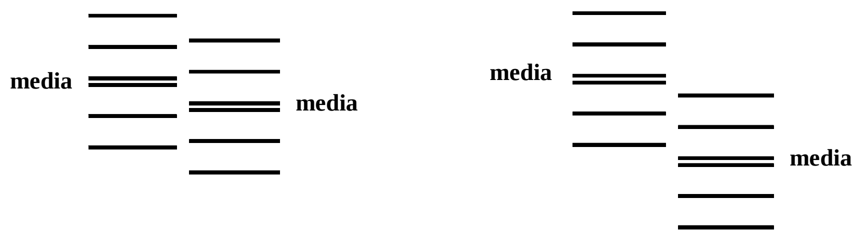
El término *población* se presta a cierta equivocidad en este contexto, sobre todo cuando hablamos de *poblaciones distintas*. En este caso llamamos poblaciones distintas a aquellas poblaciones (tal como nos vienen representadas por muestras concretas) cuyas medias difieren entre sí mas que los sujetos entre sí, aunque hablemos de poblaciones distintas con otros criterios meramente conceptuales o hipotéticos (por ejemplo los alumnos de la facultad A y los alumnos de la facultad B).

Estos ejemplos reflejan una situación sencilla porque se trata solamente de dos grupos; los grupos podrían ser tres o más. Lo que importa ahora es ver que al analizar *varianzas* podemos llegar a conclusiones sobre si hay o no hay diferencias *superiores a lo normal* entre las medias de varias muestras, considerando como *diferencias normales* las que podemos encontrar entre los sujetos del mismo grupo.

Otra manera de representar gráficamente lo que analizamos mediante el análisis de varianza la tenemos en la figura 4. Tenemos representados dos grupos:

En un caso (caso A) las medias difieren entre sí más o menos lo mismo que los sujetos entre sí; podríamos concluir que ambas muestras proceden de la misma población.

En el otro caso (caso B) las medias difieren entre sí más que los sujetos entre sí; en cambio en ambos grupos las diferencias entre los sujetos son de magnitud semejante; *dentro* de cada grupo la varianza es más o menos igual. Nuestra conclusión sería que los grupos son distintos, proceden de poblaciones con media distinta.



Caso A
Las medias no difieren entre sí más de lo que los sujetos difieren entre sí; la varianza *entre* los grupos es más o menos igual a la varianza *dentro* de los grupos.

Caso B
Las medias difieren entre sí más de lo que los suje difieren entre sí; la varianza *entre* los grupos es mu mayor que la varianza *dentro* de los grupos.

Figura 4

Expresado de otra manera: la diversidad o variación que encontramos *dentro* de los grupos (expresada por la varianza *dentro* de los grupos) es la diversidad *normal,* aleatoria; lo *normal* es que no todos los sujetos de una muestra sean idénticos en una determinada característica. Si las medias difieren entre sí (varianza *entre* grupos) más de lo que se puede esperar por azar (varianza *dentro* de los grupos), afirmaremos que las medias son distintas o, lo que es lo mismo (expresado en términos más formales), que las muestras proceden de poblaciones distintas con distinta media.

Básicamente vamos a hacer esto: la varianza *total* (del *gran grupo;* el que resultaría si unimos a todos los sujetos en un único grupo) la vamos a descomponer en dos varianzas;

a) Una varianza nos va a expresar las diferencias *entre* las medias (entre los grupos)

b) Otra varianza nos va a expresar las diferencias o variabilidad entre los sujetos, *dentro* de los grupos (y que consideramos que es la variabilidad normal)

Si la diversidad *entre* las medias (los grupos) es mayor que la diversidad entre los sujetos *dentro* de los grupos, es cuando afirmaremos que entre las medias hay diferencias superiores a lo que podemos encontrar por azar (que es lo que sucede dentro de los grupos).

El análisis de varianza, analizando varios grupos simultáneamente, nos dirá si entre las medias de los grupos hay o no hay diferencias significativas (superiores a la variabilidad normal dentro de los grupos), pero en el caso de que haya diferencias entre los grupos, el *mero* análisis de varianza no dice directamente entre qué grupos está la diferencia; habrá después que comparar los grupos de dos en dos mediante procedimientos análogos (hay varios) a la t de Student, denominados *contrastes posteriores* que expondremos a propósito del análisis de varianza para muestras independientes.

4. Conceptos y *términos* propios del análisis de varianza

Una dificultad inicial que suele presentar el estudio del análisis de varianza es el uso de términos nuevos, por eso es útil familiarizarse con estos términos ya desde el principio. Realmente los conceptos no son nuevos, solamente pueden resultar relativamente nuevos los términos para designarlos. Cuando se cae en la cuenta de que se trata de lo que ya sabemos, desaparece la dificultad.

Recordamos la fórmula de la varianza: $\sigma^2 = \dfrac{\Sigma(X - M)^2}{N - 1}$

Es decir, se trata de una razón o quebrado con un *numerador* y un *denominador* (que ahora es N-1, y no N simplemente, porque se trata de una *estimación* de la varianza de la población). A este numerador y denominador de la varianza nos vamos a ir refiriendo por separado utilizando los nuevos términos, que por otra parte no son arbitrarios y nos ayudarán a entender cómo se *analiza* o *descompone* la varianza.

El numerador de la varianza o suma de cuadrados

La *suma de las diferencias* de todos los datos con respecto a la media, elevadas previamente al cuadrado [$\Sigma(X\text{-}M)^2$] es el numerador de la varianza. A este numerador se le denomina *Suma de Cuadrados* y su símbolo habitual es SC. No es raro encontrarse con el símbolo *SS,* que significa lo mismo pero en inglés *(Sum of Squares).*

La expresión $\Sigma(X\text{-}M)^2$ también suele simbolizarse Σx^2 (la *equis minúscula,* x, es símbolo frecuente de X- M), y también se utiliza a veces Σd^2 (d = diferencia de cada puntuación individual con respecto a la media).

Como la varianza de la *muestra* es $\sigma^2 = \frac{\Sigma(X - M)^2}{N}$

podemos también expresar, y *calcular,* la *suma de cuadrados* [$\Sigma(X\text{-}M)^2$] de esta forma (despejándola de la fórmula precedente):

Numerador de la varianza o *Suma de Cuadrados:* $\Sigma(X\text{-}M)^2 = N\sigma^2$

Esta expresión del numerador de la varianza o *suma de cuadrados* ($N\sigma^2$) es muy importante porque, como ya hemos indicado, facilita mucho el cálculo de la *suma de cuadrados* cuando se dispone de una calculadora con programación estadística que nos da directamente el valor de la desviación típica (σ), como iremos viendo al explicar los diversos métodos[4].

La *Suma de Cuadrados,* o numerador de la varianza, se puede por lo tanto expresar o simbolizar de estas maneras:

Numerador de la varianza o *Suma de Cuadrados:*

$$SC = \Sigma(X\text{-}M)^2 = \Sigma x^2 = \Sigma d^2 = N\sigma^2$$

El denominador de la varianza o grados de libertad

El denominador de la varianza es el *número de sujetos menos uno,* o, según los casos, el número de grupos o número de criterios de clasificación, menos uno (N-1, k-1, etc.). Restamos una unidad porque se trata de estimaciones de la varianza en la *población*. El término habitual de este denominador es *grados de libertad* y ya nos resulta conocido. El símbolo habitual de los grados de libertad es *gl* (en inglés encontraremos el término *degrees of freedom* simbolizado como *df).*

La varianza o cuadrados medios

La varianza es la *razón* entre *la suma de cuadrados* (numerador) y *los grados de libertad* (denominador). La varianza suele denominarse, en este contexto del análisis de varianza, *Cuadrados Medios*[5], y se simboliza como CM (y a veces *MS* o *Mean Squares* en inglés).

Utilizando los diversos símbolos y expresiones habituales, tendremos por lo tanto:

[4] Por otra parte el valor de la desviación típica, y otros datos, podemos encontrarlo ya calculado y *no tenemos necesidad de conocer todos los datos de todos los sujetos* para hacer un análisis de varianza; sobre esto mismo insistiremos en otras ocasiones porque el caer en la cuenta de esta posibilidad es sumamente práctico. Si disponemos *solamente* de estos datos, N, M y σ (y puede ser un caso frecuente) no podemos hacer un análisis de varianza con los programas habituales de ordenador (como el SPSS) y sí podemos hacerlo con una simple calculadora y en algunos programas de Internet que sólo requieren esos datos.

[5] En EXCEL a la varianza o Cuadrados Medios se le denomina *Promedio de los Cuadrados.*

Varianza: $\sigma^2 = \dfrac{\Sigma(X - M)^2}{N - 1}$ = Cuadrados Medios = $\dfrac{\text{Suma de Cuadrados}}{\text{Grados de Libertad}} = \text{CM} = \dfrac{\text{SC}}{\text{gl}}$

5. Cómo podemos analizar (o *descomponer)* la varianza total

La varianza tiene una propiedad que la hace muy útil: puede *descomponerse* y esto permite numerosos análisis.

En el ejemplo de dos (o más) muestras, la varianza total (uniendo las dos muestras en una sola) puede descomponerse en dos varianzas:

1) Una varianza que indica la variabilidad *dentro* de los grupos
2) Otra varianza que expresa la variabilidad (diferencias) *entre* los grupos (entre las medias).

El que la varianza puede descomponerse podemos captarlo en un sencillo ejemplo gráfico. Es muy útil entenderlo aunque sólo sea de manera intuitiva y observando con detención la figura 5, para poder comprender toda la información que nos da el análisis de varianza.

En la figura 5 tenemos representados esquemáticamente (y de manera muy exagerada para hacer más claro el esquema):

1. Dos grupos o muestras, cada uno con su media (M_1 y M_2),
2. El grupo formado por las dos muestras con la *media del total* de ambos grupos (M_T),
3. La puntuación (X) de un sujeto del primer grupo.

Los puntos indicados en la figura 5 representan las dos medias, la media total y la puntuación X de un sujeto concreto del grupo 1 (y podría hacerse la misma representación con todos los sujetos de todos los grupos).

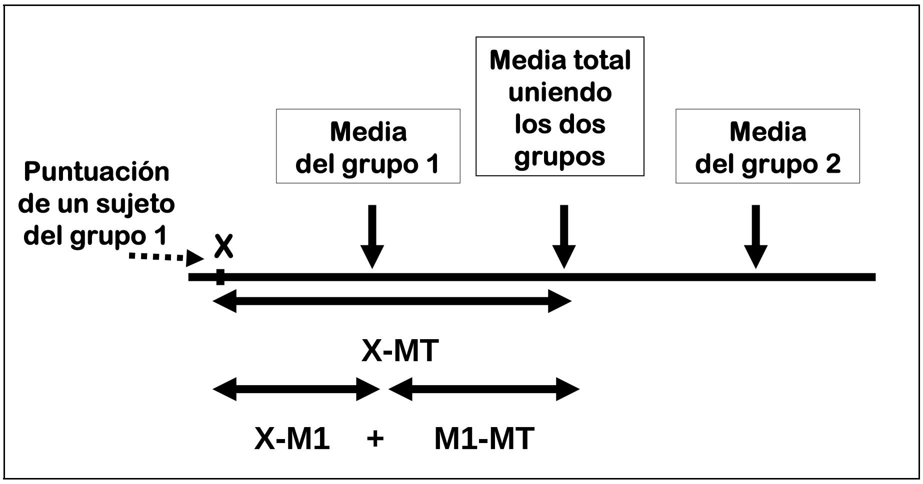


Figura 5

Si vamos a calcular la *varianza del grupo total* (el que resultaría al unir a todos los sujetos en un solo grupo) con media M_T, ésta será la fórmula:

$$\sigma^2_{total} = \frac{\Sigma[X - M_{total}]^2}{N - 1}$$

En el numerador: $\sum(X - M_T)^2$ *(suma de cuadrados)* donde X representa a todas y cada una de las puntuaciones pertenecientes a las dos (o más) muestras.

La contribución a la varianza total de la puntuación del sujeto X señalado en la figura 5 y perteneciente al grupo 1, será:

$$X - M_T$$

Esta diferencia de X con respecto a M_T puede descomponerse en dos diferencias (tal como puede apreciarse gráficamente en la figura 5):

$$X - M_T = (X - M_1) + (M_1 - M_T)$$

La diferencia de cada sujeto con respecto a la media total es igual a:

la diferencia entre esta puntuación y la media de *su* grupo ($X - M_1$)	*más*	la diferencia entre la media de *su* grupo y la media total ($M_1 - M_T$)

Observando la figura 5 se ve con facilidad cómo una diferencia se ha *descompuesto* en la *suma de dos diferencias* que *expresan dos variabilidades:*

La variabilidad que hay *dentro* de los grupos: ($X - M_1$)
La variabilidad que hay *entre* los grupos: ($M_1 - M_T$)

Esta operación la extendemos a todos los sujetos de todos los grupos, así por ejemplo:

para un sujeto del grupo 1: $X - M_T = (X - M_1) + (M_1 - M_T)$;
para un sujeto del grupo 2: $X - M_T = (X - M_2) + (M_2 - M_T)$;

Para todos los sujetos tendríamos lo mismo, tal como se indica en la figura 6.

Suma de Cuadrados ***total*** $\Sigma[X - M_T]^2$	=	**Suma de Cuadrados** ***dentro*** **de los grupos** $\Sigma[X - M_n]^2$	+	**Suma de Cuadrados** ***entre*** **los grupos** $\Sigma[M_n - M_T]^2$
variabilidad total **diferencias de los sujetos con respecto a la media** ***total***	=	***variabilidad dentro de los grupos*** **diferencias de cada sujeto con respecto a la media de** ***su*** **grupo**	+	***variabilidad entre los grupos*** **diferencias de cada media con respecto a la media total**

Figura 6: *cómo descomponemos la suma de cuadrados o numerador de la varianza*

Es decir, la *suma de cuadrados,* o numerador de la varianza, la hemos *descompuesto* en dos sumas de cuadrados:

Una *suma de cuadrados* expresa las diferencias *dentro* de los grupos
Otra *suma de cuadrados* expresa las diferencias *entre* los grupos.

Algo que conviene tener claro es que la varianza, o la variabilidad, *dentro* de los grupos *es independiente* de las diferencias o la variabilidad *entre* las medias:

Si un sujeto del grupo 1 tiene una puntuación de $X = 7$ y la media de su grupo es $M_1 = 5$, su contribución a la varianza o diversidad *dentro* de los grupos va a ser $7-5 = 2$;

Si un sujeto del grupo 2 tiene una puntuación de $X = 15$ y la media de su grupo es $M_2 = 13$, su contribución a la varianza o diversidad *dentro* de los grupos va a ser $15-13 = 2$.

Es decir, ambos sujetos contribuyen en *idéntica cantidad* a la varianza *dentro* de los grupos, aunque las medias de sus grupos respectivos sean muy distintas.

Esto es lo más importante (conceptualmente) de la varianza; son estas *distancias* las que cuantifican la diversidad expresada por la varianza; el denominador lo necesitamos porque en definitiva se trata de *medias,* para que unas sumas de cuadrados sean comparables con otras.

En el denominador, con los *grados de libertad,* sucede lo mismo; los grados de libertad de la varianza total (N -1) se pueden descomponer en grados de libertad *dentro* de los grupos y grados de libertad *entre* los grupos, tal como está resumido en la figura 7.

grados de libertad de la varianza *total* N -1	=	**grados de libertad de la varianza *dentro* de los grupos (N - k)**	+	**grados de libertad de la varianza *entre* los grupos (k - 1)**
N = suma de todos los sujetos (Σn si *n* es el tamaño de cada grupo)		**k = *número de grupos* En cada grupo los grados de libertad son n –1 (n = número de sujetos en cada grupo); si se trata de tres grupos, los grados de libertad son: $(n_1 -1)+(n_2 -1)+(n_3-1)$ o lo que es lo mismo, N- 3**		**o número de grupos menos uno**

Figura 7: *cómo descomponemos los grados de libertad o denominador de la varianza*

Si tenemos tres grupos de 10 sujetos cada uno (N=30), los grados de libertad de la varianza total serán gl = (30-1) = [30-3] + [3-1] = 29:

[30-3]: grados de libertad *dentro* de los grupos = (10-1) + (10-1) + (10-1) (n-1 son los grados de libertad de cada grupo).

[3-1]: grados de libertad *entre* los grupos: número de grupos menos uno.

Esta explicación es literalmente válida para un análisis de varianza hecho con varias muestras independientes (dos o más de dos grupos de sujetos físicamente distintos), pero de manera análoga se puede aplicar a otros modelos de análisis de varianza.

6. Qué comprobamos con el análisis de varianza

Refiriéndonos al análisis de dos o más muestras independientes (y de manera análoga hacemos lo mismo en otros planteamientos), en la Hipótesis Nula se afirma que todas las muestras proceden de la misma población, y que por lo tanto sus medias no difieren significativamente; sus diferencias se explican adecuadamente por el *error muestral* (la variabilidad normal que podemos encontrar en cualquier grupo).

Para comprobar esta hipótesis calculamos *dos estimaciones* de la varianza de esa supuesta misma población, siguiendo caminos distintos e independientes. Si realmente todas las muestras proceden de la misma población, y por lo tanto sus medias no difieren significativamente entre sí, ambos caminos nos llevarán al mismo resultado.

Las dos estimaciones de la varianza (o variabilidad, σ^2) de la población ya las hemos visto:

1º A partir de las *medias* de los grupos, de su variabilidad con respecto a la *media total;* como si asignáramos a cada sujeto la media de su grupo, prescindiendo de las diferencias individuales dentro de cada grupo. Es lo que denominamos *varianza entre grupos;* expresa lo que difieren unos grupos de otros.

2º A partir de las *puntuaciones individuales* con respecto a sus medias respectivas, dentro de cada grupo. Es lo que llamamos *varianza dentro de los grupos;* indica lo que difieren los sujetos entre sí dentro de cada grupo, prescindiendo de las diferencias entre medias, como ya hemos visto.

Estas dos varianzas (*entre* y *dentro* de los grupos), o Cuadrados Medios, las calcularemos dividiendo en cada caso la Suma de Cuadrados por los Grados de Libertad.

Si ambas estimaciones de la varianza son iguales o muy parecidas, podremos afirmar que todas las muestras proceden de la misma población (aceptamos, o no rechazamos, la Hipótesis Nula), y que por lo tanto no difieren significativamente entre sí.

Si por el contrario ambas estimaciones son muy diferentes, y la varianza *entre* los grupos es mayor que la varianza *dentro* de los grupos (es mayor la diferencia entre los grupos que la que encontramos entre los sujetos) podremos inferir que las muestras proceden de poblaciones distintas con distinta media.

Dicho en términos más simples, se trata de *verificar si las medias de los grupos difieren entre sí más que los sujetos entre sí.*

7. Cómo comparamos dos varianzas: la razón F

Para comparar dos varianzas no restamos una de la otra (como hacemos cuando comparamos dos medias) sino que dividimos una por la otra calculando la razón F de Snedecor:[6]

$$F = \frac{\sigma^2_{\text{mayor}}}{\sigma^2_{\text{menor}}} \qquad [2]$$

o según los términos convencionales del análisis de varianza,

$$F = \frac{CM_{\text{entre}}}{CM_{\text{dentro}}} \qquad [3]$$

donde CM = *Cuadrados Medios,* o varianza.

Para entender mejor lo que hacemos mediante la razón F del análisis de varianza podemos pensar en una *analogía* con la t de Student[7]. Con muestras de idéntico tamaño ésta es la fórmula que utilizamos:

$$t = \frac{M_1 - M_2}{\sqrt{\dfrac{\sigma_1^2 + \sigma_2^2}{N-1}}} \qquad [4]$$

En el numerador tenemos la diferencia *entre* las medias de dos muestras. En el denominador vemos las varianzas de los dos grupos, un indicador de las diferencias

---

6 Las tablas de la distribución de F son de Snedecor (y por eso se llaman *tablas de la F de Snedecor*) pero se basó en un trabajo previo de Sir Ronald Aymer Fisher (1890-1962), y en su honor denominó F a este cociente. El análisis de varianza se lo debemos fundamentalmente a Fisher. George Waddle Snedecor (1881-1974) fue el primero en fundar en EE.UU. un departamento de estadística en Iowa State University.

7 Tomamos esta analogía de McGuigan (1994).

dentro de los grupos[8]; es lo mismo que vemos en la fórmula [3], diferencias entre medias en el numerador y diferencias entre sujetos en el denominador.

Podemos ver sin mayor dificultad que obtendremos un valor de t estadísticamente significativo (el cociente será mayor) en la medida en que la diferencia entre las dos medias (numerador) sea mayor y las diferencias dentro de los grupos expresadas en las varianzas del denominador sean más pequeñas. No es algo muy distinto a lo que hacemos en el análisis de varianza: *verificar si las medias difieren entre sí más que los sujetos entre sí.* De hecho, y en el caso de dos muestras, ya veremos que ambos análisis, contraste de medias y análisis de varianza, nos llevan a los mismos resultados y a las mismas conclusiones (F, el estadístico propio del análisis de varianza, es, en el caso de dos grupos, igual a t^2).

¿Qué varianza se pone en el numerador y qué varianza se pone en el denominador?

a) Cuando se comparan (o *contrastan)* dos varianzas mediante la razón F, la norma general es colocar en el numerador la varianza mayor y en el denominador la varianza menor, como se indica en la fórmula [2].

b) En el *análisis de varianza* al calcular la razón F colocamos en el *denominador* la varianza considerada en cada caso como *normal* o *aleatoria,* aunque no sea la más pequeña (aunque casi siempre es la más pequeña), como se indica en la fórmula [3]. Cuando comparamos varias muestras independientes, esta *varianza aleatoria* (que expresa la diversidad o variabilidad normal) es la varianza *dentro* de los grupos, como ya hemos indicado.

En otros planteamientos (muestras relacionadas, diseños factoriales, etc.) cuál es la varianza que va en el *denominador* (la varianza aleatoria, el término de comparación) se indica expresamente en cada caso. En estos planteamientos puede haber más de un razón F pues comparamos varias varianzas (o fuentes, orígenes de diversidad) con la varianza aleatoria o *diversidad normal*.

En la terminología para designar el *denominador* de la razón F cabe cierta confusión porque se emplean indistintamente distintos términos:

Varianza (o cuadrados medios) *dentro de los grupos* (que es lo que es realmente)

Varianza del *término del error* (*erro*r es aquí lo mismo que diferencias aleatorias, normales las que hay en cualquier grupo de sujetos u objetos),

Varianza *residual* (la que nos queda cuando eliminamos otras fuentes sistemáticas de variabilidad como puede ser la pertenencia a uno u otro grupo).

La varianza que colocamos en el numerador es la que nos interesa comparar con la que consideramos normal o aleatoria. Nuestro interés está en comprobar si la varianza del numerador (que expresa las diferencias *entre* los grupos) difiere de la varianza del denominador (que expresa las diferencias *dentro* de los grupos), que es el término de la comparación porque expresa la *variabilidad normal*.

c) Si la varianza del denominador es mayor que la del numerador, no es necesario calcular la razón F; el cociente va a ser inferior a 1 y la diferencia entre las dos varianzas no va a ser estadísticamente significativa. Se puede calcular y poner el dato en su lugar,

[8] En términos propios, el denominador de la t de Student es el *error típico de la diferencia entre medias.*

pero no hace falta consultar las tablas. En lugar de poner p<.05 ó p<.01, pondremos p> .05 (si nuestro nivel de confianza es α = .05) o simplemente *no significativo.*

d) En cualquier caso, al consultar las tablas, donde dice *grados de libertad del cuadrado mayor* hay que entender *grados de libertad de la varianza del numerador* y donde dice *grados de libertad del cuadrado menor* hay que entender *grados de libertad de la varianza del denominador.*

e) Si la razón F es igual a 1, las dos varianzas son iguales. En la medida en que la varianza del numerador sea mayor que la del denominador, el cociente irá siendo mayor que 1. Si los sujetos pertenecen a la misma población y el pertenecer a un grupo u otro no tiene nada que ver con la variable dependiente, es muy improbable obtener valores de F muy grandes. La probabilidad de obtener un cociente F por azar es lo que consultamos en las tablas de Snedecor. Si nuestra F *es muy poco probable* (p < .05) en el caso de que no haya deferencias entre los grupos, nuestra conclusión será que sí hay diferencias. El razonamiento es el mismo que nos hacemos en el contraste de medias.

Normalmente vienen en las tablas dos valores; el primero es el valor que se daría por azar el 5% de las veces (p = .05) y el segundo el 1% de las veces (p = .01); si se superan estos valores lo expresamos así: p<.05 ó p<.01. En estos casos consideramos que la probabilidad de que la diferencia entre las dos varianzas *haya sido una casualidad* es muy pequeña, y por eso afirmamos que las varianzas son *distintas,* o que el valor de F es *estadísticamente significativo*.

Algunos textos traen tablas con otros valores, pero .05 y .01 son las probabilidades que utilizamos habitualmente para aceptar la decisión de no diferencia, como es usual en la investigación experimental en las ciencias sociales. En programas de ordenador (y de Internet) nos viene la probabilidad exacta, y si disponemos de esta información, es la que deberíamos utilizar y comunicar.

El ejemplo explicado corresponde al planteamiento más sencillo, en el que se comparan varias muestras independientes. Si el valor de F es estadísticamente significativo, e indica por lo tanto que la varianza que corresponde al numerador (diferencias *entre* los grupos) es mayor que lo que podríamos esperar por puro azar; pasamos ya a comprobar qué pares de medias difieren significativamente, y según los casos, podemos hacer también otros cálculos adicionales.

8. Explicación alternativa: relación entre variables *cualitativas* o criterios de clasificación (variable independiente) y variables *cuantitativas* (variable dependiente)

Otra manera de presentar lo que hacemos con el análisis de varianza, y que ya hemos enunciado antes brevemente, es ver de qué tipos de datos disponemos y qué información buscamos que nos relaciona los distintos tipos de datos. Esta manera de presentar el análisis de varianza es equivalente a la explicada en el apartado anterior, pero puede ayudar a una comprensión más cabal del procedimiento.

I. Siempre que hacemos un análisis de la varianza tenemos *dos tipos de información* o dos tipos de datos:

a) *Información cuantitativa*. Los datos en *la variable dependiente;* son los *datos que hemos obtenido y tabulado:* la medida de una actitud, una medida de rendimiento académico, etc.; *estos son los datos cuya varianza o diversidad analizamos.*

b) *Información cualitativa*. Tenemos además *otra información* sobre los sujetos: *los criterios o categorías que hemos utilizado para clasificar a los sujetos (variable independiente),* como pueden ser (en los planteamientos más comunes del análisis de varianza):

1. La pertenencia a un grupo u otro, utilizando como único criterio de clasificación el grupo al que pertenecen (en el *análisis de varianza para muestras independientes);*
2. Las preguntas a las que han respondido, experiencias o condiciones por las que han pasado, etc. (*análisis de varianza para muestras relacionadas, los sujetos son los mismos* en las diversas condiciones *o columnas)*;
3. Los dos o más criterios que nos han servido para clasificarlos, para organizar a los sujetos al disponer la tabla de datos, como en el *análisis de varianza de diseños factoriales (tablas de doble entrada);* con dos o más criterios de clasificación, en los que cada criterio está dividido en dos o más niveles, así el criterio sexo tiene dos niveles, hombre-mujer, etc.

II. *Mediante el análisis de varianza podemos relacionar los dos tipos de información:*

La información *cuantitativa,* que son los datos obtenidos y tabulados (*variable dependiente*)

La información *cualitativa,* que son los criterios para clasificar a los sujetos, como el pertenecer a uno u otro grupo (*variable independiente*).

1º Nos hacemos estas preguntas, que son equivalentes:

- La varianza, la diversidad que encontramos en la variable dependiente (la variable medida y tabulada) ¿Está *influida por, tiene que ver con* los criterios de clasificación?;
- El criterio de clasificación (pertenecer a un grupo u otro) ¿Tiene que ver con las diferencias que encontramos en la variable dependiente?
- Este criterio de clasificación, ¿Es una *fuente* de varianza, de diversidad en la variable dependiente? (los criterios de clasificación ¿Son *orígenes* o causas *hipotéticas* de varianza o diversidad en la variable dependiente?).
- ¿Son los sujetos distintos en la variable dependiente, en la variable medida, *porque también* son distintos en el criterio de clasificación (unos han seguido el método A, otros el método B, etc.)?

2º Respondemos a estas preguntas mediante la razón F:

Si la razón F es *significativa* (o lo que es lo mismo, si la varianza del numerador, y que corresponde a los criterios de clasificación, es superior a la varianza aleatoria o *normal* que hemos puesto en el denominador) entonces podemos concluir que los sujetos son distintos en la variable dependiente (la que hemos medido y tabulado) *porque también* (siendo muy prudentes en la interpretación causal) son distintos en la variable o variables que nos ha servido para clasificarlos y cuya varianza está puesta en el numerador de la razón F.

O si se quiere expresar lo mismo de una manera más *cauta,* podemos decir que una F significativa indica diferencias sistemáticas y *coherentes* o *simultáneas* en los *dos* tipos de información, en la variable dependiente que hemos medido y en el criterio de clasificación puesto en el numerador de la razón F: difieren en la variable dependiente (la que hemos medido) y *además* pertenecen a grupos o clasificaciones distintas (existe una relación *de hecho,* cualquiera que sea su explicación, entre los criterios de clasificación y la variable dependiente).

Una razón F significativa nos indica por lo tanto que hay una relación superior a lo aleatorio (o normal) entre

a) la variable que corresponde al numerador de la razón F y
b) la variable en la que hemos medido a los sujetos.

Hay diferencias entre los sujetos en la variable medida *porque* también son diferentes en el criterio de clasificación (o lo que esté puesto en el numerador de la razón F: la varianza correspondiente a un criterio de clasificación o la varianza correspondiente a la relación entre dos o más criterios).

Cuando decimos que hay diferencias en la variable medida *porque* también las hay en el criterio de clasificación no estamos implicando una relación *causal;* podríamos decir con más propiedad (o de manera más descriptiva) que si la razón F es significativa, las diferencias en la variable medida están *asociadas de hecho* a pertenecer a un grupo u otro[9].

III. Una razón F significativa nos permite afirmar que la varianza o diversidad del numerador de la razón F (el pertenecer a un grupo a otro) está relacionada con la varianza o diversidad en la variable medida. *Pero nos falta todavía información para interpretar bien los resultados.*

a) Podemos comprobar entre qué grupos hay una diferencia significativa cuando sea apropiado; la razón F nos dice que *hay diferencias* entre las medias, pero no entre qué grupos se da esa diferencia.

b) Podemos *cuantificar* la magnitud de los resultados mediante los cálculos apropiados (coeficientes de fiabilidad en el caso de muestras relacionadas, y otros coeficientes de asociación en otros planteamientos que iremos viendo). Esta cuantificación (de 0 a 1) nos ayuda a interpretar los resultados, o a comparar dos F significativas y *valorar* su importancia.

Una razón F (o un valor de t o su equivalente) no *cuantifica* la diferencia; simplemente nos permite afirmar que hay diferencias por encima de lo aleatorio, sin responder de manera clara al *mucho* o *poco* de la diferencia. Sin embargo disponemos de análisis adicionales para apreciar la *magnitud* de las diferencias (de la misma manera que en el contraste de medias disponemos del *tamaño del efecto,* concepto que también es aplicable aquí).

c) Siempre hace falta una *valoración conceptual,* lógica, de los resultados, en función del diseño, de otros datos que ya conocemos, etc., y con frecuencia nuestras conclusiones nos sugerirán otras hipótesis, otros diseños, o una repetición del experimento, con otras muestras o en otras circunstancias, para confirmar los resultados.

---

[9] Para inferir *causalidad* tenemos que poder excluir otras explicaciones, y esto lo *intentamos* a través del *muestreo aleatorio* y del *control de otras variables* con un *diseño* apropiado.

9. Diversos modelos de análisis de varianza

En esta explicación introductoria nos estamos refiriendo al planteamiento más sencillo y fácil de entender, el referido a varias muestras independientes, pero el análisis de varianza admite gran variedad de planteamientos distintos y es el método adecuado para *plantear* y *analizar* muchos diseños *experimentales* y *cuasi-experimentales,* y también estudios exploratorios.

Los que vamos a exponer son los siguientes:

1º Para varias muestras independientes
2º Para varias muestras relacionadas
3º Para diseños factoriales (tablas de doble entrada)
4º Para verificar tendencias a creer o decrecer
5º Algunas variantes de los diseños factoriales

Los textos básicos de estadística e investigación suelen traer al menos los dos primeros (para *más de dos* muestras independientes o relacionadas); en ambos modelos *encajan* muchos posibles diseños de investigación. También es frecuente encontrar el modelo más común de análisis de varianza para diseños factoriales; menos frecuente es encontrar en textos básicos el análisis de varianza para verificar *tendencias* (muy útil en investigación sociológica, educacional y psicológica) y las diversas variantes de los diseños factoriales. Hay otros muchos modelos de análisis de varianza que se pueden resolver con facilidad (o al menos entender) mediante procedimientos análogos a los explicados aquí.

El tener a la vista, y con ejemplos resueltos, varios modelos de análisis de varianza es útil por varias razones que se complementan entre sí.

1. El qué hacemos, o qué planteamos, en una investigación depende en buena medida de *qué sabemos hacer.*

Si sabemos de qué análisis disponemos, podemos pensar en planteamientos que nunca se nos habían ocurrido. Por otra parte el *qué sabemos hacer* no es exacto: nos basta saber *qué podemos aprender* o qué podemos buscar o repasar si ha habido un estudio previo; en definitiva lo que importa es saber qué análisis tenemos de alguna manera disponibles.

2. Muchos posibles análisis de varianza coinciden con *diseños experimentales* o *cuasi-experimentales* específicos.

A veces podemos pensar en diseños, estudiados a veces de un modo más teórico y abstracto pero sin referencia a planteamientos y modos de análisis específicos. Esto puede llevar después a diseños mal planteados o inabordables, en definitiva a un aprendizaje inoperante. Una manera de abordar el aprendizaje de los diseños es ver y aprender simultáneamente cómo se pueden analizar los datos en cada diseño. Lo mismo sucede con el control de determinadas variables, que puede verse también incorporado en el planteamiento de algunos modelos de análisis de varianza.

3. Con frecuencia se nos ocurren *preguntas de investigación* a las que después no sabemos dar respuesta.

Puede ser interesante examinar primero posibles repuestas y pensar después *qué preguntas queremos (o podemos) hacernos*… Los análisis estadísticos nos brindan *respuestas* a posibles *preguntas:* si tenemos un repertorio amplio de posibles respuestas, pensaremos con más facilidad en preguntas de interés.

4. En definitiva, y como ya se ha indicado, en el *análisis de varianza* disponemos de *dos tipos de datos:*

a) Datos en la variable dependiente (qué medimos, qué preguntamos, qué observamos… en los sujetos) y…
b) Cómo están *clasificados en categorías*

Lo que hacemos con el análisis de varianza es comprobar si los datos de la variable dependiente *tienen que ver* con *cómo están seleccionados y clasificados* los sujetos (u objetos). En buena medida los modelos de análisis de varianza que vamos a exponer no son otra cosa que *modos de clasificar* a los sujetos, que a su vez nos pueden sugerir numerosas *preguntas* y planteamientos de investigación.

10. Cuestiones metodológicas previas

10.1. Requisitos previos para utilizar el análisis de varianza

En los modelos teóricos en los que se basa el análisis de varianza se hacen tres suposiciones; 1) escalas de intervalo, 2) distribución normal y 3) homogeneidad de varianzas, pero las tres suposiciones son de importancia muy desigual.

1) En la variable dependiente (en la que medimos a los sujetos) tenemos *unidades de intervalo,* (y observaciones independientes)

La primera suposición es que utilizamos escalas de *intervalo,* con una unidad en sentido propio. Esto no suele ser lo habitual en los instrumentos de medición educacional y psicológica (tests, escalas de diverso tipo, preguntas con respuestas graduadas, etc.), pero la mayoría de los autores coinciden en afirmar que con estos instrumentos el análisis de varianza, como otros muchos análisis, son *seguros,* y así lo confirma la práctica más habitual[10]

2) La variable dependiente (la que medimos) sigue la distribución *normal;*

Sobre el presupuesto de *normalidad* en la variable dependiente (la que medimos), una abundante investigación confirma que *en general* la violación de estos presupuestos no invalida de manera apreciable los resultados del análisis de varianza. La violación de la *normalidad* es menos importante *(prácticamente irrelevante,* Glass y Stanley, 1974:373), como está confirmado por numerosos estudios; de hecho *las medias tienden a la distribución normal* aunque las poblaciones de donde proceden no sean normales (Guilford y Fruchter, 1973: 277).

3) Las *varianzas* de las distintas poblaciones representadas en las muestras *no difieren significativamente entre sí.*

La condición previa de *homogeneidad de varianzas* (denominada *homoestacidad*) es sin duda la más importante, aunque la distorsión en los resultados (en relación al error

---

[10] Los métodos habituales de obtención de datos (como escalas tipo-Likert, etc.) *se aproximan suficientemente* a las escalas de intervalo y las distorsiones que se pueden introducir son pequeñas; es más lo que se gana que lo se pierde con estos métodos que dan por hecho que se da una *unidad aproximada* (Guilford, 1954; Nunnally, 1978 y muchos otros). Este punto lo tratamos con más detalle en Morales (2006), cap. 1, apartado 2.3

Tipo I)[11] es pequeña *si el número de sujetos es idéntico* en todas las muestras o submuestras.

¿Qué sucede cuando las varianzas son *muy* desiguales? Al menos hacemos dos observaciones:

a) Las probabilidades que señalan las tablas de la F no son las *reales;* una probabilidad de .05 puede corresponder realmente a un valor menor o mayor, aunque la diferencia entre la probabilidad señalada en las tablas y la real suele ser pequeña.[12]

b) Con grupos de tamaño desigual y varianzas desiguales el que la probabilidad real sea mayor o menor que la indicada por las tablas depende del tamaño de los grupos:

Cuando el grupo mayor tiene también la varianza mayor el valor de F es *conservador:* la probabilidad de que la diferencia entre varianzas sea aleatoria es todavía menor de lo que señalan las tablas.

Cuando el grupo más pequeño tiene la varianza mayor, el valor de F es *liberal:* las probabilidades de que las varianzas difieran son mayores de lo que señalan las tablas[13].

En consecuencia el problema es menor cuando coinciden la muestra mayor y la varianza mayor (aunque podemos quedarnos sin demostrar nuestra hipótesis).

Dos pruebas populares para verificar la homogeneidad de varianzas son la de Bartlett y Levene (*para muestras de tamaño distinto*)[14] y la de Hartley *(para muestras de idéntico tamaño).*[15]

El test de Hartley es muy sencillo; consiste en calcular la razón F con las dos varianzas extremas, dividiendo la varianza mayor de todas por la más pequeña: si vemos en las tablas que la razón F *no* es significativa ya sabemos que se cumple la condición de *homogeneidad de varianzas* (utiliza sus propias tablas, no las convencionales de la razón F). Otra prueba muy utilizada es la de Levene, menos sensible a la no normalidad de las muestras y una buena alternativa a la de Bartlett[16].

Todas estas pruebas previas son sin embargo problemáticas por diversas razones; son muy sensibles a la *no normalidad* (menos la de Levene) y con frecuencia tienen poca

---

[11] Recordamos cuál es el error Tipo I: aceptar la diferencia cuando realmente no la hay (rechazar, o no aceptar, la Hipótesis Nula cuando es verdadera). Es el tipo de error que en principio nos interesa evitar., no equivocarnos al afirmar que hay diferencias superiores a lo normal entre los grupos.

[12] Si las varianzas son muy desiguales puede suceder que un valor de F tenga una probabilidad de .05 en las tablas y realmente esté entre .04 y .07 (Guilford y Fruchter, 1973:277) o entre .07 y .09 (Linton, Gallo y Logan, 1975, que no recomiendan la comprobación previa de estos requisitos).

[13] Lix, Keselman y Keselman (1996), Jaccard (1998: 81); sobre este punto puede verse también Hernández, Borges y Ramírez (1996).

[14] El test de Bartlett se basa en el *ji cuadrado* y se encuentra programado en Internet (*Homogeneity of Multi-variances: The Bartlett's Test* 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 se encuentra en la *Home Page* de Hossein Arsham http://home.ubalt.edu/ntsbarsh/Business-stat/home.html en JavaScript E-labs Learning Objects *(Equality of Mult-variances)*

[15] Estas pruebas se encuentran en numerosos textos; una explicación muy clara puede verse en Escotet (1980). Métodos para verificar la *homogeneidad de varianzas* hay muchos; en Zhang, Shuqiang (1998) se encuentra una exposición crítica de 14 de estas pruebas.

[16] El test de Levene está bien explicado en *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/ Levene Test for Equality of variances http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm (consultado 7, Oct., 2007).

potencia (no controlan bien el error Tipo I) cuando se utilizan como control previo al análisis de varianza.[17]

Sobre *qué hacer en la práctica* se pueden dar unas orientaciones generales que podemos ver en diversos autores. Aunque los programas de análisis estadístico como el SPSS suelen dar los dos resultados (lo mismo que cuando se hace un contraste de medias) suponiendo varianzas iguales y suponiendo varianzas desiguales conviene tener en cuenta estas orientaciones.

a) Las varianzas desiguales no deben preocuparnos si las muestras son de *idéntico tamaño*[18]; por varianzas desiguales podemos entender que la mayor no es más de tres veces mayor que la varianza más pequeña.

b) Tampoco deben preocuparnos las varianzas desiguales si las muestras son de distinto tamaño con tal de que 1) los tamaños de las muestras no difieran mucho y 2) las muestras no tengan menos de 20 sujetos (Jaccard, 1998:81).

c) Por otra parte ya veremos en su lugar que cuando no se cumplen las condiciones previas del análisis de varianza, hay contrastes posteriores muy seguros (como el de Scheffé) y al menos hay un *contraste posterior* específico para cuando los tamaños de las muestras son distintos y las varianzas son desiguales.[19]

d) En algunos casos de varianzas muy desiguales (mucho mayores que las demás) podemos examinar si en alguna muestra hay sujetos *atípicos* con puntuaciones muy extremas *(outliers).* Si hay sujetos muy atípicos, podemos considerar el eliminar de la muestra a estos sujetos (responsables de una varianza mucho mayor). En estos casos habrá que comprobar si estos sujetos tienen alguna característica común para no generalizar los resultados a ese tipo de sujetos. Por ejemplo, podemos encontrarnos con que un subgrupo de sujetos con puntuaciones muy atípicas tienen *también* una edad muy distinta a la de la mayoría, o una *procedencia* distinta, etc.

e) Cuando las varianzas son notablemente distintas, y sobre todo si los grupos son de tamaño distinto, hay otras alternativas al análisis de varianza menos conocidas[20] Con grupos pequeños, de tamaño desigual y con varianzas muy distintas, siempre tenemos las alternativas *no paramétricas*.[21]

Las violaciones de estos presupuestos afectan al análisis de varianza cuando se trata en sentido propio de estadística *inferencial,* es decir, de llegar a conclusiones acerca de las poblaciones representadas por las muestras (como es lo habitual). Si con el análisis de

---

[17] Pueden verse numerosas investigaciones citadas por Jaccard, (1998:82).

[18] Por ejemplo Hays (1981:347), Kirk (1995:101) y muchos otros autores de autoridad reconocida. Para Myers (1972:72-73) aun cuando las varianzas difieran en una proporción de 4 a 1 (o de 3 a 1, Jaccard, 1998:81), la distorsión en los resultados (en relación al error Tipo I) es pequeña pero solamente *si el número de sujetos es idéntico en cada muestra;* en todos estos autores (y en otros como Glass y Stanley, 1974:371) se citan muchas investigaciones que apoyan lo mismo.

[19] Se trata del contraste de Games y Howell, para muestras de tamaño desigual y varianzas desiguales.

[20] Pueden verse expuestas y evaluadas en Lix, Keselman y Keselman, (1996). Estos autores presentan el *estado de la cuestión* y aportan un *meta-análisis* sobre los efectos de las violaciones en el análisis de varianza, con conclusiones que aconsejan cautela cuando las varianzas son *muy* desiguales.

[21] La prueba de Kruskal-Wallis para el caso de varias muestras independientes, o de Friedman para muestras relacionadas; son las alternativas *no paramétricas* más conocidas y populares del análisis de varianza, pero hay más que reseñamos en el anexo VII.

varianza se pretende llegar a un *resumen descriptivo* de lo que está sucediendo en las *muestras* analizadas, estos supuestos dejan de ser importantes.[22]

10.2. Tamaño de los grupos y pérdida de sujetos

El *número de sujetos* en cada grupo (necesario o conveniente) lo tratamos en el anexo VI, junto con los criterios para determinar el *tamaño de la muestra*. Aquí tratamos sobre el tamaño igual o desigual de las muestras y sobre la *pérdida* de sujetos.

Cuando tenemos *varias muestras independientes* y vamos a verificar si entre las medias existen diferencias significativas:

1º Los grupos *pueden ser de tamaño distinto;* el procedimiento es el mismo (con alguna variante menor que indicaremos en su lugar; es más sencillo cuando los grupos son de idéntico tamaño). Ya veremos también que la mayoría de los *contrastes posteriores* suponen el mismo tamaño en los grupos, pero también disponemos de contrastes apropiados cuando los grupos son de distinto tamaño.

2º En principio es *preferible* utilizar muestras de idéntico tamaño por dos razones:

1ª Ya hemos visto que con muestras de idéntico tamaño el análisis de varianza *tolera* mejor el que no se cumplan los requisitos previos (sobre todo el de homogeneidad de varianzas).

2ª Con muestras iguales tenemos disponible un repertorio más amplio de *contrastes posteriores,* y en muchas ocasiones los más aconsejables requieren muestras de idéntico tamaño.

Aún así, cuando se trata de varias muestras independientes, es muy frecuente que los grupos sean de hecho de distinto tamaño, sobre todo cuando comparamos *grupos naturales* (como los alumnos de diversas clases, Facultades, etc.).

Esta recomendación (muestras de idéntico tamaño) es más pertinente en *diseños experimentales* en sentido propio; en estos casos suele ser más viable disponer de grupos de idéntico tamaño. En planteamientos experimentales, hechos frecuentemente con grupos muy pequeños, podemos con más facilidad disponer de grupos con idéntico número de sujetos, bien porque los escogemos así, o bien porque descartamos sujetos aleatoriamente.

Lo que sucede a veces es que en el proceso de la investigación *perdemos* sujetos (sobre todo si hay medidas repetidas en diversos tiempos) y al llegar al análisis nos encontremos con grupos desiguales. Si los grupos son muy pequeños (pongamos por ejemplo $n < 10$) el descartar sujetos supone una pérdida importante en el tamaño de la muestra. En estos casos lo que suele aconsejarse (por ejemplo Denenberg, 1976) es 1º substituir la puntuación que nos falta por la media del grupo pero solamente si nos falta *una* observación o sujeto y además 2º descontar *un* grado de libertad en el *término del error* (el denominador de la razón F).

En los *diseños factoriales* (tablas o cuadros de doble entrada) es más importante disponer del mismo número de sujetos en cada clasificación, como tratamos en el lugar correspondiente.

[22] Lix, Keselman y Keselman (1996: 582).

10.3. Tipos de categorías de clasificación

Las categorías de clasificación (variable independiente) pueden ser de tres tipos: *fijas, aleatorias y mixtas,* como explicamos enseguida.

En el caso de varias muestras independientes las categorías sólo pueden ser o fijas o aleatorias; las mixtas se dan cuando hay más de un criterio de clasificación, como en los diseños factoriales (cuando disponemos los datos en tablas de doble entrada) que veremos más adelante.

1º *Categorías fijas*

Son categorías fijas las escogidas arbitrariamente por el investigador, y es posiblemente el caso más frecuente; como ejemplos podemos pensar en:

a) Alternativas obvias y con frecuencia las únicas disponibles: sexo, pertenencia a un grupo, etc.

b) Variables *cuantitativas,* como tiempo dedicado a una tarea (una hora, dos horas, etc.), número de experiencias, *nota media* previa, *edad* (agrupándolas en varios niveles), etc.

En estos casos suelen escogerse como criterios de clasificación algunas *categorías-tipo* entre todas las posibles, pero no por azar (aleatoriamente) sino con algún criterio lógico. Si, por ejemplo, una categoría es el tiempo dedicado a una actividad, y el máximo tiempo posible es hora y media, se pueden clasificar los sujetos según dediquen a la actividad media hora, una hora o una hora y media. El número de niveles o subcategorías puede ser tan grande como se estime oportuno.

c) Diversas *variantes* de un método, condición, etc.

Categorías fijas son en definitiva cualquier criterio de clasificación que siga un criterio lógico como base para la clasificación (método, modalidad, grupo, etc.). Las categorías fijas se definen quizás mejor por lo que no son: no son niveles o categorías de clasificación seleccionadas aleatoriamente entre otras semejantes *de la misma población,* como explicamos a continuación.

2º *Categorías aleatorias*

Son las escogidas *aleatoriamente* entre una población mayor. Un ejemplo puede ser éste:

Se desea comprobar, por ejemplo, si el rendimiento escolar depende del tipo de centro o del tipo de profesor. Las categorías de clasificación van a ser en estos casos profesor y centro escolar, pero son muchos los posibles centros y los posibles profesores en una zona dada. Si en este caso escogemos profesores y centros aleatoriamente, tenemos categorías aleatorias, y las conclusiones podremos generalizarlas a las *poblaciones* de centros o de profesores.

Un ejemplo clásico de *categorías aleatorias* se da en las investigaciones en las que se pretende comprobar si el *orden* con que se presentan las preguntas de un cuestionario influye en cómo se responde a estas preguntas. En un cuestionario de seis preguntas se pueden hacer 720 combinaciones alterando el orden. Evidentemente son demasiadas posibilidades. El investigador puede escoger un número limitado de versiones del mismo cuestionario, por ejemplo cinco o seis, a cada versión responde una muestra de sujetos y las conclusiones pueden extrapolarse entonces con más seguridad a la *población* de posibles maneras de ordenar las preguntas[23].

3º *Categorías mixtas*.

Cuando tenemos más de una categoría de clasificación, una categoría puede ser fija y la otra aleatoria; por ejemplo métodos didácticos escogidos con un criterio lógico (o simplemente los métodos disponibles) y centros escolares escogidos aleatoriamente en los que se van a aplicar los diversos métodos.

Las categorías de clasificación más frecuentes (y más fáciles al planificar una investigación) son las categorías fijas.

En la práctica las repercusiones del tipo de categorías son dos:

a) La posibilidad de extrapolar las conclusiones a la *población de categorías* (de centros, de profesores, etc.) cuando estas han sido seleccionadas aleatoriamente.

b) Algunas peculiaridades metodológicas que iremos viendo en su lugar; en algunos casos (como en los diseños factoriales) el denominador de la razón F va a variar en función del tipo de categorías.

11. Referencias bibliográficas

DENENBERG, VICTOR H., (1976). *Statistics and Experimental Design for Behavioral and Biological Researchers.* New York: John Wiley & Sons,

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

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].

HAYS, WILLIAM L. (1981). *Statistics*. Third Edition. New York: Holt, Rinehart and Wilson. 713 pp.

HERNÁNDEZ, JUAN A.; BORGES, ÁFRICA y RAMÍREZ, GUSTAVO (1996). Análisis de la robustez del ANOVA en el caso de tamaños muestrales distintos y no proporcionales frente a los contrastes de aleatorización. *Psicológica,* 17, 127-141.

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

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

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

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.

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

[23] Este ejemplo, y algunas variantes del mismo, puede verse bien explicado en Iversen y Norpoth (1987)

MORALES, VALLEJO, PEDRO (1996). *Medición de actitudes en Psicología y Educación*. 3ª edición. Madrid: Universidad Pontificia Comillas.

MYERS, JEROME L., (1972). *Fundamentals of Experimental Designs.* Boston: Allyn and Bacon.

NUNNALLY, JUM C. (1978). *Psychometric Theory*. New York: McGraw-Hill.

PEARSON, E. S. and KENDALL, MAURICE, (1978). *Studies in the History of Statistics and Probability,* Volume I. London: Griffin & Co. Limited.

ROSENTHAL, ROBERT and ROSNOW, RALPH L. (1991). *Essentials of Behavioral Research, Methods and Data Analysis.* Boston: McGraw-Hill.

ROSENTHAL, ROBERT, (1987). *Judgment Studies, Design, analysis and meta-analysis.* Cambridge: Cambridge University Press.

ROSENTHAL, ROBERT, (1991). *Meta-analysis procedures for social research.* Beverly Hills, CA: Sage.

SPATZ, CHRIS (1993). *Basic Statistics: Tales of Distributions*, 5th Edit. Pacific Grove, CA: Brooks/Cole

ZHANG, SHUQIANG (1998). *Fourteen Homogeneity of Variance Tests: When and How to Use Them.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, California.